IMPACT ASSESSMENT DISCUSSION PAPER NO. 5

SOME USEFUL METHODS FOR MEASURING THE BENEFITS OF SOCIAL SCIENCE RESEARCH

Henry E. Kilpatrick, Jr.

Director General s Office

International Food Policy Research Institute
2033 K Street, N.W.
Washington, D.C. 20006
Tel: (202) 862-5600
Fax: (202) 467-4439
Email: IFPRI@CGIAR.ORG

October 1998

Little is known about the impact of social science research in general, and food policy research, in particular. In order to expand the scope of available academic research and to develop quantitative methods for estimating the impact of IFPRI's work, several papers were commissioned from social scientists. Furthermore, IFPRI held an essay contest to solicit research from a broader range of scientists. The resulting papers were discussed at a two-day symposium organized by IFPRI in 1997. This Discussion Paper is a revised version of a paper prepared for and discussed at the symposium. Other papers will be published in this Discussion Paper series over the next months.

CONTENTS

ABSTRACT

What are the “returns” to policy-oriented research in the social sciences? One presumes that the positive net benefits to society, or at least a certain segment of society, would be treated as returns, but how does one determine what these benefits are? Clearly benefits to some social science research are available because society continued to fund it, albeit at different levels in different locations and times. This paper cannot fully answer the questions of what it is we seek to measure in any empirical sense, although it will discuss this issue. The returns in the marketplace for social science research are those that exist in the eye of the customer who bears the cost of the research. This paper's primary goal is to offer the client some ways of measuring these returns. It does this with particular emphasis on methods that are often overlooked, even though some of them have been available to the analyst for decades. It also explains some of the costs and benefits of each method and explains how some of them may be used together in order to achieve a higher level of efficacy in measurement.

INTRODUCTION

What are the returns to policy-oriented research in the social sciences? One presumes that positive net benefits to society, or at least certain segments of society, would be treated as returns, but how does one determine what these benefits are? Clearly there must be benefits of some sort to some social science research because society continues to fund it, albeit at different levels in different locations and times. This paper cannot fully answer in any empirical sense the question of what to measure. The returns (or “expected returns”) in the marketplace for social science research are those that exist in the eye of the customer who bears the cost of the research. This paper's primary goal is to offer that customer some ways of measuring these returns. It does this with particular emphasis on methods that are often overlooked, though some of them have been available for decades.

How do we determine the returns to policy-oriented research in the social sciences? First, research should be evaluated both ex ante and ex post. An ex ante evaluation of potential benefits is needed to determine whether research should be funded to begin with. Ex post evaluations are needed to determine whether additional research should be funded (one might also call these ex ante evaluations occurring before the next round of funding). Ex post evaluations can also determine whether the research has paid off and whether it should be given additional funds. These additional ex post evaluations may be based upon evaluating programs that are put into effect as a result of the research.

Since the benefits of social science research may accrue slowly,[1] the methods used to determine whether to conduct additional research at an early stage (that is, to make a first assessment of benefits from completed research) may be different from the methods used when the results can be evaluated more fully. However, if one waits until all costs and benefits are fully measured before providing additional funding for promising research, start-up costs must again be incurred and institutional knowledge may be lost.

This paper will offer an overview of methodologies that can be used for ex post and ex ante evaluations of the benefits of social science research. Its primary focus will be on research that supports policy, but it should be noted that research must be evaluated continuously. The evaluation should begin before the actual research is undertaken and only end after a program to disseminate the benefits from research is put into place or even later.

THE LITERATURE ON MEASURING THE BENEFITS OF SCIENTIFIC RESEARCH

The literature on the measurement of the benefits of policy-oriented social science research is sparse. There is no coherent body of theory to draw upon that does not draw heavily from theory established in another field. However, literatures are available on program evaluation (including policy evaluation), research and development (R&D), basic science research evaluation, economic analysis, and policy analysis (the public policy process). As Wagner (1995, 4) contends, there is no theoretical framework within which to place quantitative tools for measuring the output of scientific inquiry, although there is a comprehensive literature that examines how to calculate the returns to scientific research and R&D expenditures.[2] These literatures will form the basis for many of the recommendations for the evaluation of social science research that are considered in this paper.

THE UNCERTAIN RELATIONSHIP OF RESEARCH TO POLICY

While policy-oriented social science research is rarely, if ever, translated directly into policy in its entirety, there are other things evaluators must be concerned about before they determine how to measure its benefits. Assuming that the particular research in question is propitious, evaluators are still faced with a number of problems beyond the control of the researchers. Good research might not become policy because of laws, institutions, culture, or opposition by interest groups or the public at large to the proposed program. As Quade and Carter (1989, 7) state, “We may be able to design an efficient system on paper, but as of yet we have no algorithms for finding ways to overcome the resistance offered by tradition, legal restrictions, and a host of privileged interests that inhibit radical or even morphological change.” Competent research may be translated into policy but the underlying conditions that the research is based upon may change. Finally, research may be translated into policy, but unanticipated consequences may negate any positive results from that policy.

DESIGNING SOCIAL SCIENCE RESEARCH EVALUATION

A design is necessary to measure the benefits and otherwise evaluate social science research. Furthermore, while it may be intuitively obvious to many, readers should recognize that an analysis to determine the benefits of social science research might also be characterized as “research” of sorts.

In evaluating social science research, one must first determine the purpose of the analysis. Will it be conducted to determine whether the research is profitable for society, or at least for the financier? Is it to decide whether further research should be funded? Next, a clear, precise, and manageable process needs to be created that will elicit information about the research that is to be analyzed (Quade 1989, 137).

Following this, the analysis must clearly define what is to be measured. It has already been pointed out that research is but one step in a process. Hence, the actual research must be isolated from the program that implements it. The research to be evaluated must also be separated from related research and the rest of the surrounding policies and processes. Furthermore, the question of whether the evaluation analysis is to cover one research project or a group of projects must be answered. If the target research is a group of projects, it must be determined whether one evaluation methodology or a mix of methodologies will be used to measure the entire range of research. After isolating the research, a metric must be designed to measure or rank it against a scale. The users of the research must be clearly defined. While this may seem obvious, there may be multiple users and spillover effects. Furthermore, as Levy and Terleckyj (1996, 7) point out, new knowledge may enhance the value of old, ignored, used up, or simply unrecognized research by lending new life to previous research results. This effect, if it is present, should be captured in any measure of benefits.

How rigorous must the analysis be? As Newcomer, Hatrey, and Wholey (1994, 1) note, rigor in program evaluation can be costly. The same is true in measuring the benefits of social science research. Hence, the rigor of the specific analysis should be defined before undertaking the analysis.

An empirical method of analysis that will accomplish the above program must be chosen. The analysis must be carried out with as little bias as possible, and the results of the analysis must be reported clearly and precisely. This report must also specify the limitations of the evaluation methodology or methodologies that are used.

With this framework in mind, the paper will proceed to the nuts and bolts of the evaluation process. It should be noted that in actual operation the methodologies to be discussed will generally fall short of the mark set by theory. The ability of an evaluator to isolate a particular body of research is questionable. Rigor is a matter of judgment. In fact, the evaluators' judgments, and sometimes the judgments of those the evaluator relies upon, are crucial elements of every methodology that is available.

THE FIRST STEP: REPLICATION OF ORIGINAL STUDIES

Replication is used extensively in research in science and technology, but it is more difficult in social science where controlled experiments are generally not as available as they are in the “purer”sciences. “Reality” is a social construct in the social sciences, unstable, and subject to continuous change. It is more dependent on culture than pure scientific research is.

One might ask why replication or other indirect tools such as bibliometrics should be used to determine the benefits of social science research. In fact, one might argue that replication is itself a form of research and not just a method of evaluation. As already noted, benefits may accrue slowly. One needs a means of determining whether to fund additional research, and a check on bias in the research is useful. Methods are needed to ensure that social science research that appears to be beneficial reports results correctly before it is used to develop policies and before the research itself is evaluated further.

Replication in the social science is often left to graduate students, if it is done at all.[3] Social science journals apparently do not wish to use scarce space for replication studies (Feigenbaum and Levy 1993, 1996). The system of professional honors and rewards in the social sciences places a lower value on replication than it does on original work. Weiss (1972) maintains that the repetition of results is the basis of scientific generalization, yet replication remains a rarity in evaluation research. As argued further by Feigenbaum and Levy (1990, 1993), the ability to reproduce a researcher's methods in order to validate results is a clearly recognized requirement of scientific inquiry and is *de rigueur* in the sciences. Hence, it should become a part of social science research evaluation as a first step in establishing the benefits of a particular work or group of studies.[4]

The simplest approach to replication is to repeat the experiment exactly using the same data set. This is often possible in social science research where one performs statistical analyses on a data set of historical events. One might also consider multiple cases just as one would consider multiple experiments (see Caudle in Wholey, Hatrey, and Newcomer 1994, 91). In this case, literal replication is defined as the selection of a case so that it predicts similar results. Theoretical replication produces contrary results but for predictable reasons.

There are several reasons for conducting replication studies. First, they provide useful checks on the work of the original researchers. In particular, they provide evidence about whether the procedures were carried out as described in the original paper. Second, a number of different procedures can be used in addition to the ones used originally. This provides evidence about whether different methods produce different results. If they do not, the study can be replicated successfully. If they do produce different results, judgments must be made about why this is true. There is a third reason for replication. Evaluating the original research will be futile if there are major problems with it. Hence, it is useful to carefully examine the work for accuracy before proceeding further.

BIBLIOMETRICS

Bibliometrics, or the measurement of published materials such as citations and publication counts, is often used as a proxy to measure returns to R&D spending and to evaluate scientific programs.[5] It has been used since the turn of the century (Melkers, 1993, 45). Garfield (1963) and Price (1961) pioneered its use in modern scholarship. According to Cozzens (1995, 26–27), this method is used extensively in the British university system as well as in the National Institutes of Health and the National Science Foundation in the United States. Bibliometrics is useful in showing what work researchers believe to be important, since citations are an indication of a researcher's familiarity with a particular literature and the relevance of a particular article or study to a researcher's own work. According to Sarafoglou and Haynes (1996, 288), the method can address several quantitative aspects of publications: the number of citations of an author, the growth rates of the literature in journals, the time lags between reception, acceptance, and publication of research, and production patterns in scientific literature. A number of methods are used to analyze citations. These range from simple counts to linking citations to each other.

However, bibliometrics has a number of flaws. The esteem with which a particular author holds a paper that he or she cites cannot be garnered simply from citations. Bibliometrics cannot measure quality. As Popper (1995, 10) points out, the discredited cold fusion paper of Pons and Fleischmann was among the 10 most heavily cited papers in the scientific literature for several years. Or, as Sarafoglou and Haynes (1996, 289) ask, “should more weight be given to Samuelson's text, *Economics*, than to his path-breaking book, *Foundations*, just because of its higher citation score?” Publication is more important in some disciplines than others (Cozzens 1995, 27). Finally, bibliometrics cannot measure the benefits to society at large from the research undertaken.

One of the characteristics that makes bibliometrics suitable to evaluation is that it is a measure of the original research rather than the programs that may result from it. Given its flaws, however, it may be useful as a supplement to other methods, but cannot stand alone as a method for evaluating social science research. Melkers lists five guidelines for enhancing bibliometric analysis in R&D research. These guidelines may be useful in any application to the social sciences:

1. Use it on the appropriate unit of analysis (groups rather than individuals);
2. Use it only as a partial indicator;
3. Use it for projects that have a strong scientific quality;
4. Do not try to account for quality with bibliometrics; and

5. Provide information about what data indicate and what they do not when presenting evaluation results.

OTHER PEER REVIEW MEASURES

Measures of rating research that are similar to bibliometrics include panels of experts convened to provide peer review. This is the most widely used methodology in the world for research evaluation, according to Cozzens (1995, 29). It was the earliest method used—in 1665—according to Barry Bozeman (1994, 80). Peer review relies upon the expert judgment of colleagues within the same field. Many R&D laboratories and government agencies use it. In fact, social science research contracts are likely to be awarded as a result of it. There are a number of different methods of peer review, which are themselves reviewed in Bozeman (1993).

The results of these measures depend upon the choice of reviewers. Furthermore, this methodology does not satisfy the yearning for a less judgment-oriented, more direct, and more empirically based means of measurement. Peer-based appraisals such as the enumeration of awards and other recognition from professional societies and patent counts (which have at best limited use for social science research) provide quantitative means for measuring output, but not scalable ones. Peer review may be used with other methods, but, like bibliometrics, it cannot stand alone as a means for measuring social science output.

CASE STUDY METHODOLOGY

Case studies make use of historical comparative data to tell a story that is richer in detail than statistical analysis, which by its nature must be somewhat abstract. Yin (1989, 23) defines this methodology as follows:

> *A case study is an empirical inquiry that* investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used. [Emphasis in original.]

Yin (1989, 146–151) also lists five components of an exemplary case study: It must be significant. It must contain all the relevant information that is necessary. The case study must consider alternative perspectives. It must display sufficient evidence, and there must be a valid chain of evidence presented. Lastly, the case study must be engaging.

Case studies can be a viable tool for the evaluation of social science research if applied carefully. For one thing, the evaluator can tell how research was translated into policy and how the policy then met the goals of the research. While modeling and advanced statistical methods are rarely used in case studies, statistical data in the form of charts and tables have long been deemed necessary (Davis 1955). Case studies can be qualitative, quantitative, or both.

Yin (1989, 21–22) notes that the primary weaknesses of case studies are that they may lack rigor, they may provide little basis for scientific generalization, and they can often take too long and result in massive, unreadable documents. Kingsley (1993, 36) lays out means of overcoming these weaknesses. As modified for use in evaluating social science research, his recommendations include, first, be specific about the research question (which would include the isolation of the research from the program that carries it out), and, second, develop a rationale for the methods used and how these methods address the specific needs of the evaluation. Kingsley also suggests that case studies examine both successful and unsuccessful projects. However, this may use resources that could be put to better use in evaluating research projects that appear to be successful in their early stages.

MORE DIRECT METHODS: USER EVALUATIONS

User evaluations entail surveying the users of the research that is implemented by a policy program to elicit information on the value of the research that went into it.[6] This is a costly form of analysis. It may involve interviews, surveys conducted by mail or telephone, and the development of a survey questionnaire. The metric is ordinal rather than cardinal. As Cozzens (1995, 35) points out, there may be an inherent conflict of interest in surveying users, particularly if they receive free or low-cost benefits. They have a vested interest in expressing satisfaction in order to keep the benefits flowing.

A survey reflects the biases of those surveyed as well as the biases inherent in the questions posed to the respondents (or even the order in which the questions are asked). None of the methods described here is perfect. However, given their costs, surveys do not appear to be among the best means of evaluating social science research, though they can determine whether the beneficiaries were satisfied, particularly if the benefits to a broader society are of greater interest than benefits to a select group.

Focus groups—informal small group discussions conducted by a moderator—are another means of user evaluation. They are somewhat less costly than surveys, but the selection of participants is not necessarily random, and the output is often difficult to present systematically (other than as an account of the reactions of the group). This technique is probably better reserved for the evaluation of programs and customer service, when the provider wishes to improve delivery of a product or service.

BENEFIT-COST ANALYSIS OR THE MEASUREMENT OF SOCIAL RATES OF RETURN

Benefit-cost analyses, which have been most popular in environmental studies, seek to measure the benefits and costs of a given project systematically and rationally. The logic underpinning benefit-cost analysis is simple. One calculates both the benefits and costs of a project or program in monetary terms. Then one compares the two. Or one may calculate and compare marginal costs and benefits.[7] Financial methods such as measurement of the "internal rates of return" (the discount rate that makes the present value of estimated benefits equal the present value of estimated costs) may be used with this methodology to calculate the benefits of a project (assuming that the benefits outweigh the costs).

While this is rudimentary enough in principle, it poses great challenges in practice. The quantification of costs and benefits may prove elusive or incomplete at best. Particularly thorny is the question of what value to put on human life and health (Weiss 1978, 84–88). Judgements must be made by the analyst, and critical assumptions will be buried in the calculations (Quade 1989, 225). Kee (1994, 487) labels evaluators' attempts to hide the messiness of the process from decisionmakers the "black box syndrome." Moreover, the process is prone to manipulation by those who are determined to reach a particular outcome. For example, the discount rate is never known with certainty and can be manipulated.[8] Finally, externalities are difficult to estimate with simple benefit-cost approaches (Stough and Rietveld 1996, 12).

While formal benefit-cost analysis has lost much of the glamour it had two decades ago, it can still be useful in evaluating policy-oriented research as long as the evaluator clearly spells out the premises that lie behind his or her judgments and spells out the calculations. It should be pointed out that it may prove difficult to separate the results of the research from the results of implementation, which usually modifies the research in practice. Furthermore, while formal benefit-cost analysis may have lost some of its luster, all methods of evaluation require the analyst to make a determination of benefits versus costs.

STATISTICAL METHODS: REGRESSION ANALYSIS

Regression analysis takes the form of the standard regression equation

$$\gamma = \alpha + \beta X + \varepsilon ,$$

where γ is the dependent variable, α is a constant, X represents the independent variables with coefficient β, and ε is the error term. The method fits a line or curve to the data by the means of least squares. It is a well known, but intricate, statistical technique. This paper cannot be a training manual for the general use of regression analysis. It will instead discuss the possibilities it has for evaluating social science research.

Popper (1995) refers to analyses of R&D using regression methods as "production function analyses," since they use production functions. The engineering and economic concept of *a* production function is simply the change in output due to a change in an index of inputs. Formally, a production function may be defined as follows:

Q = f (L, K, H) and is homogeneous of degree *h* at time period 1 where
Q = output,
K = capital input,
L = labor input,
H = human capital, and

If inputs are increased in production at time period 2 such that

$$c^{h}Q = f(cL, cK, cH)$$

there exist increasing returns to scale if $h > 1$, decreasing returns to scale if $h < 1$, and constant returns to scale if $h = 1$. Or to state it simply: if inputs are doubled, outputs will double under constant returns, more than double under increasing returns, and less than double under decreasing returns.

The production function can be estimated using the Cobb-Douglas or other forms of production function modeling. Popper (1995) contains an adequate discussion of the techniques. This methodology has been used with reasonable success in R&D research by employing aggregate data.[9] An assessment of the benefits of specific research projects through production functions may prove to be highly elusive, particularly in social science research. There is an advantage in using them to assess the benefits of R&D in private markets that does not exist for public markets—the existence of measurable input to firms and measurable output in the form of goods and services produced. However, it is possible to create a rough model of the results of social science research. A time series of program budgets minus personnel costs can serve as a proxy for capital costs, personnel costs can be a proxy for labor, and the variable the research is designed to act upon can be the output. In particular, one may want to determine whether certain social science research leads to increasing returns to scale.

Regression analysis does not need to take the form of a production function to be useful in evaluation. Furthermore, one may be able to isolate the original research from the program at least partially by proper selection of variables.[10]

The pitfalls of regression are that it relinquishes rich details to create a workable model, the model may be misspecified (through faulty judgment and subjectivity), and it may create a false sense of quantitative certainty. As in any analysis, the evaluator must be clear about the variables selected and the reasons for their selection. Nonetheless, regression analysis is likely to be a workable way to evaluate the benefits of social science research in some cases.

OPERATIONS RESEARCH MODELING

Operations research modeling techniques, which were developed in the 1940s, have also been used to assess programs. Linear programming is a quantitative methodology that makes use of the mathematical modeling of constrained optimization. It is used to find an optimal solution, given certain constraints imposed by the model. Several methods are available besides simple linear programming. These include goal programming and data envelopment analysis (DEA). DEA estimates the frontier of a production function rather than minimizing the sum of the squares as in regression analysis.

Goal programming, a multiple-criteria operations research model of decisionmaking, may be traced to a paper by Charnes, Cooper, and Ferguson (1955). It is an attempt to apply operations research techniques in modeling the concept of "satisficing." The technique puts goals and aspirations (which may not necessarily be achieved) on one side of the equation and constraints that must be satisfied on the other (see Carlos Romero 1990). The priorities of individuals and groups may be taken into account through goal programming (this cannot be done in traditional linear programming models).

This method does not appear to have been discussed much in the literature on evaluation, perhaps because it is associated with applied science and engineering, but it holds some promise as a quantitative methodology for the purposes specified in this paper. Selecting the researcher's goals using goal programming, rather than selecting the goals of a program, which may be divergent, may be a way to isolate the effects of the research from the program. This method should receive further study.

As noted, DEA includes the concept of the production function that was mentioned in the section on regression analysis. Rather than fitting a line using least squares, it fits a boundary that contains all points within it (the “production possibilities frontier” in economics jargon). This method is likely to be successful in cases where production functions are appropriate. The use of both methods will increase the rigor of the study, but, of course, this has a cost (for research applications of DEA, see Charnes, Cooper, Lewin, and Seiford 1994).

SIMULATION: RECENT DEVELOPMENTS IN OPERATIONS RESEARCH AND COMPUTATIONAL ECONOMICS

Simulation is a heuristic, probability-based methodology that conducts experiments by proxy using a model and “runs” of the values of uncontrolled variables. This is accomplished by the use of Monte Carlo analysis, or “repeated simulation of a stochastic model to investigate the properties of statistical techniques applied to it” (Cragg 1990, 171). An abstract model that contains enough stylized facts to approach authenticity is necessary for simulation. This technique has a long history in the applied sciences, particularly in defense work. Eppen, Gould, and Schmidt (1988, 674–675) list a number of the advantages and limitations that simulation has for modeling in general. Some of its advantages are that simulation can deal with uncertainty, it is versatile because it can be run many times, it can reproduce variability in the system being simulated, it can analyze multiple objectives, and it is rational. But it may foster a temptation to build a model that is too big and too expensive. Also, one could begin to create a model only to find out that no one understands the interactions of the variables well enough to build one that is workable. Judgement and elements of subjectivity enter into the specification of the model as they do in other methodologies. Finally, it must be said that simulation uses only limited real world data.

Simulation has been used most heavily in the applied sciences. It may prove to be useful in evaluating social science research when data are hard to come by or one suspects processes are at work that are difficult to model by conventional means. Simulation is currently used in social science research. Two examples are Nelson and Winter's (1982) models of evolutionary change and the sugarscape model of Epstein and Axtell (1996). The latter attempts to model social science systems from the ground up, an ambitious effort that probably falls short of its goal, but produces useful modeling techniques. Additional simulation models that use genetic algorithms, nonlinear regressions, and other modeling techniques are being developed in the growing field of computational economics (see Gilli 1996 for examples).

Simulation is useful in determining the sensitivity of a system and the range of options for intervention. This could be termed “scenario analysis.” Of course, the major problem with simulation methods is that they do not use actual observed data, although the models may take actual data into consideration when generating scenarios. Hence, the output is more likely to reflect a potential reality than measure actual circumstances. However, simulation can serve as a prelude to the use of other evaluation methods.

MIXING METHODS FOR BETTER EVALUATIONS

It is often possible to combine several methods to obtain better evaluations. If production function analysis is deemed appropriate for evaluating research, both regression analysis and DEA may be used to estimate the production function and one could extend this analysis to use a number of robust regression methods. Sarafoglou and Haynes (1990, 1996) combine bibliometrics and DEA to study the regional effects of research in the building sector and university productivity in Sweden. Simulation may use a number of techniques. It may also be used for preliminary testing before applying other methods of analysis. Combining methods is more costly than selecting only one. However, with the enormous decrease in per unit computing costs, it has become and will probably continue to be less costly than in the past.

CONCLUSION

This paper has discussed the evaluation of policy-oriented social science research by reference to literature in other fields. A number of methods to analyze the benefits of social science research have been offered and discussed. Most of these methods are adopted from other literatures such as those in policy studies and on the evaluation of programs, science and R&D. Many are well-known research methodologies that are used in research in the social sciences and evaluation in other fields. This paper advises that replication should be conducted as part of the evaluation process before reaching the question of whether the research is beneficial (although before they conduct any analysis, evaluators and their clients will have some prior belief about whether the research has produced any benefits). None of the methods presented as potential tools for conducting evaluations completely meets our theoretical needs. A mix of methods may often prove satisfactory. However, the link between research and the application of results is likely to remain tenuous. This means that an investigation of the benefits of social science research will probably remain an exercise in “muddling through.” Decisionmakers will find it necessary to select among the methodologies available for evaluation, or have the evaluators propose methodologies that they believe offer the best ways to evaluate social science research in particular cases.

NOTES

1. As Alston, Norton, and Pardey (1995, 29) frame the problem in a treatise on agricultural research: "...there are long, variable, and uncertain lags in the interval between commencing a research activity and generating useful knowledge...."

2. See Alston, Norton, and Pardey (1995) and Stephan (1996).

3. High-profile research that has great financial impact upon interest groups is eventually replicated and is often replicated numerous times. For example, see Card and Krueger (1995, 238–243) on minimum wage studies.

4. It is worth pointing out that replication is built into some research. The use of multiple case studies or the use of robust regression methods in conjunction with ordinary least squares are examples of built-in replication. Of course, this research is generally more costly than if replication is not built in.

5. See Cozzens (1995), Bozeman and Melkers (1993), Popper (1995), and Wagner (1995) for more on bibliometrics.

6. The surveyed population should not include members of the body that implements the research, but should focus on its intended beneficiaries.

7. The marginal benefit curve is downward sloping as marginal benefits decline while the marginal cost curve is upward sloping as marginal costs increase.

8. See Merewitz and Sosnick (1971) for an account of how discount rates may be manipulated to reflect the views of those doing the evaluation.

9. Mansfield (1996) and Hall (1996) discuss the uses and pitfalls of this method in greater detail.

10. That is, if multicollinearity does not present an insurmountable obstacle.

REFERENCES

Alston, J. W., G. W. Norton, and P. G. Pardey. 1995. *Science under scarcity: Principles and practice for agricultural research evaluation and priority setting.* Ithaca, NY: Cornell University Press.

Arthur, W. B. 1988. Self-reinforcing mechanisms in economics. In *The economy as an evolving complex system: SFI studies in the sciences of complexity*. Redwood City, Calif.: Addison-Wesley Publishing Company. Also in *The economy as an evolving complex system*, eds. P. W. Anderson, K. J. Arrow, and D. Pines. Redwood City, CA: Addison Wesley Publishing Company.

Begg, C. B., and J. A. Berlin. 1988. Publication bias: A problem in interpreting medical data. *Journal of the Royal Statistical Society* 15 (3, Series A): 419–463.

Boskin, M. J., and L. J. Lau. 1996. Contributions of R&D to economic growth. In *Technology, R&D, and the economy*, eds. L. R. B. Smith and C. E. Barfield. Washington, D.C.: The Brookings Institution and American Enterprise Institute.

Bozeman, B. 1993. Peer review and evaluation of R&D impacts. In *Evaluating R&D impacts: Methods and practice*, eds. B. Bozeman and J. Melkers. Boston, MA: Kluwer Academic Publishers.

Bozeman, B., and J. Melkers, eds. 1993. *Evaluating R&D impacts: Methods and practice.* Boston, MA: Kluwer Academic Publishers.

Bretschneider, S. 1993. Operations research contributions to evaluation of R&D projects. In *Evaluating R&D impacts: Methods and practice*, eds. B. Bozeman and J. Melkers. Boston, MA: Kluwer Academic Publishers.

Card, A., and A. B. Krueger. 1985. Time series minimum wage studies: A meta-analysis. *American Economic Review* 8 (May): 238–243.

Caudle, S. L. 1994. Using qualitative approaches. In *Handbook of practical program evaluation*, eds. J. S. Wholey, H. P. Hatrey, and K. E. Newcomer. San Francisco, CA: Jossey-Bass Publishers.

Charnes, A., W. W. Cooper, and R. Ferguson. 1955. Optimal estimation of executive compensation by linear programming. *Management Science* 1: 138–151.

Charnes, A., W. W. Cooper, Y. Lewin, and L. M. Seiford, eds. 1994. *Data envelopment analysis: Theory, methodology, and application*. Boston, MA: Kluwer Academic Publishers.

Cozzens, S. E. 1995. *Assessment of fundamental science programs in the context of the Government Performance and Results Act (GPRA).* Rand Domestic Research Division Report. Santa Monica, CA: Rand Corporation.

Cragg, J. G. 1990. Monte Carlo methods. In *The new Palgrave: Time series and statistics*, eds. J. Eatwell, M. Milgate, and P. Newman. New York: W. W. Norton and Company.

Davis, R. T. 1955. *Some suggestions for writing a business case.* Boston, MA: HBS Case Services, Harvard Business School.

Denton, F. 1985. Data mining as an industry. *Review of Economics and Statistics* 57: 124–127.

Eppen, G. D., F. J. Gould, and C. Schmidt. 1988. *Quantitative concepts for management: Decision making without algorithms*. Englewood Cliffs, NJ: Prentice Hall.

Epstein, J. M., and R. L. Axtell. 1996. *Growing artificial societies*. Cambridge, MA: The MIT Press.

Feigenbaum, S., and D. M. Levy. 1990. Testing the replication hypothesis when the data set is subject to gross error. *Economics Letters* 34: 49–53.

________. 1993. The market for (ir)reproducible economics. *Social Epistemology* 7: 215–232.

________. 1994. Is science self-enforcing? The reputational cost of scientific error. Paper presented to the Society for the Social Studies of Science.

________. 1996. The technological obsolescence of scientific fraud. In *Rationality and Society.*

Garfield, E. 1963. Citation indexes in sociological and historical research. *American Documentary* 14 (4): 89–102.

Gilli, M., ed. 1996. *Computational economic systems: Models, methods, and econometrics*. Boston: Kluwer.

Hall, B. H. 1996. The private and social returns to research and development. In *Technology, R&D, and the economy*, eds. L. R. B. Smith and C. E. Barfield. Washington, D.C.: The Brookings Institution and American Enterprise Institute.

Kee, J. E. 1994. Benefit-cost analysis in program evaluation. In *Handbook of practical program evaluation*, eds. J. S. Wholey, H. P. Hatrey, and K. E. Newcomer. San Francisco, CA: Jossey-Bass Publishers.

Kingsley, G. 1993. The use of case studies in R&D impact evaluations. In *Evaluating R&D impacts: Methods and practice*, eds. B. Bozeman and J. Melkers. Boston, MA: Kluwer Academic Publishers.

Levy, D. M., and N. E. Terleckyj. 1995. Appreciating technology: A test for endogenous growth. Paper presented to the American Economic Association, January.

Lynn, L. E., ed. 1978. *Knowledge and policy: The uncertain connection*. Washington, D.C.: The National Academy of Sciences.

Mansfield, E. 1996. Contributions of new technology to policy. In *Technology, R&D, and the economy*, eds. L. R. B. Smith and C. E. Barfield. Washington, D.C.: The Brookings Institution and American Enterprise Institute.

Melkers, J. 1993. Bibliometrics as a tool for analysis of R&D impacts. In *Evaluating R&D impacts: Methods and practice*, eds. B. Bozeman and J. Melkers. Boston, MA: Kluwer Academic Publishers.

Merewitz, L., and S. H. Sosnick. 1971. *The budget's new clothes*. Chicago, IL: Markham Publishing Co.

Muraskin, L. D. 1993. Understanding evaluation: The better way to prevention programs. U.S. Department of Education, Washington, D.C.

Nelson, R. R., and S. G. Winter. 1982. *An evolutionary theory of economic change.* Cambridge, MA: The Belknap Press.

Newcomer, K. E., H. P. Hatrey, and J. S. Wholey. 1994. Meeting the need for practical evaluation approaches: An introduction. In *Handbook of practical program* *evaluation*, eds. J. S. Wholey, H. P. Hatrey, and K. E. Newcomer. San Francisco, CA: Jossey-Bass Publishers.

Price, D. J. de Sola. 1961. *Science since Babylon.* New Haven: Yale University Press.

Popper, S. W. 1995. *Economic approaches to measuring the performance and benefits of fundamental science.* Rand Domestic Research Division Report, Santa Monica, California.

Quade, E. S., and G. M. Carter. 1989. *Analysis for public decisions*. New York: North-Holland.

Romero, C. 1991. *Handbook of critical issues in goal programming*. Oxford: Pergamon Press.

Sarafoglou, N., and K. E. Haynes. 1996. University productivity in Sweden: A demonstration and explanatory analysis for economics and business programs. *The Annals of Regional Science* 30: 285–304.

_________. 1990. Regional efficiencies of building sector research in Sweden: An introduction. *Computers, Environment, and Urban Systems* 14: 117–132.

Stephan, P. E. 1996. The economics of science. *Journal of Economic Literature* 34 (September): 1199–1235.

Stiglitz, J. E. 1988. *Economics of the public sector*. New York, NY: W.W. Norton and Company.

Stough, R. R., and P. Rietveld. 1996. Institutional issues in transportation systems. Paper presented at the NSF/ESF meeting on transportation research, Strasbourg, France, 6–7 October.

Wagner, C., and A. Flanagan. 1995. Workshop on the metrics of fundamental science: A summary. Rand Office of Science and Technology Project Memoranda, Santa Monica, California.

Wagner, C. S. 1995. Techniques and methods for assessing the international standing of U.S. science. Rand Domestic Research Division Report. Santa Monica, CA: Rand Corporation.

Weiss, C. H. 1972. *Evaluation research*. Englewood Cliffs, NJ: Prentice-Hall.

Wholey, J. S., H. P. Hatrey, and K. E. Newcomer, eds. 1994. *Handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass Publishers.

Wilson, R. H. 1993. *States and the economy*. Westport, CT: Praeger Publishers.

Yin, R. K. 1989. *Case study research: Design and methods*. Newbury Park, CA: Sage Publications.

________. 1993. *Applications of case study research*. Newbury Park, CA: Sage Publications.

Henry E. Kilpatrick, Jr. is professor at the Institute of Public Policy, George Mason University, Fairfax, Virginia.