Introduction

Evaluation is a powerful tool for policy and decision making, but only if it is correctly structured, managed, and applied. At the technical level, this means having a clear set of objectives, a logical framework, and valid methods and findings. At the program level, this means having knowledgeable evaluators and program administrators who can work with evaluators to structure relevant questions, relate general findings to specific agency settings, and communicate effectively with diverse audiences.

Evaluation involves methodological science and craft and organizational art. Evaluation involves the selection and implementation of systematic, valid, and appropriate methodologies. Evaluation also involves the organizational establishment, management, deployment, and dissemination of a portfolio of studies and associated findings that provide defensible and relevant information to decision makers.

Since at least the 1960s, program evaluation and its close companion, policy analysis, have become institutionalized aspects of congressional oversight and agency management of federal programs.[1] Institutionalization, however, is not synonymous with acceptance, quality, credibility, or impact.

Many federal and state agencies struggle to implement evaluation programs, which may be mandated but may, in fact, be unwelcome appendages to program operations and/or organizational decision making. Evaluation efforts are some-

[1]Joseph S. Wholey, John W. Scanlon, Hugh G. Fukumoto, and Leona M. Vogt, *Federal Evaluation Policy, Analyzing the Effects of Public Programs* (Washington, DC: The Urban Institute, 1970); A. Melt2sner, Policy Analysis in the Bureaucracy (Berkeley, CA: University of California Press, 1976).

times undertaken only when required by outside forces, such as legislation. These realities make the evaluation program of one of the nation's public-private partnership programs, the Advanced Technology Program (ATP), all the more striking. The design and implementation of its evaluation program offers a creative laboratory for learning more about evaluation, particularly in the field of science and technology.

This report addresses the science, craft, and art of evaluation in the context of ATP, a program within the U.S. Department of Commerce, National Institute of Standards and Technology (NIST). In terms of science and craft, this report describes the evolving set of methodological techniques ATP has used to monitor and assess programmatic impacts, and reports methodological and empirical advances generated by ATP. The report assembles a large body of past work into a coherent framework, making it more accessible to a diverse audience.

In terms of art, the report describes the creation and evolution of ATP's evaluation program. It describes how the program has used evaluation techniques to answer questions directed at its fundamental rationale, design features, and economic impacts.

This report also highlights ways in which early challenges to the program led to an evaluation program noted for its methodological variety, recourse to nationally prominent scholars, and a distinctive emphasis on disseminating its findings in peer-reviewed literature and in policy-relevant presentations for agency officials and political constituencies. Finally, the report shows that in politically contested arenas, methodological rigor and empirically grounded findings are necessary but not sufficient to protect or advance a program. For even as ATP drew national recognition in the evaluation community for its systematic, and rigorous evaluation program, congressional critics of the program repeatedly charged that the program lacked adequate evidence of its impacts.

The report is presented in three major parts, plus this introductory section. The remainder of the introduction presents the political underpinnings of ATP and discusses further the role of evaluation. Part I provides a general framework for evaluation, and in the context of the framework, discusses evaluation

fundamentals and methods, best practices, and ATP's evaluation program. Part II demonstrates the use of evaluation methods by citing, both through direct quotation and paraphrase, the contents of a selection of ATP evaluation studies. Part III presents the emerging body of knowledge from the studies of ATP over the past decade—knowledge about evaluation, firm behavior, and ATP's performance. Part III also presents conclusions and recommendations, summarizing key points and identifying remaining questions, issues, obstacles, and challenges, and promising opportunities to learn more through evaluation.

The Political Economy of the Advanced Technology Program

ATP is an outgrowth of a national policy dialogue directed at redressing systematic gaps in the market settings that link scientific advances to technological innovation. ATP was designed to fill those gaps by a project selection and funding process that fosters and enhances new and intensified modes of collaboration among private, not-for-profit, and public sector organizations engaged in high-risk research, all with the objective of accelerating the development and commercialization by U.S. firms of enabling technologies.

ATP was established in 1988 under Title V (Technology Competitiveness Act), Subtitle B of the Omnibus Trade and Competitiveness Act (P.L. 100–418), and received its first appropriations in fiscal year 1990. The program is a response both to a specific period in United States international economic competitiveness and to a longer-term historic perspective of the federal government's contribution to the development and commercialization of potentially significant technological advances of a broadly enabling nature.

The specific historic backdrop for ATP's creation was a pervasive concern in the United States throughout much of the 1970s and 1980s, documented by a variety of key economic indicators, that the nation's slow rate of economic growth and worsening international trade balance were attributable in part to its loss of technological competitiveness. The U.S. innovation system was widely seen as exhibiting structural flaws. Points of concern included a loss of international and domestic markets in technology-intensive products, a failure to gain

market position for products U.S. firms had helped research and develop, and a faltering standing in R&D "races" to exploit the commercial significance of emerging scientific advances.[2]

The flaws were seen as products of an emphasis in U.S. policies on scientific leadership, mission-directed R&D, and breakthrough discoveries, paired with a lack of attention to technology development and deployment. Critics believed an overemphasis in these areas was made at the expense of diffusion-oriented strategies and programs that would assist U.S. firms and laboratories to gain technologically and economically when their scientific advances were converted into new and improved products and processes.[3] Critics argued that the United States needed an "innovation policy" rather than a "science-only policy."

These flaws were described in both absolute and relative terms. In an absolute sense, the flaws were seen as reflecting specific forms of market failure, primarily the lack of adequate incentives for private firms and private capital markets to fund "high-risk, high-return research on broadly enabling, precompetitive technologies for which appropriability was at issue."[4] In a relative sense, the challenge to the United States was that major economic competitors, most notably Japan and some members of the European community, were held to be far more willing to use national funds to support the development and commercialization of civilian technologies, particularly those seen as providing "first mover" advantages in new, strategic, or large commercial markets.[5] Lack of

[2]U.S. Congress, Office of Technology Assessment, *Making Things Better: Competing in Manufacturing,* OTA–ITE–443 (Washington, DC: U.S. Government Printing Office, 1990).

[3]H. Ergas, "The Importance of Technology Policy." In P. Dasgupta and P. Stoneman, eds., *Economic Policy and Technological Performance* (New York: Cambridge University Press, 1987), pp. 51–96; R. Florida and M. Kenney, *The Breakthrough Illusion* (New York: Basic Books, 1990).

[4]L. Branscomb and G. Parker, "Funding Civilian and Dual-Use Industrial Technology." In L. Branscomb, ed., *Empowering Technology* (Cambridge, MA: MIT Press,1993), pp. 64–102.

[5]U.S. Congress, Office of Technology Assessment, *Competing Economies: America, Europe, and the Pacific Rim,* OTA–ITE–498 (Washington, DC: U.S. Government Printing Office, 1991).

comparable government programs in the United States was held to place U.S. firms at a disadvantage, because they had to bear the full burden of supporting costly and risky R&D projects with questionable appropriability of profits.

The design of ATP also drew upon emerging perspectives about the character of competition among firms. Knowledgeable observers no longer saw firms that conducted business in the same product lines as engaged exclusively in Darwinian struggles for survival, and they no longer identified firms engaged in buyer-seller relationships only as seeking to maximize profits/minimize costs from one-time transactions. Instead, both theory and increased documentation of business practice pointed to numerous forms of collaboration between such pairs.[6]

In particular, with respect to R&D, it became increasingly evident that firms sharing a common interest in selected "generic" technologies could increase their competitiveness through collaboration. IBM's agreement in 1992 to join with Toshiba of Japan and Siemens of Germany to develop memory chips illustrates what was seen as a new way of doing business, especially in R&D-intensive and technology-driven sectors.

Collaboration spread the capital costs of large-scale R&D projects among investors. It offered a way to explore otherwise out-of-reach technological frontiers, both as part of a firm's offensive strategy of finding new products and markets, and as part of a defensive strategy of better predicting and understanding the rise of disruptive technologies that threatened a firm's corebusinesses.

The growing acceptance of inter-firm and inter-sector collaboration, especially in pre-competitive, generic (or enabling) R&D, was reflected in several legislative changes, such as the National Cooperative Research Act of 1984,[7] that relaxed antitrust bars to collaborative R&D programs. The evolving framework also saw a relaxation of the analytical and ideological categories that had previously been

[6]D. Mowery, ed., *International Collaborative Ventures in U.S. Manufacturing* (Cambridge, MA: Ballinger, 1988).

[7]A. Link and L. Bauer, *Cooperative Research in U.S. Manufacturing* (Lexington, MA: Lexington Books, 1989).

used to define the boundaries of federal government and private sector roles in supporting R&D directed at civilian-oriented technology.[8]

Under the framework that dominated policy thinking following World War II, the federal government had responsibility for funding basic research and R&D related to mission-oriented national objectives such as defense, while the private sector had responsibility for R&D directed at civilian-oriented products and processes. Reinforcing the hold of this paradigm was the checkered political and technological history of efforts by the federal government to promote specific technological innovations.[9, 10] Over time, however, the paradigm began to yield to a more complex but nuanced appreciation of both the overlap and holes in these seemingly fixed boundary markers.[11] In particular, the emerging model emphasized government support of technically risky projects that had prospects for technological and commercial importance; significant amounts of industry cost sharing; sunset requirements for government funding; selection of projects for funding in a fair and open competitive process free of political influence; large spillover effects; and collaboration with other firms and organizations.

Despite growing support for public-private partnerships emphasizing high-risk, enabling technology, opposition continued.[12] Opposition embodied several

---

[8]B. Smith, *American Science Policy Since World War II* (Washington, DC: Brookings Institution, 1990).

[9]See L. Cohen and R. Noll, *The Technological Pork Barrel* (Washington, DC: Brookings Institution, 1991) for examples of inefficient government funding projects and a discussion of factors behind the inefficiencies.

[10]See Charles W. Wessner, ed., *The Advanced Technology Program: Challenges and Opportunities* (Washington, DC: National Academy Press, 1999), pp. 1–4, for a summary description of the instrumental role played by the federal government since its earliest history in the development of new production techniques and technologies. Examples include a government contract in 1798 with Eli Whitney to help lay the foundation for the U.S. machine tool industry, government funding to demonstrate the feasibility of Samuel Morse's telegraph, and government support for the development of the radio.

[11]National Academy of Sciences, *The Government Role in Civilian Technology* (Washington, DC: National Academy Press, 1992).

[12]C. Hill, "The Advanced Technology Program: Opportunities for Advancement." In L. Branscomb and J. Keller, eds., *Investing in Innovation* (Cambridge, MA: MIT Press, 1998), pp. 153–173; P. Hallacher, "Effects of Policy Subsystem Structure on Policymaking: The Case of the Advanced Technology Program and the Manufacturing Extension Partnership Program," Ph.D. Dissertation, Pennsylvania State University, 2000.

reinforcing ideological and theoretical positions.[13] It started with the premise that the national government contributes most effectively to long-term economic growth and technological innovation when it adopts a minimalist approach to involvement in the economy, focusing its activities on the enforcement of private contracts, the provision of stable monetary aggregates, the maintenance of certain and low taxes, and the provision of a small, carefully delimited set of selected public goods.[14] Opposition was also based on propositions that public sector support cannot increase the rate of commercially productive R&D because of the phenomena of moral hazard, adverse selection, rent-seeking behaviors by firms applying for support, and bureaucratic entrepreneurship on the part of agency managers pushing for overly rapid development of untested and economically problematic technologies.[15]

Opponents of federal domestic technology development programs also point to further defects. First, they believe the programs may simply substitute public-sector R&D dollars for private-sector R&D dollars. Second, they say these programs may have untoward distributive effects, unfairly benefiting some U.S. firms and/or industries at the expense of others, such as large firms at the expense

[13]Opposition extended to earlier, contemporary, and subsequent similar programs that involved efforts by the federal government to stimulate acceleration of commercially oriented technological innovation through targeted selection of industries, firms, technologies, or projects. See B. Smith, *American Science Policy Since World War II,* 1990.

[14]Federal government support of basic research rests on a broad-based, bipartisan political consensus that it is a public good that would not be adequately provided by the actions of the private sector alone, and that it contributes in significant ways to national objectives in defense, health, space, and economic competitiveness.

[15]"Moral hazard" refers to the actions by individuals or firms to increase their risk taking behavior in response to the existence of insurance or other forms of compensation for costly outcomes. Thus, while ATP purposefully underwrites a portion of an R&D project's risk to encourage private firms to take on more challenging—hence riskier—technical problems, the phenomenon of moral hazard raises the question of whether or not there is a side effect of promoting inefficient resource use. "Adverse selection" refers to a problem arising from asymmetry in the quality of information possessed by the applicant firms and ATP about the riskiness of prospective R&D projects. If ATP knows less about the risks than the companies, it will be at a disadvantage in its selection decisions. "Rent-seeking behaviors" refers to efforts directed at acquiring or making permanent a stream of payments, government program, or regulatory arrangement that yields returns to an economic group above that which they would receive under competitive market conditions. Such behavior would occur if awardees attempted continuation of their government awards beyond the scheduled end date or if political efforts were taken to maintain the program even if it were deemed not to produce its intended results.

of small firms. Additionally, opponents claim the programs may respond to pressures from bureaucratic and technological interests, and unduly push for rapid deployment of technologies before the technological feasibility or market demand has been demonstrated. They maintain that these programs may fail to select proposals with the highest likelihood of achieving technical success, commercial success, and large spillovers. They also maintain that funding a project creates vested interests that causes its funding to become self-perpetuating. Finally, opponents say the claims about the productivity of collaborative efforts, and the federal government's role in fostering them may be challenged as unproven.

The catalog of arguments on behalf of the benefits of public-private partnerships in general, and ATP in particular as well as the catalog of possible flaws are both mixes of normative perspectives on the role of the public sector, and theoretically and empirically testable propositions. The competing catalogs, in effect, describe domains of debate and decision making in which evaluation may contribute to improved public policy.

The Role of Evaluation

Evaluation has a recognized and well-understood role in the operations of most public sector agencies. As phrased by Mark, Henry, and Julnes:[16]

> Evaluation assists sense making about policies and programs through the conduct of systematic inquiry that describes and explains the policies' and programs' operations, effects, justifications, and social implications. The ultimate goal of evaluation is social betterment, to which evaluation can contribute by assisting democratic institutions to better select, oversee, improve, and make sense of social programs and policies. (p. 3)

Evaluation has served all these purposes for ATP: to oversee, improve, and make sense of the program—both within and outside of NIST. ATP's evaluations have helped program administrators revise and refine the program so that the program is harmonized with legislative intent. Evaluation has been used to respond to

[16]M. Mark, G. Henry, and G. Julnes, *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs* (San Francisco: Jossey-Bass, 2000).

Congressional and OMB questions about program characteristics and impacts. ATP's evaluation program has had a broader span of coverage than typically encountered across federal agencies. Its studies have ranged from monitoring and surveying the characteristics of early grantees to sophisticated theoretical and econometric efforts designed to tease out differences between the performance of ATP awardees and non-awardees. In pursuit of these multiple objectives, ATP has strategically employed an evolving and broad set of methodologies.

Early ATP evaluation activities focused on measuring progress, generating information about awardees, and projecting economic impact. In the context of political opposition to ATP, the program's evaluation agenda has often been shaped in response to ongoing challenges, such as whether ATP awards were being used by firms to displace private sector funds. Over time, as some funded projects have had time to progress through technical research and commercialization stages, ATP's evaluation program has centered more on measuring impacts. These include both private impacts garnered directly by the firms participating in ATP projects, and spillover impacts that are one of the justifications for the program.

Finally, evaluation has served as a test of two of ATP's core and linked premises; that is, that a federal agency program can strategically and selectively generate additional and innovative modes of R&D collaboration between and among organizations, and that this collaboration among organizations will accelerate the rate of technological innovation in the nation's economy. ATP's evaluations have provided new insights into the characteristics of workable collaborations among various sectors, the character of gaps between scientific discovery and product development, the strategies and behaviors of firms as they seek to close or bridge these gaps, and the forms and networks through which knowledge spillovers diffuse within an economy.