NIST GCR 06–891

ADVANCED TECHNOLOGY PROGRAM

Bridging from Project Case Study to Portfolio Analysis in a Public R&D Program

A Framework for Evaluation and Introduction to a Composite Performance Rating System

Rosalie Ruegg

April 2006

NIST National Institute of Standards and Technology • Technology Administration • U.S. Department of Commerce

About ATP's Economic Assessment Office

The Advanced Technology Program (ATP) is a partnership between government and private industry to conduct high-risk research to develop enabling technologies that promise significant commercial payoffs and widespread benefits for the economy.

Since the inception of ATP in 1990, ATP's Economic Assessment Office (EAO) has performed rigorous and multifaceted evaluations to assess the impact of the program and estimate the returns to the taxpayer. To evaluate whether the program is meeting its stated objectives, EAO employs statistical analyses and other methodological approaches to measure program effectiveness in terms of:

- Inputs (program funding and staffing necessary to carry out the ATP mission)
- Outputs (research outputs from ATP supported projects)
- Outcomes (innovation in products, processes, and services from ATP supported projects)
- Impacts (long term impacts on U.S. industry, society, and economy)

Key features of ATP's evaluation program include:

- Business Reporting System, a unique online survey of ATP project participants, that gathers regular data on indicators of business progress and future economic impact of ATP projects.
- Special Surveys, including the Survey of Applicants and the Survey of Joint Ventures.
- Status Reports, mini case studies that assess ATP projects on several years after project completion, and rate projects on a scale of zero to four stars to represent a range of project outcomes.
- Benefit-cost analysis studies, which identify and quantify the private, public, and social returns and benefits from ATP projects
- Economic and policy studies that assess the role and impact of the program in the U.S. innovation system

EAO measures against ATP's mission. The findings from ATP surveys and reports demonstrate that ATP is meeting its mission:

- Nine out of 10 organizations indicate that ATP funding accelerated their R&D cycle.
- The existence of a "Halo Effect." As revealed by EAO surveys, shows that an ATP award establishes or enhances the expected value in the eyes of potential investors.
- ATP stresses the importance of partnerships and collaborations in its projects. About 85 percent of project participants had collaborated with others in research on their ATP projects.

Contact ATP's Economic Assessment Office for more information:

- On the Internet: www.atp.nist.gov/eao/eao_main.htm
- By e-mail: atp-eao@nist.gov
- By phone: 301-975-8978, Stephanie Shipp, Director, Economic Assessment Office, Advanced Technology Program
- By writing: Economic Assessment Office, Advanced Technology Program, National Institute of Standards and Technology, 100 Bureau Drive, Stop 4710, Gaithersburg, MD 20899-4710

NIST GCR 06–891

Bridging from Project Case Study to Portfolio Analysis in a Public R&D Program

A Framework for Evaluation and Introduction to a Composite Performance Rating System

Prepared for
Economic Assessment Office
Advanced Technology Program
National Institute of Standards and Technology
Gaithersburg, MD 20899

By
Rosalie Ruegg
Managing Director
TIA Consulting, Inc.
ruegg@ec.rr.com

Contract SB1341–01–W–1276

April 2006

U.S. DEPARTMENT OF COMMERCE
Carlos M. Gutierrez, Secretary

TECHNOLOGY ADMINISTRATION
Robert Cresanti, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
William Jeffrey, Director

ADVANCED TECHNOLOGY PROGRAM
Marc G. Stanley, Director

Abstract

This paper presents a framework, rooted in the case study method, for evaluating both individual projects and a portfolio of projects. It introduces a prototype evaluation tool that offers new capabilities for the evaluation of a public research and development program in the intermediate period before long-run effects can be measured. An assembly of steps produces linked layers of information in a framework that results in a synergistic evaluative capability that enables bridging from project case study to portfolio analysis. This interlinked framework for evaluation can be used both by program administrators, taking a top-down approach, and by project managers, taking a bottom-up approach. It also produces results of interest to policy makers and other program stakeholders. The prototype evaluation tool, the Composite Performance Rating System (CPRS), is designed specifically for the Advanced Technology Program (ATP), but is potentially adaptable to other programs. CPRS uses uniformly collected output and outcome data (indicator metrics) to compute an overall performance rating for each of ATP's completed projects, using a four-star system. The distribution of ratings provides an easy-to-grasp measure of the overall portfolio's performance. This report, organized in two parts, presents in Part I the framework for evaluating individual projects and the entire portfolio, and in Part II a detailed account of CPRS developed specifically for ATP. Part I is expected to be of general interest to the broader evaluation community; Part II, of primary interest to ATP and to other multi-goal, public research and development program administrators interested in new ways of using indicator data for portfolio evaluation.

Key Words: Advanced Technology Program; case study; composite performance rating system; CPRS; indicator metrics; performance metrics; program evaluation; public-private partnership program.

Contents

Figures

Tables

Acknowledgements

I acknowledge the support of ATP and its Economic Assessment Office (EAO) for making this work possible. I particularly wish to thank Stephanie Shipp, EAO Director, and Connie Chang, EAO senior economist and project coordinator, for their oversight, helpful guidance, and many useful comments throughout the process. Aaron Kirtley, former EAO staff member, Lee Bowes, EAO economist, and Dennis Leber, a statistician at NIST, provided helpful comments through their review of the final draft. I also wish to express my gratitude for their thoughtful and in-depth reviews of an earlier draft manuscript to Harris Liebergot, former program manager in ATP's Information Technology and Electronics Office, and Susannah Schiller, leader of ATP's Information Resources Group. Elissa Sobolewski, Director, ATP Information Technology and Electronics Office; Lorel Wisniewski, ATP Deputy Director; and Brian Belanger, former ATP Deputy Director (and currently consultant to ATP), reviewed and provided additional comments on the final report.

I am also indebted to two groups who attended pre-report presentations on the approach. Those who attended presentations at NIST on the composite performance rating system provided valuable insight through their questions, and gave direct assistance through their helpful suggestions. Those who attended a presentation on the framework, given at the Expert Lecture Series of the American Evaluation Association Annual Conference, November 8, 2002, Crystal City, Virginia, included staff from other government agencies in the United States and abroad who provided useful insight through their questions and comments.

Executive Summary

Improved methods are needed for program evaluation. Program administrators of public research and development programs like ATP, with long lead times to payoff, multiple pathways to impact, and multiple program goals, need more effective tools to manage their portfolios of projects. The challenging evaluation needs of ATP have led it to engage in state-of-the-art application of existing evaluation methods, and to support the development of new and emerging methods as it has developed a comprehensive evaluation program. This report presents in two parts an evaluation methodology developed in the context of ATP's broader evaluation program. Part I presents a framework rooted in the case study approach that makes it possible to bridge from the individual project focus typical of case study to an assessment of a program's entire portfolio of projects. Part II presents in detail an emerging evaluation tool—the Composite Performance Rating System (CPRS)—embedded within the project-to-portfolio approach and critical to accomplishing portfolio analysis.

The project-to-portfolio approach of Part I begins with the definition of the portfolio for which the approach will be applied. In the case of ATP, the defined portfolio consists of completed projects for which approximately three to four years have elapsed since ATP funding ended. For projects in the defined portfolio, case studies are conducted, using uniform data collection to build a database of output and outcome data selected to indicate progress toward achieving program goals and referred to alternatively as "indicator metrics."

CPRS is constructed to combine selected indicator metrics in order to show overall progress in three dimensions: (1) adding to the nation's scientific and technical knowledge base, (2) disseminating the knowledge, and (3) commercializing new and improved products and processes from the technology developed. The distribution of projects by their CPRS ratings characterizes the overall performance of the portfolio against program goals. The broader framework of which CPRS is a part also provides aggregate statistics on individual indicator metrics and an estimate of minimum net benefits for the portfolio or for the program.

Carrying out the project-to-portfolio approach produces interlinked levels of information that can serve as a versatile evaluation resource. Program administrators can take a top-down approach, starting with the portfolio performance distribution and tracing down to the individual project case studies for additional amplification. Project managers, on the other hand, can take a bottom-up approach, starting with the case-study details of projects assigned to them, and tracing up to the portfolio level to find out how their projects or technology areas are performing relative to the broader portfolio.

Part II of the report moves from this framework of evaluation to focus in detail on the CPRS tool, which is embedded in the framework. CPRS was developed to allow program administrators to respond to a question of some urgency: How are projects in ATP's portfolio performing overall against ATP's mission-driven multiple goals in the intermediate period after project completion and before long-term benefits are realized?

CPRS assigns 0 to 4 stars to each completed project based on overall performance across multiple program objectives, with the weakest performers assigned 0 stars, and the strongest, 4 stars. CPRS allows the portfolio of projects to be characterized in terms of the resulting distribution. For example, application of CPRS to the first 50 completed ATP projects generated the following distribution of projects by performance: 16%, 4 stars; 26%, 3 stars; 34%, 2 stars; and 24%, 1 or 0 stars. This performance distribution was considered in line with expectations of program administrators for the high-risk research projects selected by ATP. Indeed, it is a tenet of the program that some degree of failure must be expected if projects are tackling difficult-to-solve technical challenges.

The CPRS star ratings are based on uniformly compiled output and outcome data (indicator metrics), plus an assessment of the outlook for further developments drawn from case studies conducted for all completed ATP projects several years after ATP funding ends. In the intermediate period of focus, evidence that (1) a project added to the nation's scientific and technical knowledge base, (2) the knowledge created is being disseminated to others, and (3) the innovators and their collaborators are moving toward commercializing new and improved products and processes from the project's technology, and the outlook for continuing commercialization efforts is positive is taken as indicative that the project is continuing on the path toward delivering potential economic benefits.

CPRS as depicted in Part II has been developed specifically for ATP on an exploratory basis. CPRS represents an initial baseline evaluative tool for assessing overall project and portfolio performance during the intermediate period of an ATP-funded project's life cycle. CPRS would have to be reformulated to be applicable to other public R&D programs with a different mission, goals, and time horizon than ATP, or to rate the performance of projects substantially earlier or later in their life cycles.

A search of the literature revealed counterpart development of composite rating systems in a variety of applications. They shared the characteristic of an ad hoc development stemming from the absence of existing theory to direct the selection of variables and the specification of weights used in their construction. Similarly, the formulation of CPRS is necessarily ad hoc in nature—an empirically based, rather than a theoretically based, method. The prototype CPRS formulation was constrained by the kinds of indicator data available for ATP's first 50 completed projects, and influenced by the range of values for those first 50. The use of indicator data means that the CPRS scores signal intensity of progress toward goals but do not reveal estimates of economic benefits. Projects with similar CPRS ratings may differ in their ultimate long-term net economic benefits for this reason and also because project performance may change after the case-study data are collected—for better or worse.

Limitations notwithstanding, CPRS provides an easy-to-grasp management and communications tool capable of highlighting overall portfolio performance. The star ratings show which projects exhibit strong and which exhibit weak or moderate outward signs of progress toward program goals during the intermediate period of the project's life cycle. ATP has used CPRS to brief ATP oversight and advisory bodies, public policy analysts, evaluation groups, the broader scientific and technical community, and the general public about ATP's performance.

The CPRS was developed in 2000–2001 using the first 50 ATP projects completed in conjunction with the writing of status reports (mini-case studies) on these projects. Since then, ATP's Economic Assessment Office has computed CPRS ratings and published over 100 additional status reports. All completed status reports and CPRS ratings can be accessed on a searchable website (http://statusreports.atp.nist.gov/) and in the following publications:

- Performance of 50 Completed ATP Projects: Status Report — Number 2,[1]
  NIST SP 950–2, 2001
- Performance of 50 Completed ATP Projects: Status Report — Number 3,
  NIST SP 950–3, 2006

1. Status Report — Number 2 contains all projects from Status Report — Number 1 (38 projects), as well as the additional 12 projects used in the CPRS formulation.

Introduction

Administrators of public programs need evaluation tools that can help them manage their programs effectively and provide answers to stakeholders. A number of evaluation tools exist that can guide management decisions, help assess program performance, and provide answers to a variety of questions. However, available tools are not adequate to meet all the requirements asked of them. Advances in evaluation that increase the understanding of public programs, help determine what is working and what is not, analyze why, and measure effectiveness are important for management, oversight, and public support.

Measuring the impact of the Advanced Technology Program (ATP)[2] is challenging to evaluators. ATP is a complex program with multiple mission-driven goals. ATP partners with industry to develop enabling, early-stage technologies with the potential for delivering broad-based economic benefits to the United States. As of January 2006 ATP has 768 projects in its portfolio, with multiple pathways of impact—some direct and some indirect—multiple recipients of benefits, and multiple dimensions of success, all played out over a number of years.

Existing evaluation methods are able to answer many questions about ATP's performance and impact reasonably well. Surveys, descriptive statistics, economic case studies, econometrics and sociometrics, bibliometrics, and peer reviews are all evaluation methods and tools applied by ATP to answer stakeholder questions and to manage the program. Evaluation challenges such as measuring spillover effects, capturing portfolio performance in the intermediate period, and other requirements led ATP to also support the extension of existing methods, the development of new and emerging methods, and the compilation of databases to advance evaluative capabilities. The fact that ATP operated under a local climate of experimentation in trying out new technical approaches, and the focus of its parent organization, NIST, on measurement, contributed to a favorable environment for advancing program evaluation methodology, as well as applying it. Now in the middle of its second decade of operation, ATP has strategically deployed a comprehensive evaluation program, and the framework and method presented here are encompassed within that larger evaluation program.[3]

2. ATP is operated by the National Institute of Standards and Technology, within the Technology Administration of the U.S. Department of Commerce. ATP was authorized by the Omnibus Trade and Competitiveness Act of 1988 (P.L. 100–418), as amended by the American Technology Preeminence Act of 1991 (P.L. 102–245).

3. For an account of ATP's evaluation program over its first decade, its use of existing methods, support of developing new methods, and principal study findings, see Ruegg and Feller (2003).

This report is composed of two parts. Part I presents an eight-step framework, rooted in case study, that substantially extends the evaluative capability of the case study method. The framework provides interlinked, multiple layers of information useful to a variety of program stakeholders, including program administrators, project managers, policy makers, and others. Part I provides an account of how and why the framework was developed within the context of ATP, describes its components, and explains how it can be used to extend understanding of a public research and development (R&D) program. Part I contains Sections 1–3 of the report.

Part II details the development of one component of the eight-step approachæa composite performance rating system (CPRS), which is a key link in bridging from a case study of a project to an analysis of a portfolio of projects. It describes the formulation of CPRS for use by ATP, based on indicator metrics from ATP's first 50 completed projects. It also shows the results of sensitivity analyses, reports the results of applying the CPRS method to ATP, and discusses its strengths and limitations. Part II contains Sections 4 through 9 of the report.

Part I

A Framework Linking Evaluation of Individual Projects to Assessment of a Portfolio of Projects

Traditionally, project case study and program portfolio analysis have been pursued as separate lines of evaluative inquiry. Project case studies are commissioned by nearly all public R&D programs. They offer rich, detailed information about a project or cluster of projects, but limited information about overall program performance. In contrast, evaluative methods aimed at portfolio analysis, such as survey, focus on a whole body of projects, but, by necessity, eliminate individual project details in favor of overall measures of a program's performance.

Program administrators, policy makers, and other stakeholders typically need information at both the project and the portfolio levels to understand a program, manage it, and report on it. They need both the details and the overview. Ideally, they would be able to traverse smoothly between the project, or bottom level, and the portfolio, or top level of their programs. But, heretofore, a tight linkage between the bottom and top levels has not been provided in the evaluation of public R&D programs.

For public R&D programs with new project starts each year, multiple program goals, multiple years of funding per project, multiple beneficiaries from project success, and multiple paths of impact, years must elapse before it is possible to measure retrospectively long-term impacts for a portfolio of projects. To help inform program developments during the intermediate period—after government funding of research has ended, but before long-term impacts have been realized—ATP adopted several evaluative strategies during the 1990s. One strategy was to conduct prospective case studies, which attempted to broaden the analysis from single projects to clusters of projects.[4] Another strategy was to conduct surveys that

4. The case studies conducted by ATP during the 1990s were single project studies with one exception. Research Triangle Institute conducted a cluster study of the first seven tissue engineering projects funded by ATP. See Sheila Martin et al., (1998). In its second decade, ATP has supported case study of several additional technology clusters, including component-based software projects and advanced composite projects.

included questions about accomplishments, plans, and expectations. A related strategy was to aggregate selected individual output and outcome measures that relate to program goals and use them as "indicator metrics" to signal project and program progress.[5]

Partly through forethought, and partly through a series of evolutionary steps, ATP developed a framework linking evaluation of individual projects to assessing a portfolio of projects. First, ATP saw the advantage of conducting descriptive case studies systematically for a broadly defined population of its projects. The approach called for flexibility in telling each project's story, and consistency in data collection across all the projects. Uniformity in data collection provided the opportunity to report select aggregate output and outcome data for the body of case studies. Parallel pursuit of detailed prospective economic case study of some of the projects in the portfolio provided an opportunity to estimate minimum net benefits for ATP's portfolio of projects. Finally, the need for a consolidated performance measure for portfolio assessment prompted the development of CPRS, a new evaluation tool. Interlinked, the steps associated with this framework of evaluation produces a synergistic capability, well beyond the value of the steps taken piecemeal.

1.0 Developing the Project-to-Portfolio Approach

The framework builds from a case-study base to portfolio analysis, with linkages from bottom to top. This section first shows how the various layers of information are linked in this approach. Then it lists and discusses the steps for constructing the framework. Next, it assesses the evaluative outputs from applying the approach.

1.1 Multi-Tiered Analytical Capability

With reference to Figure 1, the bottom tier contains the foundational project case studies and details of all the projects in the portfolio. These case studies document the what, how, and why of completed projects. In addition to this information, for a select sub-group of the projects, detailed economic benefit estimates are provided, principally through prospective economic case study.

The second tier from the bottom of Figure 1 indicates the body of uniformly collected output, outcome, and outlook data for each project and, for a subset of projects, more exten-

5. ATP, like other agencies, included select indicator metrics in its performance reports under the Government Performance and Results Act of 1993 (GPRA), which requires federal agencies to develop strategic plans, relate budgetary requests to specific outcome goals, measure performance, and report on the degree to which goals are met. An overview of the GPRA is provided in Appendix 1 of the U.S. General Accounting Office's *Executive Guide, Effectively Implementing the Government Performance and Results Act* (GAO 1996).

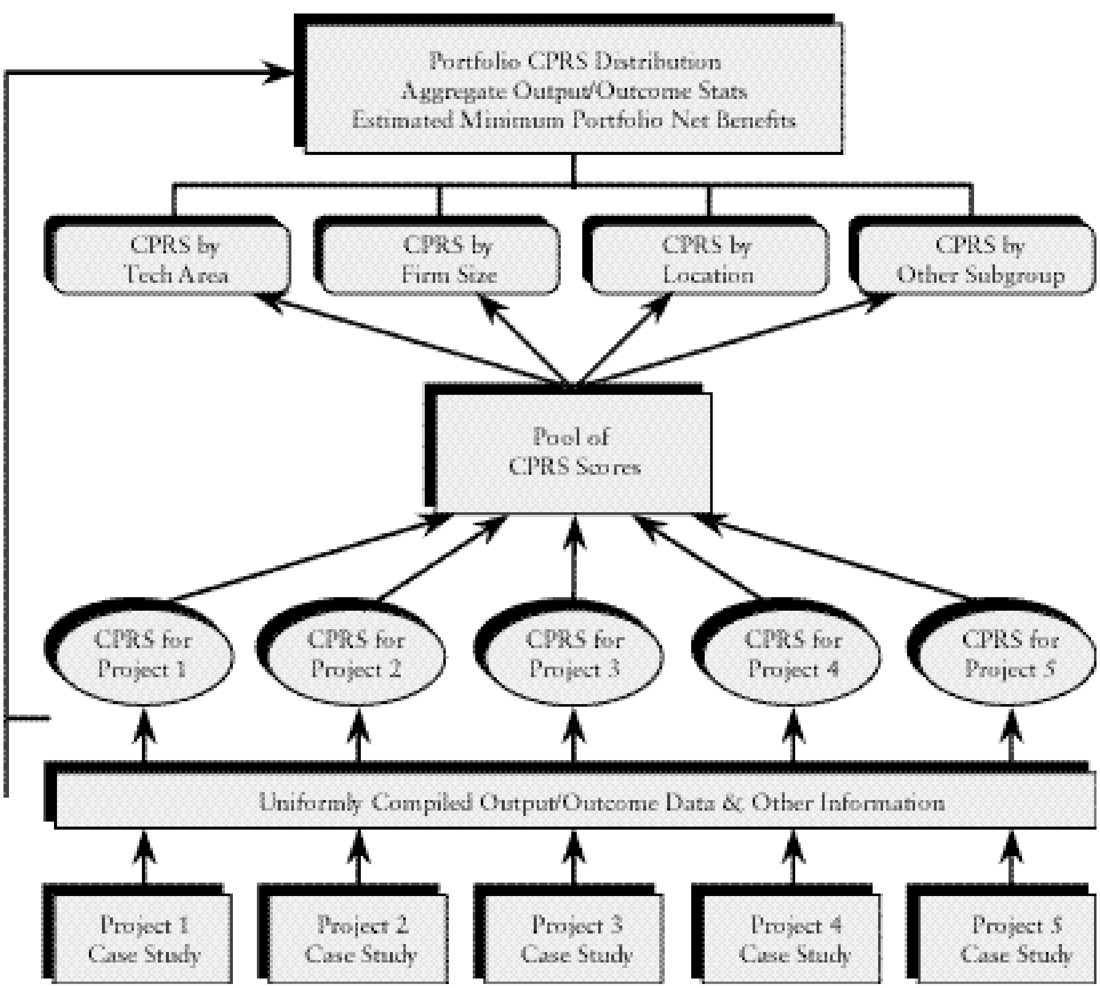
sive economic estimates of impact. These data are tied to the individual project, but can be aggregated by data category.

The third tier from the bottom of Figure 1 indicates the CPRS ratings. There is a star rating for each project, computed from weighted values of its output/outcome/outlook data.

The individual CPRS scores form a pool of CPRS scores in the fourth tier from the bottom of Figure 1. The results can be broken out by technology area, firm size, firm location, or other areas of interest as indicated in the fifth tier from the bottom of Figure 1.

The sixth, or top, tier contains portfolio-level information. It shows the overall distribution of CPRS ratings for all projects in the completed project portfolio. It also shows an estimate of minimum net benefits for the program, computed by aggregating benefits from the

Figure 1. Multi-Tiered Analytical Capability



sub-group of detailed economic case studies, and subtracting total program costs from the sub-group's total benefits. The top tier also provides output and outcome data from all the case studies, aggregated by data type, such as numbers of patents or commercial products.

The resulting analytical capacity under this framework allows either a top-down or bottom-up approach. Program administrators generally start with the top levelæthe distribution of performance ratings and estimated minimum net benefits of the programæand move down the tiers to the case studies for more detailed information. In contrast, project managers generally start with the projects they are assigned to oversee, and go up the chain to determine how their projects are performing relative to the overall program portfolio. In either case, the analytical capability of this approach facilitates managerial action to improve project and portfolio performance, and provides performance metrics to stakeholders. This ability to move up and down the various tiers makes this framework of analysis potentially useful both to higher-level program administrators and to project managers.

1.2 Steps Involved in Developing the Framework

The analytical capability depicted in Figure 1 can be developed by carrying out the steps listed in Table1. The first step consists of several preparatory tasks:

- defining the program's major goals;
- defining the portfolio of projects for which case studies will be conducted;
- identifying the projects to receive more detailed economic case studies;
- identifying the targeted time within the projects' life cycle for conducting the case studies; and
- gaining at least a preliminary idea of the output and outcome metrics that appear best to indicate progress during the specified time period against the identified program goals and that are feasible for uniform collection.

The second step is to formulate CPRS, reexamining the indicator metrics to be used, and deciding the weights to assign them and how to combine them to signal a project's overall progress toward achieving the program's multiple goals.

The third step is to prepare a data collection template, as well as a broader case-study template, for use by case-study analysts. A template is particularly important to ensure the consistent collection of information when there are multiple case-study analysts. Consistency in the data collection is critical to implementing CPRS and satisfying other study objectives. The detailed economic case studies may also benefit from a study template or guide, but it is important to afford the analysts sufficient latitude to respond to the unique modeling requirements of each economic case study.

TABLE 1

Eight-Step Process in Developing Project-to-Portfolio Evaluation Framework

Step Number	Step Description
1	Perform preparatory analysis: identify program goals, project populations, study timing, and relevant output/outcome metrics.
2	Formulate CPRS including indicator metrics to be used and weights to be assigned for project performance scoring.
3	Prepare templates for use by case-study analysts/writers.
4	Conduct case studies, including detailed economic estimates for a select project sub-group.
5	Apply CPRS using the compiled project data.
6	Compute and display the distribution of scores for the portfolio and sub-groups of interest.
7	Compute aggregate output/outcome statistics.
8	Compute aggregate project benefits attributed to ATP for the detailed economic case studies and compare total project benefits against total program costs (either for the portfolio or for the entire program).

The fourth step is to conduct case studies for the defined population using the template. To keep evaluation cost down and to allow the cases to be carried out faster, the bulk of the case studies can be short, descriptive "mini-cases." To enable the quantitative estimation of minimum program net benefits, it will be necessary to conduct detailed economic assessments for a few projects in the portfolio.[6]

6. Specification of the selection plan for the detailed economic case studies is beyond the scope of this paper. In the example used in this paper, the cases were selected from among what were expected to be relatively strong projects—an effective way of estimating minimum net portfolio benefits, given the distribution of projects by performance.

The fifth step is to apply the rating system to the projects in the portfolio using the compiled indicator metrics and the CPRS formulation as carried out in step 2. Each project then carries a star rating together with the project story and key data.

The sixth step is to calculate and depict in tabular and graphic form the distribution of project ratings for the portfolio. Distributions of performance ratings grouped by variables of special interest, such as technology area or company size, also can be calculated and depicted.

The seventh step is to compute aggregate indicator metrics for the portfolio. These metrics might include, for example, the number of publications, the number of collaborations formed, and the number of products commercialized.

The eighth step is to estimate minimum net benefits attributed to the public program. This rough estimate is derived by aggregating benefits attributed to the program for the sub-group of projects for which detailed economic case studies have been conducted and subtracting total program costs from the result. Because the portfolio examined may encompass less than the entire program portfolio, and detailed economic benefit estimates are likely only available for a fraction of the projects, the resulting estimate of program net benefits is considered a minimum value.

1.3 Resulting Evaluative Products

Carrying out the eight steps listed in Table 1 produces a number of distinct products that have potential value for project and program evaluation. As summarized in Table 2, these evaluation products are:

- a detailed descriptive and explanatory case study account of each project in the portfolio;
- a set of detailed economic cases that provide estimates of project benefits;
- aggregate statistics for the various indicator metrics;
- a CPRS rating for each project;
- the distribution of projects in the portfolio by their CPRS ratings; and
- an estimate of minimum net benefits or net losses for the entire program.

The six products are individually useful for managing the program and reporting its results to stakeholders. The linkages among the products increase their potency and usefulness as management tools.

TABLE 2
Six Products From Applying the Eight-Step Process

Number	Product Description
1	A set of unique project case studies.
2	Prospective estimates of economic benefits from a sample of projects.
3	Databases of aggregate statistics.
4	A CPRS rating for each project.
5	Distributions of projects and groups of projects by CPRS ratings.
6	Estimated minimum net benefits for the entire program.

2.0 Pros and Cons of Rooting the Approach in Case Study

ATP, like most other public R&D programs, made the case study method—both descriptive and economic—one of the mainstays of its evaluation program from the outset.[7] There are a number of reasons for the popularity of the case study method—particularly with programs that fund scientific research.[8] Table 3 summarizes both the advantages and disadvantages of the case study method.

7. A recent benchmarking workshop among five S&T programs in the U.S., and several S&T programs in other countries revealed that they all used case study as one of their main evaluation methodologies. See Ruegg (2003). Program administrators from the U.S. National Science Foundation, National Institutes of Health, the Department of Energy's Office of Science and Office of Energy Efficiency and Renewable Energy, the National Institute of Standards and Technology's Advanced Technology Program, Finland's Tekes Program, Canada's Industrial Research Assistance Program, and Israel's MAGNET Program all noted the importance of case study methodology to their evaluation efforts.

8. Branch et al., (2001).

2.1 Case Study Advantages

Cast in a narrative format, descriptive case studies can make complex science and technology (S&T) projects accessible and interesting to a non-scientist audience. The potential scope of the case study format is flexible and broad, ranging from brief descriptive summaries to long, detailed accounts. Using a "storytelling" approach, the evaluator may present the genesis of ideas, explore what happened and why, give an account of the human side of a project, explain goals, explore project dynamics, investigate particular phenomena, and present outcomes in their complexity without being subject to the confines inherent in most other evaluation methods. The freedom to collect multiple kinds of information makes the case study method useful for exploring ideas and constructing theories about program or project dynamics.

Pushing beyond descriptive case studies, economic case studies can combine the storytelling richness of the descriptive case study with the analytical rigor of economic analysis. Economic benefit-cost case studies typically combine qualitative and quantitative results, reflecting the difficulty of translating all important effects into monetary units.

2.2 Case Study Disadvantages

Despite its advantages, the case study method is traditionally considered to have several major shortcomings as an evaluation tool. Descriptive case studies are qualitative and anecdotal. Economic case studies, though highly quantitative, usually are successful in capturing only partial benefits in monetary units and these are usually predicated on a number of assumptions. And, whether qualitative or quantitative, case studies typically pertain to single projects, or, at best, small clusters of projects, such that their results usually cannot be generalized to the entire portfolio of projects.

2.3 Case Study, A Serious Research Tool?

Noting that distinguished scholars frequently use case study as a method of analysis, Yin, in his landmark book on case studies, asks, "If the case study method has serious weaknesses, why do investigators continue to use it?"[9] Among the possible explanations Yin considers are that people are not trained in the use of other methods, or, for federally sponsored research, that the difficult clearance procedures required for surveys and questionnaires have made their use "a bureaucratically hazardous affair," leading researchers to favor the case study method, which is relatively unencumbered by restrictions and requirements. Yin, however, rejects these explanations for the popularity of the case study methodology. As evidence, he points out that expert analysts skilled in the use of a variety of evaluation methods use the case study method

9. Yin (1994).

TABLE 3

Advantages and Disadvantages of Case Study Methodology

Advantages

- Easy to understand and remember.
- Richness of detail may be useful in formulating theories and hypotheses and for understanding how and why questions.
- Good for documenting experiences, providing a benchmark for other projects, and capturing a holistic view.
- An economic case study relates project outcomes to inputs and provides quantitative measures in the language of finance.

Disadvantages

- Descriptive case study generally provides anecdotal evidence, which lacks robustness as evidence of a program's effectiveness.
- The focus is on the individual project, not on the program's portfolio of projects.
- Results for single projects and small clusters of projects generally cannot be generalized.
- Important benefits in an economic case study may be difficult or impossible to capture in monetary terms.

and that "federally sponsored research does not dominate the social sciences—certainly not in Europe and other countries" where the case study method is also widely practiced.

Identifying three types of case studiesædescriptive, exploratory, and explanatory—Yin argues that the case study method qualifies as a serious research tool. He states, "In general, case studies are the preferred strategy when 'how' or 'why' questions are being posed, when the investigator has little control over events, and when the focus is on a contemporary phenomenon within some real-life context."[10]

10. Yin (1994), p. 1.

Many public R&D programs, such as ATP, have numerous applications that meet these conditions. In the case of ATP, case study investigators typically are in a third-party status with no control over the contemporary developments that occur in real-life project contexts. In the case of the mini case studies, the investigators have the task of finding out and documenting how the research projects turned out and why; and why the firms turned to the government for funding. Although the survey method could be alternatively used to collect data on what happened, it is less suited for capturing data to address how and why questions.[11] Moreover, surveys, unlike case studies, do not capture well complex human and organizational elements of the projects, important to understanding project dynamics. In Yin's words, "... case study allows an investigation to retain the holistic and meaningful characteristics of real-life events...."[12] He continued, " 'How' and 'why' questions are likely to favor the use of case studies, experiments, or histories."[13]

3.0 ATP's Development and Use of the Project-to-Portfolio Approach

In 1996, ATP and NIST managers requested ATP's Economic Assessment Office[14] to produce streamlined case studies for all completed projects, written for a lay audience, with consistent inclusion of specific information. This request was driven primarily by stakeholder requests for more information on the performance of ATP's portfolio of projects after ATP's period of funding for the projects has ended.

Prior to the development of these case studies for all completed projects, brief project descriptions prepared and made public at the time the ATP awards were announced were the only information provided consistently to the public on every project funded. Survey data provided information for all or most projects, but the data were released only in the aggregate in order to protect proprietary project or company information. Workshops and conferences provided the public with information on many projects, but not in a written and consistent format that was widely available, and generally not for projects in the post-completion period.

11. ATP used the case study method to capture all the output data for the first 50 completed projects, but now is positioned to use a combination of surveys and case studies to collect the output or "what" data. The case study method continues to be used to capture the 'how' and 'why' of project developments.

12. Yin (1994), p. 3.

13. Yin (1994), p. 3.

14. ATP's Economic Assessment Office (EAO) is comprised of a staff of economists, statisticians, and information specialists that organizes and manages evaluation activities for ATP, conducts studies in-house, and commissions studies with universities, consulting firms, other assessment organizations, and with individual economists, in part through the National Bureau of Economic Research (NBER).

Economic case studies focused on a single or small cluster of projects; econometric case studies used specialized databases in ways that did not reveal in-depth information about individual projects.

The new case study product was added to a rich array of existing evaluation methods that provided substantial valuable information about ATP's performance. Results from surveys, for example, showed what percentage of the project portfolio had published, patented, sold products, collaborated, attracted funding from other sources, experienced employment growth, and other interim achievements that were evidence that ATP was meeting its mission objectives. Econometric studies showed, among other things, that ATP was selecting projects with larger than average spillover potential, providing a halo effect as award-recipient firms were more successful than non-recipient firms in attracting additional private sector funding, and having a positive effect on the rate of patenting by award recipients. Detailed benefit-cost studies provided estimates of returns on investment for selected projects. Descriptive case studies for selected projects provided details about those projects and explanations about ATP's role. A variety of analytical and conceptual modeling studies advanced fundamental understanding of the program and its operations. Indicator data signaled that projects were making progress in a number of areas. The new case study product provided additional capability to ATP's evaluation program.

3.1 Mini Case Studies Using Uniform Data Collection for All Completed Projects

ATP's Economic Assessment Office responded to the request for a new evaluation product by commissioning descriptive mini case studies with uniform data collection for all completed ATP projects, written approximately 3 to 5 years after project completion. There were several reasons for setting the cases 3 to 5 years after project completion. First, because of ATP's relatively short history; few projects had progressed far enough at the time the mini case study effort was begun. Second, fixing the case studies 3 to 5 years after project completion had the advantages of yielding a manageable number of projects to get the effort underway, allowing adequate time for companies to make post-project progress, and avoiding problems associated with attempting to obtain detailed information long after a project has been completed.

The mini case studies told the project story, describing what was done, how and by whom the project goals were accomplished, the role of ATP, and assessed the outlook for further developments. The mini case studies focused on the original award recipients, but branched out to include licensees and takeover companies if these organizations provided the focus of current activity. In a few instances and to a limited extent, the mini case studies followed the mobility of people.

Analysis of ATP's mission and program goals, and specifically requested information by NIST and ATP administrators, guided the identification of data to be uniformly collected for all the completed projects. The specific data collected are listed and discussed further below under the section, "Defining Mission-Driven Goals for CPRS Development."

3.2 Publication of ATP's First Status Report

There were 38 projects in the first group for which mini case studies were developed. These projects were collectively published in a report. [15] The project story was given in 4–5 pages of text and key data were highlighted for each mini case study. An overview chapter reported on aggregate output and outcome data for the portfolio of completed projects, and provided an estimate of minimum net benefits from the program. ATP called these mini case studies, "Status Reports," to underscore the fact that they provided a snapshot view of project accomplishments at a specified time, and that subsequent developments were likely to occur.

The first volume of Status Reports proved useful, but it did not provide program administrators a composite measure of portfolio performance. Policymakers and other stakeholders were interested in the overall performance of the program and ATP managers were interested in what aspects of the portfolio were performing well and which ones were falling short. With the program's multiple goals and multiple output and outcome measures, it was difficult to get a clear reading of performance across the portfolio, even from the aggregate indicator metrics. Some projects scored high in one outcome area; and low in another. Program administrators were unable to interpret the results in terms of overall portfolio performance. A pressing question of policy makers could not be answered using the evaluation tools at hand, "How are projects in ATP's portfolio performing overall against ATP's mission-driven objectives in the intermediate period after project completion and before long-term benefits are realized?" And more specifically, "What percentage of the portfolio is made up of strong performers, weak performers, and those showing a moderate level of performance?" Thus, the first volume of Status Reports, while valuable from the standpoint of reporting on the post-ATP performance of individual projects, could not be used as a management tool for addressing portfolio performance.

3.3 New Features in ATP's Second Status Report

To address portfolio performance, an easy-to-grasp overall performance measure for each project needed to be added to the richness of the individual project details so as to offer a means of comparison across the completed projects examined.

In 2001 a replacement volume containing status reports for the first 50 completed projects was published.[16] Like the first volume, it contained an overview chapter with aggregate indicator metrics and estimated minimum net benefits from the program's investment to date. In addition to the earlier volume, it contained a new feature: Composite Performance Rating System (CPRS) ratings. Each project was assigned a 0 to 4 star rating, where a low star rating indi-

15. Long (1999).

16. ATP (2001). This volume contained the first 38 status reports plus 12 new ones, in addition to the CPRS (Composite Performance Rating System) ratings.

cated weak performance, and a high star rating indicated strong performance against ATP mission objectives. The overview chapter showed the distribution of projects by their composite performance. An appendix listed the 50 projects in terms of their individual CPRS ratings. Another extension provided by the second volume was inclusion of patent citation trees to illustrate visually knowledge spillovers from the project.

The latest batch of 50 mini case studies for completed projects is now available. A data template is being used to ensure that data for the new cases are consistent with the first 50 projects. The newly assessed projects carry CPRS ratings, and the distribution of projects by performance has been computed for the larger group. The entire portfolio of ATP Status Reports of completed projects are available on ATP's website with a searchable feature that facilitates retrieval of status reports by state, technology area, CPRS rating, size of company (for project lead), and other characteristics of the mini case studies.[17]

17. The searchable website is available at www.atp.nist.gov, or specifically at http://statusreports.atp.nist.gov/

Part II

Development and Application of the Composite Performance Rating System (CPRS)

Part II focuses on the Composite Performance Rating System, or CPRS, embedded in the framework described in Part I of the report. CPRS is an evaluative tool for portfolio management. It is still in the prototype stage of development. As related in Part I, the impetus for its development came from the need for an effective way to characterize the interim, overall performance of a large portfolio of projects aimed at achieving multiple program goals. None of the methods used at the time showed how projects in ATP's portfolio were performing overall in the intermediate period after project completion and before long-term benefits are realized against ATP's mission-driven goals.

As an emerging evaluation method, CPRS stands on the shoulders of existing methods and uses their outputs as its inputs. It is not a stand-alone approach. Specifically, development of CPRS followed a sequence of related advances in ATP's case study methodology outlined in Part I.

The intended audiences for this part of the report are limited and specific. The primary audience is ATP. This report documents CPRS computation and hence serves as a user guide. Evaluators who follow new methods may also be interested as well as administrators of other public R&D programs who wrestle with problems similar to those of ATP—that is, difficulties in clearly characterizing the interim performance of their project portfolios. Policy makers and other program stakeholders who use evaluation results from a variety of methods may also find the details of the method useful as a reference resource. A goal is to make the method clear for users and to provide sufficient background to allow for critical assessment of the method.

Part II is organized into six sections. The next section, "Prior Practice Using Composite Ratings," discusses prior practice with composite ratings for performance evaluation. The section, "CPRS Development: First Steps," covers several preparatory steps to the CPRS development, namely the general formulation, defining the relevant goals, and reviewing available data. The section, "Specifying Indicator Variables for Use in the CPRS Formulation," explains the selection of indicator variables from the available data to correspond to each of three defined program goals. The section, "Applying Weighting Algorithms to the Selected Indicator Variables," presents the weighting algorithms for each indicator variable used to calculate the performance ratings. It illustrates the composite ratings of ATP's first 50 projects. The section, "Critique of CPRS" provides a critique of the method, pointing out data limitations and methodological issues. And, finally, "Summary and Conclusions" provides a summary and conclusions for both parts of the report.

4.0 Prior Practice Using Composite Ratings

While the CPRS represents an emerging method for analyzing portfolio performance of a public R&D program, the use of composite ratings in evaluation and program management is not new. Though not an exhaustive treatment, this section explores prior development and application of composite rating systems to establish the state of prior practice.

4.1 Composite Scoring by U.S. Education Department to Assess if Participants Meet Regulatory Requirements

The first example is a composite scoring system used by a federal government agency to meet legislative requirements. Developed by Bearing Point (formerly KPMG Consulting Inc.) for use by the U.S. Department of Education to meet responsibility under the Higher Education Act of 1992, the method constructs a composite score from financial ratios to test the financial strength of institutions participating in Title IV programs.[18]

The method calls for the user to calculate three customized financial ratios; assign "strength factors" to the ratios to place all ratio results on a common scale so they can be combined; multiply the strength factors by weights that reflect their relative importance; and sum the resulting products to form a composite, single-number score of the institution's overall financial health.

Application of the composite scores places all the institutions in one of four categories of financial performance. Institutions that have a composite score above a regulatory standard established by the Education Department are considered to meet the test of financial responsibility.

The rationale for using a composite score was that the institutions participating in Title IV programs vary in their operating size, mission, ownership structure, and operating environment, and yet may have similar overall financial health. The developer of the rating system

18. The approach is described in KPMG (1997). The report is found online at http://www.ed.gov/offices/OPE/PPI/finanrep.html.

explained that the composite score allows a host of factors to be taken into account in computing the overall financial-health rating critical to decisions made by government program officials.

4.2 Composite Scoring Proposed to Improve Healthcare Performance in OECD Countries

A second example is a composite scoring system proposed to improve the performance of health systems in OECD countries.[19] The rationale for composite indicators is that performance in healthcare is multi-dimensional, a "rounded assessment of performance" is needed, and composite indicators are needed "to make comparisons systematic."

The general form given for the composite indicators is the following:

$$C = \_{1}P_{1} + \_{2}P_{2} + \ldots + \_{n}P_{n},$$

where $\_{1}$ indicates the value attached to an extra unit of indicator P1.

For Canadian regions, it is noted that there are 15 indicators of performance organized in six categories, and that the indicators are combined using weights based on expert judgment.

4.3 Composite Scoring of Hospital Performance

A third example is a composite performance-rating tool for use by hospitals to reduce the number of patients who die each year from preventable medical errors. The tool was developed by the First Consulting Group, a provider of information-based consulting, integration, and management services in the life sciences, in collaboration with the Leapfrog Group, a consortium of more than 90 Fortune 500 companies and other large purchasers of healthcare for employees. Funding support was provided by the California HealthCare Foundation and the Robert Wood Johnson Foundation.[20]

A focus of the composite scoring for hospital performance was the Computer Physician Order Entry (CPOE). The assessment calls for a randomly generated set of physician orders for patients to be downloaded to cover each of the different order categories being evaluated. The orders are entered into the CPOE rating system along with information on the corresponding patients. The results are evaluated, and scores are generated against a weighting scheme.

The aim of the CPOE rating system is to intercept orders most likely to cause harm to patients. Each order drawn in the sample is assigned two scores. One score indicates the likely

19. See Smith (2001).

20. Kilbridge et al. (2001). The report may be viewed at www.fcg.com by clicking on the report cover.

severity of adverse reaction if the prescribed medication reaches the patient, based on "commonly used rankings" as cited in the literature. The other score indicates the frequency of the adverse event, based on opinions of expert advisers (several people who are named in the study). It is noted that expert opinion is used in the absence of a definitive literature. The hospital receives feedback on the details of the scores, and a composite score is used for public reporting on individual hospital performance.

4.4 Composite Scoring for Investor Stock Ratings

Other examples that construct composite performance scores from indicator data can also be found in the field of investment. One such example is the Quadrix® Stock-Rating System, a proprietary stock screening tool that provides composite rankings for over 5,000 stocks through the Dow Securities Review Service.

The rating system was developed by Richard Moroney, editor of Dow Theory Forecasts, a financial newsletter providing investing tools, services, and investment advice to subscribers.[21] The Quadrix uses more than 100 variables to score stocks in seven categories. The resulting percentile rating is used to compare a company's stock performance against that of industry peers.

4.5 Composite Scoring of the Entire S&T Innovation Process

The fourth example, called the Metric of Process Outcomes, is of particular interest because it is drawn from the field of science and technology (S&T) program evaluation.[22] Though it resembles the CPRS approach in several ways, it has critical differences. It is more sweeping and ambitious in its intended coverage, which is the entire innovation process. More importantly, it does not address the question of interim portfolio performance that the CPRS was designed to address. Though limited in real-world applications, the Metric of Process Outcomes has been used to compare the performance of two laboratories.

The approach, developed by Professor Eliezer Geisler of the Stuart Graduate School of Business, Illinois Institute of Technology, groups measures of output from S&T into four categories: (1) immediate outputs, (2) intermediate outputs, (3) preultimate outputs, and (4) ultimate outputs. It then assigns weights to the various outputs and computes an index for each category of outputs. It next computes an overall index (macroindex or "Omega Factor") for all categories of outputs.

21. More about the Dow Theory Forecast and the Quadrix Stock-Rating System can be found at the newsletter Website, www.dowtheory.com.

22. Geisler (2000), pp. 257–262.

The Process-Outcomes approach is intended to act as a substitute for traditional benefit-costs analysis. Geisler, the developer, aims to capture in index form the "total impacts of S&T on the economy and on society."[23]

As is the case with the CPRS, the weights to be assigned to the indicators must be determined, and it is fairly clear from the model description that there is no underlying theory which guides this step. According to Geisler:

> Clearly, the weights applied in building the indices and the macroindex are the result of our analysis of relative importance (and other factors) in each stage. In some ways this depends on the viewpoint of the evaluator. If the emphasis is on the downstream stages, then the preultimate and ultimate outputs will be awarded higher levels of importance and higher weights. However, the contributions of earlier stages must never be totally excluded from the macroindex.[24]

While they share certain features, the CPRS and the Process-Outcomes model are quite different in their construction and use. One major difference is the CPRS's organization of indicators around mission-driven goals at one process stage, as compared with the Process-Outcomes' organization of indicators by multiple process stages. Hence, the Process-Outcomes model is described as facilitating cross-agency comparisons, while the CPRS is aimed at comparisons of project classes within a single program or agency.

Another contrast of the two methods is their different focus and units of analysis. The CPRS builds from the individual project level to the portfolio level, and aims at signaling the interim performance of a portfolio of projects. The Process-Outcomes model is directed at an agency's overall S&T investment and aims at providing an overall measure of an agency's ultimate S&T impacts.

There is not a direct correspondence between the two in terms of the data used. The CPRS uses some of the indicators and measures drawn from both the first and second stages of the Process-Outcomes model, but not all of them, and it uses some categories of data not included in the Process-Outcomes model.

The Process-Outcomes model has the conceptual advantage of providing a framework that spans the entire innovation continuum, but in its selection of indicators and specification of weights, the approach remains ad hoc. Furthermore, implementation of the model would be extremely difficult and require many years of data collection. In any case, it does not provide an answer to the primary question posed here, "How are projects in ATP's portfolio

---

23. Geisler (2000), p. 257.

24. Geisler (2000), p. 258.

performing overall in the intermediate period against ATP's mission-driven, multiple goals?" In contrast to the Process-Outcomes model, the CPRS has the practical advantage of a more narrow focus that can be supported with empirical data and used for within-portfolio comparisons to answer the question of interest.

4.6 Conclusions From Reviewing Prior Practice

The five examples of prior practice summarized in Section 4 were drawn from recent, diverse applications in the fields of education, health, finance, and science. They were variously commissioned by public officials and private companies, and were carried out by analysts/ consultants in private companies and by academics. In each case, the problem addressed was important, the need for a better management tool was the driving force for development of the composite rating, the rating tool developed was specific to the given circumstances, and the approach taken seemed credible for those circumstances. These examples establish precedence for using a composite rating system to consolidate multiple kinds of information so that decision makers can more easily grasp the overall effect.

The examples demonstrate considerable care and thought in the development of their composite ratings. Nevertheless, there is a pervasive ad hoc or improvisational aspect to each case. The examples point to a common characteristic: the lack of existing theory to direct the selection of variables and formulation of weights in composite rating systems.

In the absence of existing theory, the developers in each of the illustrative cases followed an empirically based approach to structuring the composite ratings, drawing on existing data when available to assign weights to the variables used, and, in most cases, relying heavily on expert judgment for the selection of variables and the assignment of weights.

5.0 CPRS Development: First Steps

This section presents the preliminary steps to developing a composite rating system for ATP. It first presents the general formulation separately from the detailed formation because the general formulation has broad applicability to multi-goal programs while the later-presented detailed formation is specific to ATP. Second, it discusses the definition of mission-driven goals against which progress is assessed—another generic, early step which must be done regardless of the program for which a composite rating system is constructed. Third, this section discusses data constraints that applied to the construction of ATP's CPRS, and which may apply to other programs. It sets the stage for the detailed specification of ATP's CPRS that follows in Section 6.

5.1 CPRS General Formulation

In its most general form, potentially applicable to any multi-goal program, the CPRS is formulated as follows:

$$CPRS = \sum_{i=1}^{K} \sum_{j=1}^{N} (\;)\_{ij} A$$

where

CPRS = Composite Performance Rating System,

$_{i}$ = the ith of N indicators of progress toward achieving the jth of K mission-driven goals,

$_{ji}$ = the weighting factor applied to $_{i}$ indicator of progress,

N = the number of progress indicators for a given mission-driven goal, counting from i = 1 to N.

K = the number of mission-driven goals for which there are progress indicators, counting from j = 1 to K.

A = an adjustment factor for converting the total raw score to a 0–4 point scale.

Thus, for each program goal, a set of indicator variables are selected, each of them is weighted, the weighted values are summed, the process is repeated for each of the subsequent goals, and then the aggregated values for each goal are summed, and an adjustment factor is applied to convert the composite raw score to a 0–4 point scale used to assign 0 to 4 stars.

Challenges are to define appropriate program goals for the relevant evaluation period, identify what should be—and feasibly can be—the indicators of progress for each goal, and decide how much each indicator should count, i.e., what weighting factor should be applied to each indicator variable, in deriving the composite rating.

5.2 Defining Mission-Driven Goals for CPRS Development

The purpose of the CPRS is to provide a management and evaluation tool for the intermediate period after ATP funding ends and before long-term national benefits have had time to appear. Table 4 shows in three columns ATP's goals in relationship to its time horizon. It illustrates the period of focus.[25] In the left-hand column, the table shows the short-term period during

25. The table suggests well-defined stages of project/technology development, but in practice there are overlaps and non-linearities. A patent may be filed early in a research project, a spin-off product opportunity may occur early; opportunities for research publications may occur at various times during research and technology development, sometimes overlapping commercialization activities. The table is intended only to suggest tendencies for certain activities to concentrate during certain periods.

TABLE 4
ATP Goals at Different Stages of the Project Life Cycle

Short-term ATP Goals	Intermediate-term ATP Goals	Long-term ATP Goals
• Collaborative formations	Progress toward:	Broad-based social benefits
• Proposal development	• Knowledge creation	• Private returns
• Accelerated and leveraged research	• Knowledge dissemination	• Spillover returns
• Accomplishment of research tasks	• Commercialization	

which the principal activity is research, and the principal short-term program goals are to foster collaboration, stimulate proposal development, accelerate the awarded research, and accomplish the research tasks of funded projects. This period, during which ATP funding occurs, ranges from one to five years for each project, and lasts an average of about three years across all projects. In the right-hand column, the table shows the period during which the long-term goals of the program are to be realized for successful projects. In this period—approximately 10–12 years after the project ends—ATP's ultimate goals may be stated in terms of "stimulating prosperity through innovation" and "broad-based benefits for the nation." The long-term goals are to achieve widely distributed social benefits, comprised of both private returns and spillover returns. Neither of these periods is the focus of this CPRS evaluation tool. Rather, it is the intermediate or interim period that is of central focus for the CPRS.

During the interim period, the program's funding is completed and work is no longer closely monitored by ATP project managers, who have newly funded projects assigned and others still underway to monitor. During the interim period, achievements are no longer adequately captured in terms of numbers of applications, awards, initial collaborative formations, research acceleration, and accomplishment of technical tasks. But, achievements cannot yet be captured in terms of national impacts.

During the interim period, assessment on the technical side shifts from evidence that individual project research tasks are being met, to external evidence that the project has created a body of significant new knowledge, and that the knowledge is being disseminated. On the economic side, in the absence of long-term impacts, assessment centers on evidence that someone is pursuing commercial applications of the new knowledge—particularly the innovators and their collaborators and licensees because they are expected to be positioned for early use of the technology.

Hence, the mission-driven goals for ATP's interim period are defined as (1) creating knowledge, (2) disseminating knowledge, and (3) commercializing results, and the indicator metrics for constructing the CPRS are selected to correspond to these interim goals. Although spillovers—critical to ATP's rationale and long-term success—are not called out directly, the characterization of interim goals and metrics is consistent with ATP's attention to spillovers—knowledge spillovers through knowledge creation and dissemination, and market spillovers through commercialization activities.

During the interim period, stakeholders expect ATP administrators to know how the body of projects that have completed their ATP funding are progressing and if they appear on track to deliver on long-term expectations. To this end, the CPRS was developed.

5.3 Data Constraints in Constructing ATP's CPRS

As related previously, ATP's need for metrics during the intermediate period had led, prior to the construction of the CPRS, to identification of a variety of output and outcome data to serve as indicator metrics.[26,27] In Geisler's words, "Core indicators and measures provide a picture of the outputs from S&T—as they evolve through the flow of the innovation process.…"[28]

The categories of data uniformly collected by the ATP status reports for the first 50 completed projects included the following:

- Technical awards (including name of award and presenting organization)
- Business awards (including name of award and presenting organization)
- Patents filed—granted and not yet granted (including patent numbers and titles of those granted) and patent citation trees for projects with patents granted
- Publications and presentations (including only counts)
- Products and processes on the market or expected soon (including trade names of items on the market and counts of those expected)
- Collaborations (including types of collaborations—with joint venture research partners, subcontracts, university partners, licensing arrangements, and collaborations for commercial activities)

26. The items to be collected through the mini case studies of completed projects were worked out between ATP's evaluation staff and the contractor engaged to perform the first status reports, taking into account program goals and feasibility of collecting the data.

27. Reports in fulfillment of the Government Performance and Results Act (GPRA) included a variety of metrics such as patents and projects under commercialization. In addition, results of economic case studies and econometric studies contributed to GPRA reporting.

28. Geisler (2000), p. 255.

- Attraction of additional capital (including sources of capital, but not consistently the amounts)
- Change in small-company employment (large-company employment was excluded because of the difficulty of linking it to a single project; joint venture employment was excluded because of the complexity)
- Outlook for future developments (the case analysts' qualitative assessments of outlook)

A constraint in developing the CPRS was that it be formulated to use existing variables and data compiled for the first 50 completed projects.

6.0 Specifying Indicator Variables for Use in the CPRS Formulation

This section identifies the specific indicator variables selected for use in the CPRS formulation from those available, and discusses the selection decisions. Indicator variables were selected to signal progress toward the three major program goals identified as being particularly prominent during the intermediate period.

6.1 Variables to Indicate Progress Toward Knowledge Creation

Knowledge creation is an essential component in the rating system because it lies at the heart of each ATP-funded project. Each project, to be approved, must provide convincing evidence that it has "strong potential for advancing the state of the art and contributing significantly to the U.S. scientific and technical knowledge base."[29] The technology must be highly innovative and the research challenging. A qualified research team with access to necessary research facilities must carry out the research. It is the creation of knowledge through ATP-funded research that fuels subsequent developments leading ultimately in successful projects to broad-based, national economic benefits. The challenge is to choose from the available indicator variables those that best indicate that a project has created significant scientific and technical knowledge.

6.1.1 Technical-Award Indicator

Receiving a technical award from a third-party organization provides a robust signal not only that new technical knowledge has been created, but also that it is of particular importance or

29. See Chapter 2, Section E ("Project Narrative") of the *ATP Proposal Preparation Kit,* obtainable by request from ATP by e-mail at atp@nist.gov or by phone at 1-800-ATP-FUND, and also available on ATP's Website at www.atp.nist.gov.

significance. Many projects that successfully develop new scientific or technical knowledge will not be recognized with an award. Hence, award recognition for technical and scientific achievement was selected as a way to highlight strong performers in knowledge creation.

6.1.2 Patent Indicator

Patents codify the new knowledge of an invention or technology, and thus signal that new knowledge has been created. However, patents are an imperfect indicator because companies use strategies other than patents to protect their intellectual property, including secrecy and speed to market. Hence, the presence of patents indicates knowledge creation, but the absence of patents does not necessarily indicate that knowledge has not been created.

The status reports separated out patent data into patents filed and granted, patents filed and not yet granted, and all patents filed. It is questionable which version serves best as an indicator of knowledge creation. On the one hand, the granting of a patent is a more reliable signal of knowledge creation than the filing of a patent application, because it is possible an application will be turned down on grounds that it does not contain original ideas. On the other hand, the granting of patents can take years, and for a young program like ATP, many of the completed projects had patent filings that had not yet been granted only because sufficient time had not yet elapsed. For this reason, patents filed was selected as the better indicator for ATP of knowledge creation, based on the assumption that the probability of overstating patents by using patent filings was smaller than the probability of understating patents by using patents granted, because many ATP project patents were still under review at the U.S. Patent and Trademark Office. Only patents filed during or after the ATP project, and that were a result of ATP-funded research, are counted.

6.1.3 Publications and Presentations

Publishing and making presentations are hallmarks of most research and generally accompany knowledge creation; hence, this data category also deserves consideration as an indicator of knowledge creation. Although publishing and patenting may be inhibited in cases where secrecy is an important strategy for protecting the company's intellectual property, company researchers are often able to publish and present nonproprietary aspects of their research.

The available status report data for publications and presentations consisted only of counts, providing no ability to adjust for differences in quality. The data do not distinguish between a peer-reviewed paper in a leading technical journal and an unreviewed paper in trade magazines or conference proceedings. Further, the distinction between published papers and oral presentations is not consistent. For these reasons, the combined number of publications and presentations is simply taken as a rough indicator of knowledge creation, even though it is not an ideal measure.

6.1.4 Products and Processes

To compensate for the fact that some companies strive to keep their knowledge creation secret, declining to publish, present, or to patent even when they have created new knowledge, it is important to look downstream to see if products and processes result from the research project. The emergence of new and improved product and processes can help to pick up knowledge creation missed by the other indicators.

The status reports captured both new and improved products and processes in the market and those expected to be in the market soon. To be counted, it was necessary that the products and processes be well defined and the evidence convincing to the analysts that commercialization had occurred or was imminent. The status-report case analysts provided trade names and product/process descriptions. But the indicator variables do not reliably distinguish between those products and processes that represent revolutionary or breakthrough ideas and those that represent only modest contributions. Hence, the number of new and improved products and processes provides only a rough indicator of knowledge creation.

6.1.5 Summary

The average project length is approximately three years and the post-project period before the data are compiled is 3 to 5 years; this means that most projects have 6–8 years after project award to show one or more of the above indicators that knowledge has been created. The CPRS gives the project credit for knowledge creation if there are awards for new technologies created, patent filings, publications or presentations, and new products or processes on the market or expected soon after the end of the period covered.

6.2 Variables to Indicate Progress Toward Knowledge Dissemination

The dissemination of knowledge is an important pathway for generating spillovers and the broadly diffused benefits that are an ultimate goal of ATP. Even if the award recipient fails to carry the technology into the marketplace, others may take up knowledge disseminated from the project and make something of it. And, should the award recipient successfully carry the technology into the marketplace, the dissemination of knowledge will provide an additional pathway for spillover benefits, a pathway that increases the overall potential for broad-based national benefits.

The challenge is to select from available indicator variables those that can best serve as indicators of knowledge dissemination. Some of the same variables are indicative of both knowledge creation and dissemination. Differences in the weights assigned are used to take into account differences in the strength of the relationship of a given indicator variable to knowledge creation versus knowledge dissemination.

6.2.1 Patents

Patents, by codifying new knowledge, provide a means for disseminating it. As indicated in the list of data categories provided in Section 5.3, in addition to patent counts, patent trees were constructed for the projects which had patents granted. The trees show who cited the patents, and who, in turn, cited those citations, and so forth. Because they show the intensity of citing and who is citing, patent tree data offer a potentially better indicator of knowledge dissemination than patent count data. But there are several challenges. One is how to convert the complex citation data to an indicator measure, including whether to treat foreign organizations that cite the patents differently than domestic organizations. Another challenge is what to do about the substantial numbers of patent filings, not yet granted, for which there are no patent trees. Because of unresolved issues associated with using the patent citation data and the incomplete nature of the data, the prototype CPRS was formulated to use the number of patents filed as an indicator of knowledge dissemination.

6.2.2 Publications and Presentations

Publications and presentations are a primary means through which knowledge is disseminated. Furthermore, it is a means of dissemination that is easily and inexpensively accessed. As noted earlier, the data collected support only counts—not quality—of publications and presentations.

6.2.3 Products and Processes

Through inspection and reverse engineering, knowledge can be gleaned from products and processes in the market. Obtaining project information in this way, however, tends to require more effort, be more costly, and to entail a greater lag from the time of the initial research than patents and publications. Nevertheless, the existence of products and processes appears a valid indicator that knowledge may be disseminating.

6.2.4 Technical Awards

Technical awards serve to call attention to new technology and, thereby, further knowledge dissemination. Hence, a technical-awards variable is included among the dissemination indicators.

6.2.5 Collaborations

An additional indicator of knowledge dissemination is the existence of collaborative relationships. Through collaborative relationships among researchers, and between researchers and commercial partners, project knowledge is shared. Though the extent of knowledge dissemination through collaborative relationships may be constrained in terms of the size of the affected population, the effectiveness of the dissemination tends to be strong.

6.3 Variables to Indicate Progress Toward Commercialization

6.3.1 Products and Processes

Products and processes in the market or expected soon are a direct indicator of commercial progress. Market presence signals that a project has progressed to the point that economic benefits may begin to accrue. It is taken in the CPRS formulation as the principal available indicator of commercial progress.

6.3.2 Attraction of Additional Capital

Attraction of additional capital is considered a useful indicator of progress toward commercialization because it shows that additional resources are being made available for further development and commercial efforts. Attraction of capital is generally taken as a signal that the level of technical risk has been sufficiently reduced that others are willing to invest to take the technology into use.

The available collaboration data included to some extent identification of the sources of funding: innovator financing through public stock offering or retained earnings, funding by other federal agencies and by state government investment funds, and funding through collaborative commercialization agreements. In some cases, the amount of funding by source was also provided, but not consistently. Hence, there was no way to compare the resources resulting from the alternative sources and no way to know if having multiple funding sources indicated more or less resource strength than having a single funding source. For these reasons, the indicator that was adopted for the CPRS formulation was simply whether or not additional funding had been obtained for continuation of the objectives. Information on the various sources of funding, the number of funding sources, and the partial data on the amounts of funding were not included.

6.3.3 Employment Change

A potentially useful indicator of commercialization is employment gains, but linking employment changes to a particular project may be difficult or impossible. In small companies, it is more reasonable to link a particular project to company growth. In large companies a host of other factors typically influence employment, making employment change an unreliable indicator of a project's commercial progress. In the case of joint ventures, tracking employment changes associated with a given project along task lines tends to be complex.

For these reasons, employment change data was collected by the status reports only for small, single-applicant companies. For these small companies, employment data were recorded at the project start and after the project ended by case-study analysts, and the percentage change was recorded. The small-company employment change is included in the CPRS formulation as an indicator of commercialization progress for these companies. Because most of the participants in the first 50 completed ATP projects were small companies, this indicator was available for most of the projects in the first application of the CPRS. For the large companies and joint ventures, the employment change indicator was assigned a default value.

6.3.4 Business Awards

In addition to awards for scientific and technical achievements, awards are given by third-party organizations to businesses that are demonstrating unusual business acumen. These awards are often made to small companies that are growing at a rapid rate. For small emerging businesses with a single technology focusæwhich describes many participants among the first 50 completed ATP projects—business awards appeared closely linked to the commercialization of the ATP-funded technologies. Hence, the CPRS was formulated to include a business-award variable for use as an indicator of commercial progress.

6.3.5 Outlook

For each of the first 50 completed ATP projects, the case-study analyst provided a qualitative assessment of future prospects. This allowed the analyst to bring in information beyond that revealed by the other indicator data. For example, the analyst might have found outputs and outcomes suggesting a relatively robust project, but also discovered that an alternative approach was expected soon to displace the project's technology, such that the outlook for long-term benefits was pessimistic. Or, there might be little in the data to suggest a robust project, but the case analyst might uncover a new development in the works that would give cause for optimism.

To facilitate translating the outlook descriptions into an indicator of commercial progress, the projects were divided into three groups with respect to outlook. Group 1 included those projects whose outlook the case-study analyst described as highly promising, or excellent, or on track. Group 2 included three subgroups: those whose outlook was described as neither particularly strong nor weak; those whose outlook was described as promising but with serious reservations or qualifications added; and those whose outlook was uncertain. Group 3 included those projects whose outlook was portrayed in clearly pessimistic terms. The CPRS was formulated to include a numerical outlook variable, reflecting the outlook group to which a project is assigned.

7.0 Applying Weighting Algorithms to the Selected Indicator Variables and Calculating Scores

Weights were assigned to each of the selected indicator variables to determine how they figure in the composite rating. As was the case with the other composite scoring systems reviewed in Section 7, expert judgment was used to determine the weights for the CPRS.

The range of values observed for each of the variables in the database compiled for the first 50 completed ATP projects influenced the specification of weights. An objective was to dampen the effect of outlier values. Applying the assigned weights to each indicator variable produced components of the raw score. Summing the components of the raw scores resulted in the composite raw score that was then converted to a star rating.

7.1 Knowledge Creation Scoring

Table 5 summarizes the weighting of indicator variables selected to indicate progress toward knowledge creation. Column 1 lists the variables in declining order of their assumed importance as indicators of knowledge creation. Column 2 shows the range of values for each variable observed for the first 50 completed ATP projects. Column 3 shows the weighting algorithm for each variable. Column 4 gives the range of raw scores calculated for the first 50 completed projects by applying the weighting algorithm to each indicator variable. The bottom row of the table shows the range of aggregated raw scores for knowledge creation.

7.1.1 Weighting the Technical-Award Indicator

Technical awards among the first 50 completed ATP projects ranged from zero to four in number. That is, some projects received as many as four awards from different organizations. Technical awards are assumed to serve as the best indicator that significant knowledge was created. The awarding of multiple awards by different organizations seemed largely independent of one another, and the decision was made to assign equal weights to each additional award. The raw score for technical awards is calculated by multiplying 1, the weight, times N, the number of science and technical awards received. The raw scores for awards ranged from 0 to 4 for the first 50 completed ATP projects.

7.1.2 Weighting the Patent Indicator

Patent filings ranged in number from 0 to 26 per project, with the range reflecting patenting strategy as well as the amount of knowledge created. That is, one project may file a single patent to capture its knowledge creation, while another may file many patents to capture a comparable or different amount of knowledge. The weighting algorithm values the first

TABLE 5

Calculation of Raw Scores for Knowledge Creation

Selected Indicator Variable (col. 1)	Range of Observed Values (col. 2)	Weighting Algorithm Applied to Indicator Variable Value (N) (col. 3)	Range of Calculated Raw Scores (col. 4)
Technical awards	0 to 4	1 * N	0 to 4
Patent filings	0 to 26	0.5 * Square root (N)	0 to 2.5
Publications & presentations	0 to 214	0.5 * 4th root (N)	0 to 1.9
Products & Processes on the market or expected soon	0 to 5	If N>1, assign value of 0.5; otherwise, 0	0 or 0.5
Aggregate raw score, knowledge creation			**0 to 8.9**

patent at half that of a technical award, and additional patents at a sharply decreasing rate. The raw score for patent filings is calculated as 0.5 times the square root of the number of patent filings. The raw score for patent filings ranged from 0 to 2.5 for the first 50 completed ATP projects.

7.1.3 Weighting the Publication/Presentation Indicator

Publications and presentations ranged in number from 0 to 214, with one project having far more than the rest. The weighting algorithm values the first publication or presentation the same as the first patent, and additional publications and presentations at an even more sharply decreasing rate. The raw score is calculated as 0.5 times the fourth root of the number of publications and presentations. The raw scores ranged from 0 to 1.9 for the first 50 completed ATP projects.

7.1.4 Weighting the Product/Process Indicator

Products and processes in the market or expected soon ranged in number from 0 to 5. Some projects had more than one product, some a combination of product and process. Where there were multiple products or processes for a project, it is assumed they all serve to indicate the same underlying body of knowledge creation. Hence, having multiple products and processes does not increase the raw score. The weighting algorithm is binary: 0 if there are no products or processes; 0.5 if there are. This assigns the same value to having any products and processes as to having a single patent or publication. The raw scores ranged from 0 to 0.5 for the first 50 completed ATP projects.

7.1.5 Aggregate Raw Scoring of Progress Toward Knowledge Creation

The aggregate score for each project is computed by summing across the scores for the four indicators. The aggregate raw scores for knowledge creation for the first 50 completed ATP projects ranged from 0 to 8.9.

7.1.6 Sensitivity Testing of Knowledge Creation Scores

Table 6 summarizes the results of sensitivity testing of the aggregated raw score for knowledge creation to changes in the values of each of the indicator variables.[30] An inspection of the tabular values reveals the marginal contribution of each indicator to the aggregate raw score for knowledge creation. For example, rows 1–4 show that the first technical award increases the raw score by 1.0; the first patent filing, 0.5; the first publication or presentation, 0.5; the first product or process, 0.5. If a project had one of each of the indicator variables, the raw score would be 2.5. Comparing rows 1–4 with rows 6–9 shows the marginal contribution of the second unit of each indicator. Receiving a second technical award adds 1.0 to the raw score; filing for a second patent adds 0.2 to the raw score; publishing or presenting a second paper adds 0.1 to the raw score; adding another product or process in the market adds 0.0 to the raw score. The weighting scheme moderates the effect of unusually high rates of publishing and patenting for one of the projects in the sample, as can be seen by comparing rows 17–19 with rows 2 and 3. Having 1 publication yields a score of 0.5, whereas having 200 yields a score of 1.9. Having 1 patent filing yields a score of 0.5, whereas having 20 patents yields a score of 2.2. Row 19 shows the effect of inserting values toward the upper end of the observed ranges for each of the indicator variables at once.

7.2 Knowledge Dissemination Scoring

Table 7 summarizes the application of weights to the variables selected to indicate knowledge dissemination. Column 1 lists the variables in declining order of their relative weights. Column 2 shows the range of values for each selected variable observed in the database. Column 3 shows the weighting algorithms. Column 4 gives the range of raw scores for each variable calculated by applying the weighting algorithm to the corresponding indicator variable. The

30. Use of particular combinations of values for the indicators does not imply that those combinations are likely to occur in real projects. At the same time, all the values used in the testing are within the ranges observed in the first 50 completed ATP projects.

TABLE 6
Sensitivity of Knowledge Creation to Changes in Indicator Values

Row #	Technical Awards (col. 1)	Patents Filed (col. 2)	Publications & Presentations (col. 3)	Products & Processes (col. 4)	Aggregate Raw Score (col. 5)
1	1	0	0	0	1.0
2	0	1	0	0	0.5
3	0	0	1	0	0.5
4	0	0	0	1	0.5
5	1	1	1	1	2.5
6	2	0	0	0	2.0
7	0	2	0	0	0.7
8	0	0	2	0	0.6
9	0	0	0	2	0.5
10	2	2	2	2	3.8
11	3	0	0	0	3.0
12	0	10	0	0	1.6
13	0	0	10	0	0.9
14	0	10	10	3	3.0
15	3	3	3	3	5.0
16	4	0	0	0	4.0
17	0	20	0	0	2.2
18	0	0	200	0	1.9
19	4	20	200	5	8.6

bottom row of the table shows the range of aggregated raw scores for knowledge dissemination across the first 50 completed ATP projects.

7.2.1 Weighting the Publication/Presentation Indicator

A step function was used in weighting the dissemination value of publications and presentations. A single publication or presentation will result in a weight of 1. Additional units up to 10 add to the score at a sharply declining rate; and units in excess of 10 add at an even slower

TABLE 7
Calculation of Raw Scores for Knowledge Dissemination

Selected Indicator Variable (col. 1)	Range of Observed Values (col. 2)	Weighting Algorithm Applied to Indicator Variable Value (N) (col. 3)	Range of Calculated Raw Scores (col. 4)
Publications & presentations	0 to 214	1 * Square root (N_1 to N_{10}) + 0.1 * Square root ($N_{>10}$)	0 to 4.6
Patent filings	0 to 26	1 * Square root (N_1 to N_{10}) + 0.1 * Square root ($N_{>10}$)	0 to 3.6
Collaborations	0 to 3	1 * N	0 to 3
Products & processes on the market or expected soon	0 to 5	0.5 * Square root (N)	0 to 1.1
Technical awards	0 to 4	0.25 * Square root (N)	0 to 0.5
Aggregate raw score, knowledge dissemination			**0.7 to 12.8**

rate. The assumption is that when a project produces multiple publications and presentations they are closely related, and each subsequent publication or presentation does not convey as much information as the previous one.[31] The approach avoids having one project's extremely large number of publications/presentations overwhelm all other variables in the aggregate raw score. The starting range of 0 to 214 translates to a range of weighted raw scores of 0 to 4.6 after the application of the weighting algorithm.

7.2.2 Weighting the Patent Indicator

The step-function weighting algorithm for patent filings is identical to that for publications and presentations, based on a similar rationale that as the number of patents in a project

31. At the same time, it could be argued to the contrary that multiple publications and presentations may exhibit a critical mass effect, where the dissemination value increases at an increasing rate rather than at a decreasing rate as the number increases. This is an example of an area that could benefit from further research.

increases, the contribution to knowledge dissemination of additional units declines. The range of weighted raw scores is 0 to 3.6.

7.2.3 Weighting the Collaboration Indicator

Three forms of collaborative relationships are tracked, namely (1) collaborations between award recipients and university researchers, (2) R&D collaborations among award recipients and other firms, state and federal laboratories, and other non-university organizations, and (3) collaborative ties between award recipients and other firms for technology commercialization. The weighting algorithm assigns a weight of 1 to each of these forms of collaborative relationship found in a project. The rationale is that each type of collaboration provides a different pathway of knowledge flows. Among projects in the sample of 50, some had no collaborative relationships; some had one, two, or three of the three forms listed. Hence, the range of weighted raw scores is 0 to 3.

7.2.4 Weighting the Product/Process Indicator

The weighting algorithm for products and processes effectively treats products and processes as half as important as patents and publications/presentations as knowledge disseminators. The rationale for the lower weight is the additional effort and difficulty in extracting knowledge by inspection and reverse engineering of product and processes. The weighted raw score ranges from 0 to 1.1, depending on the number of products and processes.

7.2.5 Weighting the Technical Award Indicator

The weighting algorithm for receipt of technical awards effectively treats technical awards as one-fourth as important as patents and publications/presentations. The weighting algorithm reflects the fact that technical awards raise awareness of a new technology, and hence, further disseminate knowledge, but the award does not itself typically convey much detailed knowledge. Since multiple awards may call greater attention to the new technology, the weight increases slightly with increasing numbers of awards. The weighted raw score ranges from 0 to 0.5, based on awards that range from 0 to 4.

7.2.6 Aggregate Raw Scoring of Progress Toward Knowledge Dissemination

The aggregated weighted raw score for knowledge dissemination ranges from 0.7 to 12.8. The lower end of the range is positive, despite the fact that the lower end of the ranges for each of the component indicator variables is 0, and, further, that the low-end of the range for the aggregate weighted raw score for knowledge creation is 0. This may seem a contradiction, but is not for two reasons. (1) The knowledge dissemination score contains collaboration as an indicator of progress, and one of the joint-venture projects in the sample exhibited none of the measured indicators except collaborative activity. (2) Although one or more projects had 0 values for each of the other variables, no project had 0 values for all of the variables.

7.2.7 Sensitivity Testing of Knowledge Dissemination Scores

Table 8 summarizes sensitivity testing of the aggregated raw score for knowledge dissemination to changes in the values of indicator variables. Rows 1 through 5 show the value of the first unit of each variable, with a single publication or presentation resulting in a score of 1.0; a single patent, 1.0; a single type of collaboration, 1.0; a single product or process, 0.5; and a single technical award, 0.3. Row 6 shows that if one of each of the indicators were obtained, a raw score of 3.8 results.

Comparing rows 7–12 with 1–6 shows the marginal effect of adding the second unit of each variable. The second publication or presentation adds 0.4 to the score. The second patent filed adds 0.4 to the score. The second type of collaboration adds 1.0 to the score. The second product or process adds 0.2, and the second technical award adds 0.1. Having two versus one of each of the indicators results in a score of 5.9, compared with 3.8.

The small differences revealed by comparing rows 13, 14, and 15 reflect the step function used in the weighting algorithm for publications (and also patents). Increasing the number of publications and presentations from 10 to 11 increases the raw score by only 0.1. Increasing the number from 11 to 200 increases the raw score by only 1.2. Rows 15–19 show the resulting raw score if values near the upper end of the range are used in turn for each of the indicator variables. Row 20 shows the resulting raw score of 12.6 from using values near the upper end of the range for all of the indicator variables at once.

7.3 Commercialization Progress Scoring

Table 9 summarizes the application of weights to the variables selected to indicate commercialization progress. Column 1 lists the four variables used to indicate progress toward commercialization. Each of the variables used has a potentially substantial impact in the scoring. Column 2 shows the range of values observed for each indicator variable in the database of the first 50 completed ATP projects. Column 3 presents the weighting algorithms applied to the indicator variables to calculate the raw scores, shown in column 4. The bottom row shows the range of aggregated raw scores for commercialization.

7.3.1 Weighting the Product/Process Indicator

Products and processes in the market now or expected soon are given a relatively heavy weight, because it indicates the likelihood that a project has progressed to a stage in which economic benefits may result. A primary interest is that there be a product or process. The weighting algorithm reflects the relatively large importance attached to this indicator, with the

TABLE 8

Sensitivity of Knowledge Dissemination to Changes in Indicator Values

Row #	Publications & Presentations (col. 1)	Patents Filed (col. 2)	Collaborations (col. 3)	Products & Processes (col. 4)	Technical Awards (col. 5)	Aggregate Raw Score (col. 6)
1	1	0	0	0	0	1.0
2	0	1	0	0	0	1.0
3	0	0	1	0	0	1.0
4	0	0	0	1	0	0.5
5	0	0	0	0	1	0.3
6	1	1	1	1	1	3.8
7	2	0	0	0	0	1.4
8	0	2	0	0	0	1.4
9	0	0	2	0	0	2.0
10	0	0	0	2	0	0.7
11	0	0	0	0	2	0.4
12	2	2	2	2	2	5.9
13	10	0	0	0	0	3.2
14	11	0	0	0	0	3.3
15	200	0	0	0	0	4.5
16	0	20	0	0	0	3.5
17	0	0	3	0	0	3.0
18	0	0	0	5	0	1.1
19	0	0	0	0	4	0.5
20	200	20	3	5	4	12.6

first product/process receiving a score of 4.25. Additional products/processes contribute less. The raw score ranges from 0 to 5.8.

7.3.2 Weighting the Capital Attraction Indicator

Capital attraction indicates that additional resources are being made available for further development and commercial efforts. Due to data difficulties discussed earlier, a single-value weight is assigned to a project if it has attracted funding from one or more of the sources identified, regardless of amount. The indictor variable is assigned a weight of 0 if a project has attracted no additional funding, and a weight of 3 if it has attracted funding from any of the identified sources.

7.3.3 Weighting the Employment Gains Indicator

As indicated earlier, employment data were compiled only for the small, single-applicant companies. Because most of these companies were very small at project start, the percentage increases for a number of the projects tended to be very large. Growth rates of 50% were routine, and rates approached 2000% for several projects. For this reason, the weighting strategy for small companies is to treat rates of employment change up to 50% as the norm, assigning a weight of 0 for rates of change of 50% or less. If employment increased by more than 50%, a weight of 2.5 times the fourth root of the gain in excess of 50% is assigned, providing a relatively large weight, while preventing the extremely large gains in several cases from overwhelming other indicators. In the case of bankruptcyæeither of a small company, single applicant, or a leader of a joint ventureæa negative weight of –6 is assigned to signal that there is a serious impediment to commercial progress through the direct path of the innovating companies. In cases other than bankruptcy, a default weight of 1.5 is assigned to projects for which employment data were not collected (i.e., large companies and joint ventures). Applying the weighting algorithm resulted in raw scores ranging from –6 to 5.2.

7.3.4 Weighting the Business Awards Indicator

The weighting algorithm for business awards is to assign 0 if there are no awards and 3.25 for one award. Additional awards add 0.25 each to the score. The weighted raw scores range from 0 to 3.8.

7.3.5 Weighting the Outlook Indicator

As indicated earlier, the subjective information in the cases was used to group the projects into three groups by outlook. The weighting strategy assigns a raw score of +4 to those in group 1, whose outlook is strong; 0 to those in group 2, whose outlook is neither clearly strong nor poor; and –4 to those in group 3, whose outlook is poor.

7.3.6 Aggregate Raw Scoring of Progress Toward Commercialization

The aggregated weighted raw scores for commercialization range from -10 to 21.7. The range for this score is much wider than for knowledge creation and dissemination scoring, and may be negative if the company goes out of business, or if the outlook is poor, or if both occur.

TABLE 9

Calculation of Raw Scores for Commercialization

Selected Indicator Variable (col. 1)	Range of Observed Values (col. 2)	Weighting Algorithm Applied to Indicator Variable Value (N) (col. 3)	Range of Calculated Raw Scores (col. 4)
Products & Processes on the market or expected soon	0 to 5	3 + 1.25 * Square root (N)	0 to 5.8
Capital attraction	0 or 3	If none, 0 If yes, 3	0 to 3
Employment gains	% change for small firms only	If bankruptcy, -6, If JV or large firm, 1.5 If employment change <=50%, 0 If employment change >50%, 2.5 * fourth root of gain in excess of 50%	-6 to 5.2
Business awards	0 to 3	3 + 0.25 * N	0 to 3.8
Outlook	Qualitative analysis translated to a value	-4 = poor outlook; 0 = neither strong nor poor outlook; +4 = strong outlook	-4 to +4
Aggregate raw score, commercialization			**-10 to 21.7**

Scoring at the low end of the range are projects conducted by companies that went bankrupt and for which the outlook revealed no active alternative champion of the technology to take it forward. At the high end were projects conducted by award-winning businesses—particularly those that are fast growing—with commercialized products or processes based on the technology, available resources to continue development and commercialization of the technology, and an outlook for continued robust progress.

7.3.7 Sensitivity Testing of Commercialization Scores

Table 10 summarizes the results of testing the sensitivity of the aggregate raw score for commercialization to changes in the value of each of the indicators, and to several changes in combination. The use of certain combinations of values here and in the previous sensitivity testing is not to imply that they are expected to occur in the combination shown, but rather the intention is to test the sensitivity of results to extreme values.

For the purpose of clearer exposition, two of the variables, employment gains and outlook, are entered in descriptive terms rather than in terms of the numerical value into which the description is translated. The model includes four possible conditions of employment gain, only one of which allows the score to change as a function of the amount of increase in the gain. It contains three possible outlook states.

The sensitivity testing begins with the condition of employment gain. Rows 1–3 show the effect of holding the outlook constant at Group 2 (neither clearly positive or negative and having a 0 effect on the raw score), holding other variables except employment gains constant at 0, and changing only the condition of employment gains. A joint venture or large company for which meaningful employment data were not collected receives a default score of 1.5. A small company, drawn from a group in which up to 50% employment gains were commonplace, receives a score of 0. Beyond a 50% employment gain, the score is positive and increasing at a decreasing rate, with a 100% gain receiving a score of 2.1, and a 200% gain, a score of 2.8.

Comparing rows 1 and 5 reveals the effect of changing only the outlook from Group 2 (neither clearly positive nor negative) to Group 1 (positive). The raw score increases by 4.0, showing a relatively large impact on scoring of the outlook.

Comparing rows 6–8 with row 1 shows the contribution to the raw score of the first unit of each of the other indicator variables. Commercializing a product or process adds 4.3 to the raw score; attracting additional capital adds 3.0; receiving a business award adds 3.3. Row 9 shows a raw score of 16 resulting from the coincidence of a joint venture or large company having one unit of each indicator variable, combined with a positive outlook.

Comparing row 10 with row 6 reveals the effect of commercializing a second product or process to be 0.5. Comparing row 11 with row 8 reveals the effect of a second business award to be 0.2. Because capital attraction is modeled as either yes or no, it is not tested for other possible values.

Comparing rows 12 and 13 with row 3 isolates the effect of changing only the outlook. There is an 8 point drop in the raw score as the outlook changes from positive to negative.

Rows 12, 14, and 15 combine increasing changes in small-company employment with a positive outlook, producing raw scores of 6.1 for a 100% employment gain/positive outlook,

TABLE 10

Sensitivity of Commercialization Scores to Changes in Indicator Values

Row #	Products & Processes (col. 1)	Capital Attraction (col. 2)	Employment Gains (col. 3)	Business Awards (col. 4)	Aggregate Outlook (col. 5)	Raw Score (col. 6)
1	0	0	JV/ Large Co.	0	Group 2	1.5
2	0	0	Small Co., 50%	0	Group 2	0
3	0	0	Small Co., 100%	0	Group 2	2.1
4	0	0	Small Co., 200%	0	Group 2	2.8
5	0	0	JV/ Large Co.	0	Group 1	5.5
6	1	0	JV/ Large Co.	0	Group 2	5.0
7	0	1	JV/ Large Co.	0	Group 2	4.5
8	0	0	JV/ Large Co.	1	Group 2	4.8
9	1	1	JV/ Large Co.	1	Group 1	16.0
10	2	0	JV/ Large Co.	0	Group 2	6.3
11	0	0	JV/ Large Co.	2	Group 2	5.0
12	0	0	Small Co., 100%	0	Group 1	6.1
13	0	0	Small Co., 100%	0	Group 3	-1.9
14	0	0	Small Co., 500%	0	Group 1	7.6
15	0	0	Small Co., 2000%	0	Group 1	9.3
16	0	0	Small Co., Bnkrpt	0	Group 3	-10.0
17	5	1	Small Co., 2000%	3	Group 1	21.8

Note: JV denotes a joint-venture project; Co. abbreviates company; and Bnkrpt abbreviates bankruptcy. Employment-gains data were unavailable for joint ventures and large companies, and a proxy value of 1.5 is used in lieu of real data. Group 1 had a positive outlook, Group 2 a neutral, clouded, or indeterminate outlook, and Group 3 a poor outlook.

7.6 for a 500% gain/positive outlook, and 9.3, for a 2000% gain/positive outlook. In contrast, row 16 combines a 100% loss in employment, signaling bankruptcy, with a negative outlook, producing a raw score of –10.

The last row of the table, row 17, combines strong performance values for each of the five commercialization indicator variables, producing a raw score of 21.8. One of the projects achieved a score very close to this.

7.4 Composite Scoring and Star Rating

The composite raw score is calculated for each project in the portfolio by summing the project's raw scores for knowledge creation, knowledge dissemination, and commercial progress. The composite raw score is factored by 6 to facilitate dividing the projects into five groups of adjusted composite scores. The group with a score of 4 or higher receives the highest rating of 4 stars; a score less than 4 but at least 3, 3 stars; less than 3 but at least 2, 2 stars; less than 2 but at least 1, 1 star; and less than 1, 0 stars.

7.4.1 Range of Ratings for First 50 Completed Projects

For the first 50 completed ATP projects, the computed composite raw scores ranged from –9.0 to 30.8, and the range of adjusted composite scores ranged from 0 to 5. The number of stars assigned these 50 projects ranged from 0 to 4.

7.4.2 Sensitivity Testing of Composite Scores and Star Ratings

This section adds to the previous sensitivity testing by examining how the overall scores and star ratings change in response to alternative input values of indicator variables. As a starting point, consider the raw scores necessary to produce the various star ratings. A 4-star rating requires a composite raw score of 24 or greater. A 3-star rating requires a composite raw score of at least 18 but less than 24. A 2-star rating requires a composite raw score of at least 12 but less than 18. A 1-star rating requires a composite raw score of at least 6, but less than 12. And, a 0-star rating requires a composite raw score less than 6.

Table 11 shows the composite scores and star ratings associated with different combinations of values for the indicator variables. To simplify comparisons, all seven hypothetical cases presented are based on a very small single-applicant firm. A degree of collaboration is assumed for all cases, because most companies in the ATP, particularly very small companies, have some form of collaboration.

The first case shows early publication and patenting progress, but no commercial follow-through and a cloudy outlook casting doubt on further progress. It receives a zero star rating. The second case has the same outputs as the first case, but it has a favorable outlook that boosts its star rating to 1. The third case shows some progress across the board and a favorable outlook, but only modest firm growth (from 2 to 4 persons). It receives a moderate star rating of 2. The fourth case shows no patents or publications but relatively active progress on the commercial side, more robust company growth (from 2 to 12 persons), more collaboration, and a favorable outlook. It receives a relatively robust score of 3. The fifth case is essentially the same as the fourth case, except that it substitutes knowledge outputs for one of the commercial outputs. Like the preceding case, it receives a star rating of 3. The sixth case is the

TABLE 11

Sensitivity Testing of Composite Scores and Star Ratings to Variations in Values of Indicator Variables*

Indicator Variable	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Patents	1	1	1	0	1	1	4
Publications	1	1	1	0	1	1	4
Products	0	0	1	2	1	1	2
Attraction of capital	0	0	1	1	1	1	1
Collaboration	1	1	1	2	1	1	3
Employment gain (%)	50	50	50	500	500	500	500
Technical awards	0	0	0	0	0	1	0
Business awards	0	0	0	0	0	1	0
Outlook	0	+	+	+	+	+	–
Composite score	0.67	1.33	2.71	3.10	3.32	4.07	2.89
Composite star rating	0	1	2	3	3	4	2

* All cases are based on a small, single-applicant firm.

same as the fifth case, except that it adds technical and business awards. It receives the highest star rating of 4. The seventh case strengthens the outputs and collaboration, but drops the awards and changes the outlook from favorable to unfavorable. The star rating drops to 2 despite the past accomplishment. These cases show that different combinations of output, outcome, and outlook data can produce the same or different composite ratings.

7.4.3 CPRS Calculation Database

The CPRS is implemented via an Access database with the weighting algorithms embedded in calculation queries. Values of the indicator variables for each project are entered into a form, along with other non-CPRS project data. Data for additional completed projects are added as the portfolio grows. All indicator variables, CPRS adjusted scores, and corresponding star ratings can then be analyzed by building queries and running reports.

7.4.4 Illustration of Four-Star and Zero-Star Projects

Table 12 summarizes scoring information for a project receiving 4 stars. The project was an information technology project conducted by a small company, Engineering Animation, Inc. The company won technical and business awards, and had many collaborations, large employment growth, and several products on the market. Table 13 summarizes scoring information for a project receiving no stars. The project was an electronics joint venture project led by a small company, Hampshire Instruments, Inc., which went bankrupt just after the project completed.

7.4.5 Distribution of First 50 Completed ATP Projects by Performance

Figure 2 shows the distribution of the first 50 completed ATP projects by performance as scored by the CPRS. As may be seen, the largest group of projects, 32 percent, scored in the two-star category—accomplishments, but not particularly robust progress overall. Twenty-six percent scored in the bottom category (1 star or less). Sixteen percent scored in the top category, receiving 4-stars, while an additional 26 percent also showed relatively robust progress, scoring in the 3-star category. These results are consistent with the program's expectation that not all the project will be strong performers, given the challenging nature of their undertakings.

Figure 2. Distribution of Completed ATP Projects by CPRS

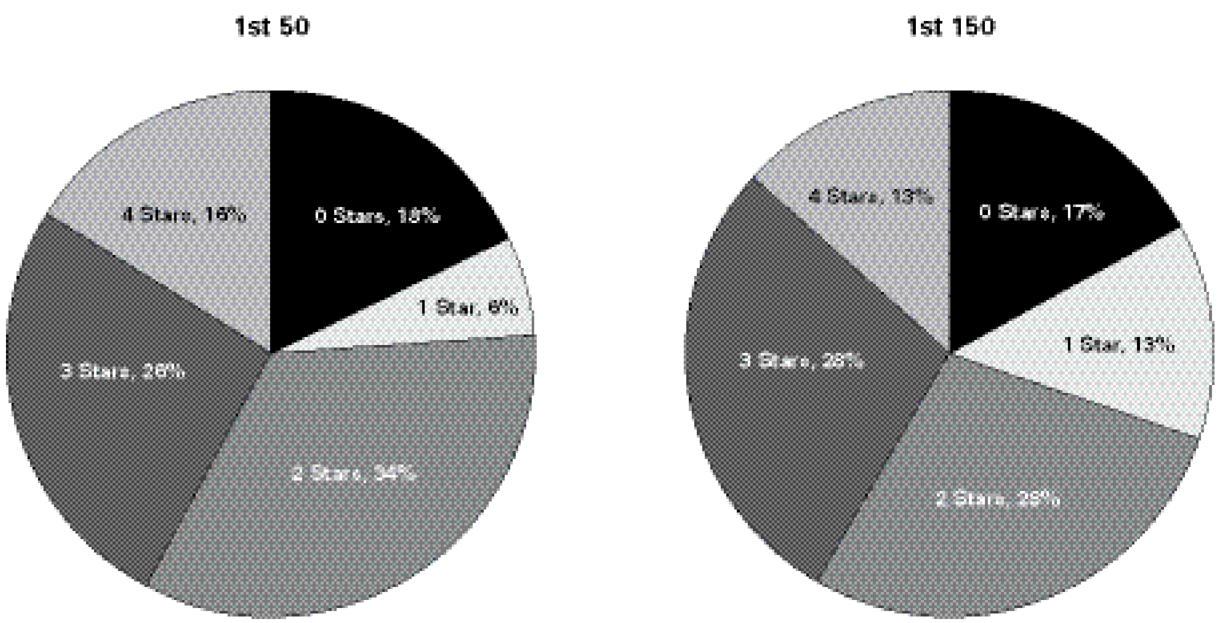


TABLE 12

Four-Star Project (Engineering Animation, Inc.)

Indicator Variable	Value of Indicator Variable	Raw Score
Knowledge Creation		
Technical awards	4	4.0
Patent filings	0	0.0
Publications & presentations	0	0.0
Products & processes on the market or expected soon	5	0.5
Total raw score		4.5
Knowledge Dissemination		
Technical awards	4	0.5
Collaborations	3	3.0
Patents	0	0.0
Publications & presentations	0	0.0
Products & processes on the market or expected soon	5	1.1
Total raw score		4.6
Commercialization Progress		
Products & processes on the market or expected soon	5	6.0
Capital attraction	Yes	3.0
Employment gains	Starting at 20; ending at 400 employees	5.2
Business awards	3 awards	3.8
Outlook	Strong outlook	4.0
Total raw score		22.0
Composite Scores		
Composite raw score		31.1
Composite adjusted score		5
Composite star rating		****

TABLE 13

Zero-Star Project (Hampshire Instruments, Inc.)

Indicator Variable	Value of Indicator Variable	Raw Score
Knowledge Creation		
Technical awards	0	0
Patent filings	0	0
Publications & presentations	0	0
Products & processes on the market or expected soon	0	0
Total raw score		0
Knowledge Dissemination		
Technical awards	0	0
Collaborations	1	1
Patents	0	0
Publications & presentations	0	0
Products & processes on the market or expected soon	0	0
Total raw score		1
Commercialization Progress		
Products & processes on the market or expected soon	0	0
Capital attraction	No	0
Employment gains	Lead company bankrupt	-6
Business awards	0	0
Outlook	poor outlook	-4
Total raw score		-10
Composite Scores		
Composite raw score		-9
Composite adjusted score		0
Composite star rating		0 stars

Those projects with one or no stars were generally those that showed few outward signs of progress (in terms of the indicator metrics) toward contributing to technology creation, dissemination, and commercialization. This group also included those that showed early signs of progress, but then faltered. The two-star category generally included projects that showed modest but no overly robust signs of progress and projects that had shown progress but whose future prospects seemed clouded or unfavorable. The two-star category also may include technologies that are just slow to develop or that take more time to develop than allowed by the assessment time frame. The three- and four-star categories included projects that made sustained progress, continuing into commercialization, with favorable prospects for the future.

8.0 Critique of CPRS

The CPRS is a management and communications tool particularly useful for providing an easy-to-grasp assessment of project and portfolio performance in the intermediate period of a public R&D program. The tool facilitates bridging from the project level to the portfolio level. The CPRS can be practically implemented as demonstrated by ATP's experience. It is intended to distinguish among varying degrees of progress toward achieving a program's multiple goals based on indicator metrics that relate to those goals.

8.1 Data Limitations

The CPRS was specifically formulated to use indicator variables available from ATP's case studies of its first 50 completed projects. Once developed, the CPRS can be exercised with new data derived either from case studies or from surveys such as ATP's Business Reporting Survey (BRS), provided data compatibility is maintained.[32] The formulation of the CPRS to use available data meant that it would be practical to implement, but it also meant that the system might use less than ideal data. For example, the available data on publications were simple counts, whereas, data adjusted for quality and significance might provide a better indicator of knowledge creation and dissemination. Similarly, available data on capital attraction indicated the source but not consistently the amount of funding; yet, resource amount is likely

32. For a description of the BRS and findings from that survey, see Powell and Lellock (2000). The BRS is a primarily electronically administered survey of ATP project participants that has a post-project component that collects survey data overlapping to some extent the data collected by the status reports. It may be used alternatively to provide part of the data needed to implement the CPRS. Without the accompanying case-study approach, however, the framework described in Part I and the linkages between the distribution of CPRS ratings for the portfolio and the published detailed case studies of projects comprising the portfolio will not be realized.

more important than funding source as an indicator of commercial progress. The available data on products and processes at the time the CPRS was formulated identified products by name and description, but did not consistently provide sales volumes or commercialized value; yet, if market data had been consistently available, it would have provided a stronger indicator of progress toward commercialization than a count of products and processes in the market or expected soon.

Some relevant categories of data were missing altogether. For example, the available data did not provide an indicator of knowledge creation embodied as human capital only, that is, knowledge that may reside in the minds of research staff and may show up at a later time in outputs associated with other efforts. For some programs that emphasize university research, a measure sometimes used to capture human capital is the number of graduate students trained, but this measure is not generally applicable to ATP projects and was not captured in the case-study data. To the extent that knowledge is created and embodied only in the minds of researchers, it is omitted from the CPRS formulation even though it is potentially important.

"Outlook" is highly subjective, and the analysts may not have consistently captured outlook over the same time periods for all the projects. In some cases, case analysts spoke of the future in terms of relatively short-term events; in other cases, they spoke of future prospects in more sweeping terms. Possible variability in time horizons covered by outlook calls into question the actual time horizon covered by the CPRS.

In short, there may be other variables, or refined specifications of variables, that would better indicate progress toward program goals during the intermediate period than those specified for the current CPRS formulation. On the other hand, those variables used have proven feasible to compile.

8.2 Methodological Issues

There are methodological issues associated with construction of composite ratings. The construction of the CPRS is ad hoc and improvisational, reflective of the absence of underlying theory to guide composite ratings. On the other hand, there is precedence for developing empirically based composite rating systems and for using expert judgment to assign weights to the selected indicator variables. The selection of indicator variables and the weighting algorithms specified in the CPRS are based on expert judgment informed and constrained by observations of actual data from the first 50 ATP completed projects. There is a lack of empirical verification of the relationships modeled; problems may lie with the formulation. For example, perhaps one goal should have received more or less weight relative to another than is built into the model. Perhaps one indicator should be given more or less weight than another in the scoring of progress toward a given goal. At this time, there is no way to know.

Construction of the CPRS entails aggregation of diverse data. Although we have learned that apples and oranges cannot be added, general pieces of fruit can be. That is, shifting the category up a level can reduce the incompatibility problem. Here we define the variables as indicators of progress toward a common set of goals. These indicators are related; combining them does not result in a single measure that is incomprehensible.

8.3 An Inexact Measure

Due to the limitations discussed above, CPRS ratings should be viewed as roughly indicative of overall project performance in terms of progress toward intermediate program goals. It should be noted that the rating system sorts projects into high, medium, and low performers; it does not provide dollar estimates of their contributions to long-run national economic benefits. For this reason, projects with the same composite ratings are not necessarily equal in their potential to deliver long-run benefits. Projects with similar ratings have comparable composite levels of outputs/outcomes/outlooks during the relevant time, but their actual economic impacts may differ substantially.

A high rating signals strong expectations about a project's progress toward contributing to ATP's goals. A low rating casts doubt on that expectation. But in neither case do the ratings rule out the possibility of surprise. The ratings are based on information compiled at a point in time. Projects advance at differing rates. Future developments could alter expectations about a project's long-run success.

A further point to note is that the rating system does not incorporate a separate measure of the role of ATP in the score, that is, it does not separate out that part of progress that is directly attributable to ATP. ATP, for example, may cause a technology to be developed that otherwise would not have been developed. It may accelerate technology development by a given number of years. It may change the scale and scope of projects. The CPRS examines project progress against program goals per se. Data on ATP's role has been collected by status-report case studies and by survey for most completed projects thus far, but attempting to combine a measure of ATP's role with project progress indicator data did not seem feasible. Data on the role of ATP, presented as a separate factor, can be viewed in conjunction with the performance ratings.[33]

To the extent that a project's performance radically changes in the out years in ways not captured by the case-study analyst's outlook, the CPRS ratings will be a poor predictor of a project's longer-run performance. Furthermore, to the extent that a project with little in the way of the outputs/outcomes measured has at least one output/outcome that eventually yields unusually high benefits, the ratings will not have good predictive value. Consider, for example, the case where a project's only measured output/outcome is a single publication that some years later has a profound impact on another organization's accomplishment. The project's performance rating will have been low, yet its ultimate benefits may be high. This possibility

33. See, for example, the table on ATP effects reported in ATP (2001), p. 24.

suggests that one should be careful not to dismiss low-scoring projects prematurely. On the other hand, it is thought that most projects that have produced few if any of the specified indicators throughout the project and extending 3 to 4 years into the post-project period will be unlikely to suddenly bloom.

A potential test of the predictive value of the rating system would be to monitor a selection of projects in each performance category over time. One such test would be to determine whether the four-star projects continue to progress and deliver significant benefits at a higher rate than the 0- and 1-star projects.[34]

The long-term predictive value of the CPRS for 2-star performers may be more difficult to determine, mainly because the category includes projects with three different types of outlook, all of which receive a neutral rating: (1) those whose future outlook was considered neither strongly positive nor strongly negative, (2) those for which there were both positive and negative elements which seemed largely offsetting, and (3) those for which the outlook was considered too uncertain to call. Hence, it would not be counter to the rating system's findings if a 2-star performer eventually emerged as either a highly successful project, a largely unsuccessful project, or remained a moderate performer. It may be possible to refine the CPRS to better distinguish across projects with the different types of outlook. One step would provide a finer breakout of outlook categories in the CPRS scoring; a related step would support the finer breakout by improving the quality of outlook data provided by case-study writers.

9.0 Summary and Conclusions

This report has presented a new framework of evaluation, together with a new evaluation tool embedded in the framework. Together, the framework and the tool boost the potency of the case study methodæone of the mainstays of program evaluation. The result is an evaluation methodology that allows program administrators and project managers to bridge from individual project case study to portfolio analysis, and to answer a question of central importance to public policy makers and other stakeholders, namely, how are projects in ATP's portfolio performing overall in the intermediate period against ATP's mission-driven multiple goals? The methodology provides a practical tool that can facilitate a deeper understanding of a program's portfolio of funded projects, and yet convey an easy-to-grasp measure of overall performance.

---

34. In fact, there has already been a partial test of the predictive value of the rating system. Data for 38 of the 50 completed ATP projects in the sample used to develop the CPRS had been collected by Long in 1997 (see Long 1999). Ruegg made a check four years later, in conjunction with preparing the overview chapter for the new status report (see ATP 2001) of top-rated projects from the earlier group. The check showed them all to be continuing their relatively strong performance in terms of further commercial progress.

Case study is just one of a set of evaluation methods that ATP and most other public R&D programs use for assessment. The framework and new tool presented here are rooted in case study and, hence, stand on the shoulders of an existing approach.

ATP developed the framework partly as a result of foresight and partly through a series of evolutionary steps. The effect was to move from conducting single-project and cluster-of-project case studies presented individually, to defining a workable portfolio of projects, all of which would receive a "mini case study" and be subject to uniform collection of a set of progress indicator data. The period of focus was after project completion and prior to long-term benefits realization. Detailed economic case studies for a subset of the projects in the portfolio allow the estimation of minimum net benefits for the portfolio or the program. Aggregation of the indicator data by category shows outputs related to each program goal. Development of a composite performance rating system (CPRS) allows the indicator data to be combined to provide an easy-to-grasp overall performance measure across multiple program goals. The distribution of CPRS ratings within the portfolio gives program administrators a handle on the overall performance of the portfolio and an easy way to communicate that performance. At the same time, linking the composite ratings back to the individual case studies facilitates further investigation into the impact of funded projects.

The CPRS tool is still in a prototype development stage for application specifically to ATP. It is undergoing review and critique for possible improvements or extensions. At the same time, the CPRS has been used extensively to monitor project and portfolio performance during the intermediate period after project completion and before long-term benefits have had time to be realized and measured. It was used to rate ATP's first 50 completed projects and is being used to rate the next group of completed projects. ATP has used the CPRS tool to brief ATP oversight and advisory bodies, public policy analysts, evaluation groups, the broader S&T community, and the general public about ATP's performance.

There is considerable precedence for using composite scoring as a management tool. Composite rating systems have recently been developed or proposed for use by other federal agencies, international bodies, hospitals, and businesses, several examples of which are presented in the report. For example, a composite rating was developed to score the stability of financial institutions of variable size, location, and other characteristics to meet federal guidelines. In the prior cases examined, a composite rating system was developed to make complex information about multiple aspects of an issue more understandable to program administrators and other stakeholders in order to facilitate decision making.

A common characteristic of composite rating systems is the lack of an existing theoretical basis on which to base its development. The counterpart composite ratings examined were unavoidably ad hoc in nature, based on empirical experience rather than an existing theory or literature. Like the CPRS, the other rating systems examined relied heavily on expert judgment to select the indicator variables used for the composite measure and to assign weights to the indicator variables to determine how much each would count in the composite measure. This is a methodological limitation that may be reduced over time by further analysis and definition of underlying, functional relationships between alternative indicator metrics and the goals to which they relate.

The CPRS was developed after the first 50 mini case studies of completed ATP projects had been completed, and it was formulated specifically to use the data on outputs, outcomes, and outlook uniformly collected in those case studies. The available data were examined and those variables that appeared best to serve as indicators for each of three program goals were selected. Weights were assigned to the indicator variables according to expert judgment of the analyst in consultation with program administrators. The CPRS was designed to measure overall progress during the intermediate performance period toward accomplishing the following three goals: creating knowledge, disseminating the knowledge, and commercializing the technologies created from the knowledge base. It was not intended to provide a measure of long-run economic benefits. Given that the CPRS does not provide a measure of net benefits, and, in any case, project performance may change after the case-study data are collected, projects with similar CPRS ratings may differ in their long-term net economic benefits.

By weighting and combining progress indicators, a star rating was computed that provides a composite view of each project's progress overall during the interim period toward accomplishing program mission. From the individual project ratings a distribution of star ratings for the portfolio was computed to provide an overview of performance across the whole portfolio. By reducing a large amount of detail to a single symbolic rating for each project—0 to 4 stars—the CPRS conveys a snapshot of project performance. By reducing an even larger amount of often conflicting detail to a distribution of symbolic ratings across the portfolio—16% with 4 stars, 26% with 3 stars, 34% with 2 stars, and 24% with 1 or no stars—the CPRS conveys an immediate picture of portfolio performance. At the same time, all the details are preserved in the underlying project case studies to allow one to probe the specifics of each project.

The CPRS is consistent with the idea that there are varying degrees of project success that can be distinguished at a given time and signaled by indicator metrics. It is also consistent with the idea that cumulative project accomplishments at each stageæbeginning with knowledge creation, continuing with knowledge dissemination, and progressing further with commercializationærepresent an increasing degree of project success. Though surely an imperfect measure, the CPRS as formulated distinguishes among projects in ATP's portfolio in terms directly tied to the program's mission-driven goals and that are meaningful to its stakeholders. It provides a composite performance measure that is practical to construct and easy to understand and communicate.

The CPRS presented here has been custom designed for ATP's application and would not be suitable for direct transference to other public R&D programs with different goals and different time horizons. However, the CPRS concept and the eight-step framework of which it is an element can be adapted to fit other programs.

The CPRS was developed in 2000–2001 using the first 50 ATP projects completed in conjunction with the writing of status reports (mini-case studies) on these projects. Since then, ATP's Economic Assessment Office has computed CPRS ratings and published over 100 additional status reports. All completed status reports and CPRS ratings can be accessed on a searchable website (http://statusreports.atp.nist.gov/) and in the following publications:

- Performance of 50 Completed ATP Projects: Status Report — Number 2,[35] NIST SP 950–2, 2001
- Performance of 50 Completed ATP Projects: Status Report — Number 3, NIST SP 950–3, 2006

35. Status Report — Number 2 contains all projects from Status Report — Number 1 (38 projects), as well as the additional 12 projects used in the CPRS formulation.

References

Advanced Technology Program (2001). *Performance of 50 Completed ATP Projects*. Status Report Number 2, NIST SP 950–2. Gaithersburg, MD: National Institute of Standards and Technology.

Advanced Technology Program (Recent edition). *Proposal Preparation Kit*. Information about proposing to ATP is available on line at http://www.atp.nist.gov.

Branch, K., M. Peiffers, R. Ruegg, and R. Vallario (2001). *The Science Manager's Resource Guide to Case Studies*. DE-ACO6–76RL01830. Washington, DC: U.S. Department of Energy, Office of Science.

Dow Theory Forecast and Quadrix Stock-Rating System can be found at the Dow Theory Newsletter Website, http://www.dowtheory.com.

Geisler, E. (2000). *The Metrics of Science and Technology.* Westport, CT: Quorum Books, 2000, pp. 243–266.

KPMG, now Bearing Point (1997). *Methodology for Regulatory Test of Financial Responsibility Using Financial Ratios*. The report may be viewed at http://www.ed.gov/offices/OPE/PPI/finanrep.html.

Long, W. (1999). *Performance of Completed Projects,* Status Report Number 1. NIST SP 950–1. Gaithersburg, MD: National Institute of Standards and Technology.

Kilbridge, P., E. Welebob, and D. Classen (2001). *Overview of the Leapfrog Group Evaluation Tool for Computerized Physician Order Entry.* December 2001. The report may be viewed at http://www.fcg.com by clicking on the report cover.

Martin, S., D. Winfield, A. Kenyon, J. Farris, M. Bala, and T. Bingham (1998). *A Framework for Estimating the National Economic Benefits of ATP Funding of Medical Technologies; Preliminary Applications to Tissue Engineering Projects Funded from 1990 to 1996*. NIST GCR 97–737. Gaithersburg, MD: National Institute of Standards and Technology.

Powell, J. and K. Lellock (2000). *Development, Commercialization, and Diffusion of Enabling Technologies: Progress Report for Projects Funded 1993–1997*. NISTIR 6491. Gaithersburg, MD: National Institute of Standards and Technology.

Ruegg, R. (2003). *Benchmarking Evaluation of Public Science and Technology Programs in the United States, Canada, Israel, and Finland; Proceedings of the Impact Assessment Benchmarking Workshop*. Embassy of Finland, Washington, DC, September 25, 2002. Helsinki, Finland: Tekes—National Technology Agency of Finland, 2003.

Ruegg, R. and I. Feller (2003). *A Toolkit for Evaluating Public R&D Investment: Models, Methods, and Findings from ATP's First Decade.* GCR 02–842. Gaithersburg, MD: National Institute of Standards and Technology.

Smith, P. (2001). "Component Indicators of System Performance." Presentation at the OECD Conference, *Measuring UP; Improving Health Systems Performance in OECD Countries,* Ottawa, Canada, November 5–7, 2001.

U.S. Congress (1988). Omnibus Trade and Competitiveness Act of 1988 (P.L. 100–418).

U.S. Congress (1991). American Technology Preeminence Act of 1991 (P.L. 102–245).

U.S. General Accounting Office (USGAO) (1996). *Executive Guide, Effectively Implementing the Government Performance and Results Act.* GAO Report GGD-96–118. Washington, DC.

Yin, R. (1994). *Case Study Research—Design and Methods.* Applied Social Methods Series, vol. 5. Thousand Oaks, CA: SAGE Publications, 2nd edition.

About the Advanced Technology Program

The Advanced Technology Program (ATP) is a partnership between government and private industry to conduct high-risk research to develop enabling technologies that promise significant commercial payoffs and widespread benefits for the economy. ATP provides a mechanism for industry to extend its technological reach and push the envelope beyond what it otherwise would attempt.

Promising future technologies are the domain of ATP:

- Enabling or platform technologies essential to development of future new products, processes, or services across diverse application areas
- Technologies where challenging technical issues stand in the way of success
- Technologies that involve complex "systems" problems requiring a collaborative effort by multiple organizations
- Technologies that will remain undeveloped, or proceed too slowly to be competitive in global markets, in the absence of ATP support

ATP funds technical research, but does not fund product development—that is the responsibility of the company participants. ATP is industry driven, and is grounded in real-world needs. Company participants conceive, propose, co-fund, and execute all of the projects cost-shared by ATP. Most projects also include participation by universities and other nonprofit organizations.

Each project has specific goals, funding allocations, and completion dates established at the outset. All projects are selected in rigorous competitions that use peer review to identify those that score highest on technical and economic criteria. Single-company projects can have duration up to three years; joint venture projects involving two or more companies can have duration up to five years.

Small firms on single-company projects cover at least all indirect costs associated with the project. Large firms on single-company projects cover at least 60 percent of total project costs. Participants in joint venture projects cover at least half of total project costs. Companies of all sizes participate in ATP-funded projects. To date, nearly two out of three ATP project awards have gone to individual small businesses or to joint ventures led by a small business.

Contact ATP for more information:

- On the Internet: www.atp.nist.gov
- By e-mail: atp@nist.gov
- By phone: 1-800-ATP-FUND (1-800-287-3863)
- By writing: Advanced Technology Program, National Institute of Standards and Technology, 100 Bureau Drive, Stop 4701, Gaithersburg, MD 20899-4701

About the Author

Rosalie Ruegg, managing director of TIA Consulting, Inc., specializes in the economic assessment of new technologies. Prior positions include director of ATP's Economic Assessment Office, senior economist in NIST's Center for Applied Mathematics, and financial economist for the Federal Reserve System's Board of Governors. She has more than 60 publications, including an economics textbook; has served on editorial boards, most recently as economics editor of Macmillan's *Encyclopedia of Energy,* and on advisory committees. A former member of the Federal Senior Executive Service, she received both the Department of Commerce's Gold and Silver Medal Awards, and the Institute of Industrial Engineers' 2001 Wellington Award for contributions to the field of engineering economics. She holds degrees in economics from the Universities of North Carolina and Maryland, and an M.B.A. from The American University.