Best Practices in Assessment of Research and Development Organizations

ISBN
978-0-309-26626-0

90 pages
8 1/2 x 11
PAPERBACK (2012)

Panel for Review of Best Practices in Assessment of Research and Development Organizations; Laboratory Assessments Board; Division on Engineering and Physical Sciences; National Research Council

Best Practices in Assessment of Research and Development Organizations

Panel for Review of Best Practices in Assessment of
Research and Development Organizations

Laboratory Assessments Board

Division on Engineering and Physical Sciences

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the panel responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract No. SB134106Z011 between the National Academy of Sciences and the U.S. Department of Commerce. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the agency that provided support for the project.

International Standard Book Number-13: 978-0-309-26626-0
International Standard Book Number-10: 0-309-26626-2

Copies of this report are available from

Laboratory Assessments Board
Division on Engineering and Physical Sciences
National Research Council
500 Fifth Street, NW
Washington, DC 20001

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); http://www.nap.edu.

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

PANEL FOR REVIEW OF BEST PRACTICES IN ASSESSMENT OF RESEARCH AND DEVELOPMENT ORGANIZATIONS

JOHN W. LYONS, National Defense University, *Chair*
EDWARD A. BROWN, The MITRE Corporation
W. WARNER BURKE, Columbia University
ROSS B. COROTIS, University of Colorado at Boulder
WILLIAM W. CRAIG, Lawrence Livermore National Laboratory
C. WILLIAM GEAR, Princeton University
WESLEY L. HARRIS, Massachusetts Institute of Technology
ELENI KOUSVELARI, Sandia National Laboratories
BERNARD S. MEYERSON, IBM Corporation
ELSA REICHMANIS, Georgia Institute of Technology
JOEL M. SCHNUR, George Mason University
LYLE H. SCHWARTZ, Air Force Office of Scientific Research (retired)

Staff

JAMES P. MCGEE, Board Director
CY BUTNER, Senior Program Officer
LIZA HAMILTON, Administrative Coordinator
EVA LABRE, Program Associate

LABORATORY ASSESSMENTS BOARD

JOHN W. LYONS, National Defense University, *Chair*
CLAUDE R. CANIZARES, Massachusetts Institute of Technology
ROSS B. COROTIS, University of Colorado at Boulder
JOSEPH S. FRANCISCO, Purdue University
C. WILLIAM GEAR, Princeton University
HENRY J. HATCH, U.S. Army (retired)
LOUIS J. LANZEROTTI, New Jersey Institute of Technology
ELSA REICHMANIS, Georgia Institute of Technology
LYLE H. SCHWARTZ, Air Force Office of Scientific Research (retired)
CHARLES V. SHANK, Howard Hughes Medical Institute
DWIGHT C. STREIT, Northrop Grumman Space Technology

Staff

JAMES P. MCGEE, Board Director
CY BUTNER, Senior Program Officer
LIZA HAMILTON, Administrative Coordinator
EVA LABRE, Program Associate

Acknowledgment of Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for clarity, objectivity, and responsiveness to the charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

David E. Borth, Motorola, Inc. (retired)
Jennie S. Hwang, H-Technologies Group and Case Western Reserve University
Miriam E. John, Consultant
Cherry A. Murray, Harvard University
Richard R. Paul, Unites States Air Force (retired)
Julia M. Phillips, Sandia National Laboratories
Dwight C. Streit, University of California, Los Angeles

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, and they did not see the final draft of the report before its release. The review of this report was overseen by Joseph S. Francisco, Purdue University. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Contents

Summary

Research and development (R&D) organizations are operated by government, business, academe, and independent institutes. The success of their parent organizations is closely tied to the success of these R&D organizations. In this report, *organization* refers to an organization that performs research and/or development activities (often a laboratory), and *parent* refers to the superordinate organization of which the R&D organization is a part. When the organization under discussion is formally labeled a laboratory, it is referred to as such.

The question arises: How does one know whether an organization and its programs are achieving excellence in the best interests of its parent? Does the organization have an appropriate research staff, facilities, and equipment? Is it doing the right things at high levels of quality, relevance, and timeliness? Does it lead to successful new concepts, products, or processes that support the interests of its parent?

The management of the National Institute of Standards and Technology (NIST) asked the National Research Council (NRC) to study methods of assessing research and development organizations. To conduct the study, the NRC appointed the Panel for Review of Best Practices in Assessment of Research and Development Organizations. This report summarizes the findings of that panel.

The report offers assessment guidelines for senior management of organizations and of their parents. The report lists the major principles of assessment, noting that details will vary from one organization to another. It provides sufficient information to inform the design of assessments, but it does not prescribe precisely how to perform them, because different techniques are needed for different types of organizations.

Three key factors underpin the success of an R&D organization: (1) the mission of the organization and its alignment with that of the parent; (2) the relevance and impact of the organization's work; and (3) the resources provided to the organization, beginning with a high-quality staff and management. Other resources include its budget, facilities, and capital equipment.

Consideration of the alignment of the organization's mission and the relevance and impact of its work requires assessing the relationship that the laboratory and its parent have with their customers and stakeholders. Definitions of customer and stakeholder vary. A customer is often viewed as someone either in or outside of the organization who purchases products from the organization or its elements. A stakeholder may be viewed as an entity that can impact the organization's vision, mission, plans, or resources. Customers may differ from or be a subset of stakeholders. Although these definitions are arguable, the point remains that an effective assessment determines whether the organization has developed a clear and meaningful identification of its set of customers and stakeholders and a means for identifying and satisfying their needs.

THE CONTEXT OF EVALUATION

The context in which an organization is being evaluated relates first to the mission and vision of the parent organization. It is essential that the organization align its programs to be consistent with the parent's mission and vision.[1,2] Additionally, the organization may write its own mission and vision statements. (It is important to keep in mind, when discussing these missions, which is meant.) The output of an organization depends on the kind of work that it is commissioned to do. R&D organizations perform a variety of technical work. Some conduct fundamental, long-term research; some do applied research; others do developmental work; still others support technical efforts leading to production and marketing or to implementation of new processes. Some organizations do all of the above. Effective assessments are structured to take into account what the organization is aspiring to do.

Research and development can be examined by considering three phases of the R&D: the planning stage, ongoing research, and evaluation of the relevance and impact of the R&D activities. In the planning stage, prior to launching a project, an organization develops goals for the projects, selects strategies and tactics intended to reach these goals, identifies needed personnel, and lists methods, including metrics, to assist in evaluating progress. Planning is done and assessed in the context of the organization's mission. Assessing ongoing research, the most common subject of assessments, includes reviewing and evaluating the technical projects and considering the quality of the research staff and management, the facilities, and the capital equipment. An effective assessment compares the program to the parent's mission and vision. Relevance can be assessed by comparing the organization's portfolio with expressed needs of customers in terms of the substance of the work and of its priorities. Retrospective analyses of programs may be made at various times following the completion of research and development activities. Many of the same metrics used to evaluate work in progress are useful in examining an R&D program after its completion.

THE THREE MAIN ELEMENTS OF AN ASSESSMENT

A comprehensive assessment evaluates three elements: management, technical quality, and impact. Aspects of these three elements may overlap. For example, the quality of the workforce falls under both management actions and quality of the work. Quality of the work will also be covered in considering impacts. Determining the relevance of the work is a key role of management.

Assessing Management

Customers and Stakeholders

An effective assessment begins with considering how well management of the organization has identified its mission and vision as they are aligned with those of the parent. The assessment involves identification of the stakeholders and customers of the organization. A key question is how management stays in close contact with these groups and how well it responds to changing demands. For organizations that are focused on a limited number of customers, contacts can be made directly with the ultimate users of the results. For other

[1] National Research Council, 2002. *Future R&D Environments: A Report for the National Institute of Standards and Technology*. The National Academies Press, Washington, D.C.
[2] J. Sommerer, 2012. "Assessing R&D Organizations: Perspectives on a Venn Diagram." Presentation at the National Research Council Workshop on Best Practices in Assessment of Research and Development Oganizations, March 19, Washington, D.C.

organizations, such as NIST, there are too many customers for individual contacts. NIST's clientele includes the scientific and engineering communities; a variety of federal, state, and local organizations; and other countries. NIST of necessity deals through intermediary organizations such as trade groups, scientific and engineering societies, and aggregations of state and local government interests. It is important that an assessment of an organization include the efficacy of such interactions, using techniques such as polling and face-to-face meetings.

Resource Management

Any effective assessment is done in the context of the organization's mission. In order to satisfy its mission, an organization needs to be prepared to handle its current and future workload. This means that it will have a successful combination of the following:

- *R&D portfolio*—a collection of projects that are most likely to lead to successful accomplishment of the organization's mission;
- *Resources*—a workforce with an appropriate skill balance; the needed physical plant and equipment; and sufficient funding to enable accomplishment of the mission;
- *Organizational leadership and management structure appropriate to the mission*; and
- *Planning for the future*—the preparedness needed to ensure that the required resources will be in place as the mission evolves.

These elements are properly considered in context. Academic research focuses on generating new knowledge with relatively few mission objectives, whereas government and industrial research organizations have fairly clearly defined missions. Any effective management assessment also recognizes externally imposed limitations, including but not limited to regulatory and budgetary restrictions.

Portfolio

At all stages of R&D, it is important that the institution construct and manage its portfolio to maximize the probability of success. In basic research, it is important that the portfolio cover those areas that are likely to be important in the long term to achieving the mission, and that the assessors look at the portfolio and comment on whether there are areas that may be missing and whether there are areas that may be covered but not be relevant. In product development the areas are often well specified, but it is important to consider whether or not the correct set of technologies is being applied to achieving the desired results. As noted above in the discussion on the context of the evaluation, portfolios can be assessed during the planning phase, ongoing research phase, and retrospectively.

There are three elements to consider when assessing the quality of the research portfolio: (1) current projects and their relevance to the mission; (2) anticipation of opportunities; and (3) alignment of the planned future portfolio to mission, opportunities, and budget. Every R&D organization has some systematic way of listing its investments. At the extreme of offering little specificity is identification of funding per group, with descriptions of the group responsibilities and recent accomplishments. This format is most common at the more basic or fundamental end of the R&D spectrum. At the more applied extreme are examples in some industrial organizations in which each project is specified in great detail, including time lines and anticipated return on investment.

In many respects surveying the research portfolio of an organization or of a unit within that organization is straightforward. Individual projects are grouped under programs and evaluated in concert with stakeholders' and customers' needs and expectations. This process is often done through organized meetings. It may be a continuing process that includes formal oversight from stakeholders as well as outside reviewers and consultants. Industrial organizations may rely heavily on metrics involving financial return, whereas government organizations may focus more on delivering needed value to stakeholders consistent with mission statements.

Resources

Fulfilling the organization's mission requires a high-quality workforce with an appropriate mix of skills, an appropriate physical plant and laboratory equipment, and sufficient funding to accomplish the tasks.[3]

Managing the Workforce. The importance of the quality and expertise of people within an organization cannot be overemphasized. Of special significance for many R&D organizations is the composition within any particular division, laboratory, or project, of a staff with deep and creative technical capabilities. People with deep specialties but also broad perspectives and a history of varied assignments help prepare the organization for future assignments. At the heart of looking forward to the next generation of scientific and technological opportunities are the organization's scientists and engineers. Their knowledge of cutting-edge research is the starting point in all such efforts. To maintain and expand their knowledge, scientists and engineers require opportunities to attend scientific and technical meetings and to participate in the international community of scholars. Supportive management will also encourage the staff to think about next-generation efforts and reward them for that effort by bringing resources to bear on the most promising ideas.

It is important to assess the organization's policies and actions aimed at steadily building upon the sets of capabilities associated with individuals possessing both breadth of experience across multiple projects and depth in one or more systems and disciplines. An effective organization enables its staff to capture new skills as required for a given set of tasks at hand, while over the long term building the network required to make team members effective participants in global efforts to achieve the overall goals of the organization. To facilitate the creation of such capabilities, a diversity of personnel and work experience is vital. Effective assessments of an R&D organization include consideration of a diverse workforce whose contributions may affect and advance the R&D mission of an organization.

It is important that management continually plan for the future so that when the future arrives the laboratory is in a position to fulfill its mission. This means that the right resources and leadership must be in place when needed. The demographics of the workforce must be tracked so that there will be leaders in place as retirements and departures occur, and people with new skill sets must be recruited to be ready to deal with new technologies. An effective

[3] J. Lyons, 2012. *Reflections on Over Fifty Years in Research and Development: Some Lessons Learned*. National Defense University, Washington, D.C.

assessment considers the adequacy of this planning as well as plans for the necessary physical infrastructure to support the new skill sets and technologies.

Physical Resources. Technical facilities encompass the organization's physical space and how it is occupied. Does the workforce have what it needs to carry out the research program as identified or planned? An evaluation of facilities independent of clear understanding of program content is just as imprudent as an evaluation of program content without recognition of the need for appropriate facilities. Metrics concerning the facilities are not readily generalized, but it is important that the assessment of resources include capital equipment. It is important that management seek funding for purchases of modern items or for effective upgrades of existing equipment as well as maintenance of existing equipment. There is no established rule for how much of the budget management should allocate for equipment. The need for such funding depends on the nature of the work. Some work requires major purchases of equipment, but other work is not very dependent on expensive devices. Often, organizing equipment in the facilities to maximize utilization is desired, so individual "ownership" may need justification. "Home-built" equipment, properly identified and documented for the assessment team, is often a useful measure of a researcher's creativity.

Organizational Leadership and Management

An effective management structure will be consistent with the nature of the work. Basic research typically requires a very flat management structure, significant individual freedom in selection of research directions, and a management very receptive to suggestions (although projects involving very large experimental resources such as accelerators may require more structure). Product development typically requires a more hierarchical structure in order to ensure mission progress.

It is important that assessment of management include the effectiveness of the two-way communication between management and the workforce. How well does management explain the mission, vision, and strategy of the organization? How well does the management explain the importance of the work and why the work has been assigned to the organization? Does the management explain external changes that affect the organization? Is there a clear operational plan for executing the technical work? Does management provide copies of its planning documents to the staff?

It is important that assessors try to identify the culture of the organization, including how well the staff understand how "things are done here"; whether they feel that the organization is "a great place to work"; whether staff members are treated with dignity and respect; how well diversity is encouraged; and how conflict is surfaced and managed. It is also important to determine whether and how managers are noted for their ability to intercept and handle bureaucratic demands from above, thereby shielding research staff from administrative burdens.

Other important assessment items include whether an organization's leaders have experience matched to their assigned groups and experience in leading groups of professionals; whether there are programs to prepare staff members for future assignments involving more managerial functions; and whether training assignments, mentoring, and coaching are a part of personnel development.

Assessing Quality

An assessment of the quality of a research organization's work involves the consideration of a number of factors. Some can be measured quantitatively; others require more subjective judgments. Effective assessments will also include the quality of managers; the quality of research staff members; the output of the organization in terms of papers, patents, presentations, and handoffs to clients; and the adequacy of facilities and equipment. Some of these factors can be quantified in metrics; others require hearing presentations, walking through the organization, and speaking with staff and managers. Some of these assessments will also be made when looking at management and at impact.

The quality of the research staff is thought by many to be the most important factor in an organization's success. The assessment of the quality of staff is therefore most important. Some measures address the staff as a whole—for example, the percentage of Ph.D.'s or the number of postdoctoral fellows. Production by individual staff members includes the number of papers or patents per staff year, although these metrics do not really tell the quality of the individual—papers can be routine, and patents can be trivial and not used. However, assessing the quality of papers can be done by subject-matter experts on the assessment panels. The assessment of ongoing work is accomplished by hearing presentations by the individuals and visiting with them informally in their workplace or laboratory modules. In this way the overall capability of the laboratory staff can be estimated. An important assignment of the assessment panels, while touring the organization, is an evaluation of the state of the infrastructure—facilities, capital equipment, and support services.

Conducting the Assessment

The way that an assessment is done depends on what the nature of the organization is, the time frame for which the review applies, and who designs and manages the assessment. Assessments can be done within the organization or by outside parties. There is a trade-off between inside and outside evaluations. The inside assessors would have more detailed knowledge of the roles of the organization and the projects under review, but insiders may possess a bias with respect to the organization. External panels of independent experts would need to develop enough knowledge to make the assessment but would necessarily assess it with less intimate knowledge of the organization. Generally, there is more credibility attached to an independent external assessment. A fully independent assessment is arranged by and managed by an independent contractor. The appointment of assessment panel members with requisite expertise is crucial. It is essential that candidates for membership on assessment panels are required to present any biases and potential conflicts of interest to the contractor so that the appointment decision can take into consideration such potential conflicts.

Before the assessment is carried out, it is important that the panel's members receive briefing materials covering the background of the organization, including some history, a discussion of the parent organization, a number of quantitative measures (metrics) that the organization's management maintains, and any special charges to the assessors from the organization's management. It is essential that an assessment panel spend sufficient time visiting the organization to be able to perform an assessment at the desired scope. Normally, selection of the topics to be addressed is discussed with the management, but the decisions are generally best made in collaboration with panel members, guided by the panel chairs.

Outputs of the Organization

The results of an organization's work will normally be available for the assessment panels to review. These may be papers, presentations, or other means of conveying the nature of the technical work. The quality of the finished work is evident through the study of the documents and the discussion of them with those who did the research and, if possible, with the customers for it. It is important that evidence of the opinions of the customers be sought by either the panel chair or the managers at the organization. It is essential that the organization provide anecdotes of successful work by, for example, citing significant scientific advances (well-cited papers, awards, or other recognition), or, if appropriate, new products in the marketplace, new processes for producing the products, or new software in use by the technical communities.

Benchmarking

One commonly used technique of assessing quality is to compare one R&D organization with others judged to be at a high level of performance. Benchmarking is usually done with metrics, which have to be normalized to account for size and budget differences of the organizations examined. For example, one may cite the number of archival publications for each technical professional. Using percentages also avoids the problem of size differences—for example, the percentage of doctorates among the professional population. It is important to make comparisons among R&D organizations operating in similar contexts. Comparing an engineering research organization with an academic department would usually be inappropriate. A problem with benchmarking by metrics is that such assessments do not get at the effectiveness of the organization being assessed. There are examples of first-class organizations working in a parent organization that has failed to capitalize on the organization's breakthroughs. Nonetheless, benchmarking can be a useful addition to the assessment tool kit.

Assessing Impact

Measuring the impact of R&D activities is an important aspect of assessment. An insightful definition of impact was posed by William Banholzer of the Dow Chemical Company in his presentation to the National Research Council's workshop on Best Practices in Assessment of Research and Development Organizations: "What would not have happened if you did not exist, and how much would society have missed?" [4]

One looks to the customers and stakeholders for an evaluation of the impact of a research program. Supporting this evaluation, an organization will put in place and use on a regular basis a systematic process of outreach to this clientele. Polling by questionnaire and polling by interviews are alternatives. Sometimes impact can be assessed by talking to industrial and technical organizations that are able to represent individual companies or other groups. Holding

[4] National Research Council, 2012. *Best Practices in Assessment of Research and Development Organizations—Summary of a Workshop*. The National Academies Press, Washington, D.C., p. 10.

periodic sessions with clientele would seem to be helpful not only for analysis of impact but also for validating the current research program against the needs of clientele.

Meaningful evaluations also include analysis of completed R&D, by examining the recent past and at other times through retrospective analyses of more distant history. R&D organizations are part of a system or process that leads to products that both organizational management and stakeholders will ultimately use as the basis for judgments about the worth of the R&D. In attempts to project the future impact of current or proposed programs, the R&D organization is hampered by the inevitable fact that it may take many years, or even decades, before the full impacts of current programs are realized by other organizations. And when that eventually does come to pass, there have usually been so many different organizations involved in developing, engineering, producing, and fielding the end item that its identity with the research organization is lost. Regardless of how the story is formulated, the stakeholders' confidence in the organization and its management will be bolstered by demonstration that the decision making and processes of the present are comparable to, or better than, those of the past that led to measurable impacts. To tell this story properly, many organizations have had recourse to looking backward and tracing the consequences of R&D events long past.

With respect to applied research and product and process development, industry will appropriately focus on its return on investment (ROI) for the R&D. Feedback from both failures and successes may be communicated to stakeholders and used to modify future investments. Government organizations rarely have such a direct metric, and it is important to search for more information and a structure to communicate to their myriad stakeholders.

An example of this approach for learning about impact is Project Hindsight.[5] Carried out by the Department of Defense in the mid- to late 1960s, Project Hindsight was a study of the development of 22 different weapons systems drawn from across the military services. It involved more than 200 personnel over a period of approximately 6 years. For years afterward the observations and conclusions of Project Hindsight guided military R&D planning and decision making. In 2004, recognizing that much had changed in the intervening years, the U.S. Army commissioned a new study, Project Hindsight Revisited.[6]

The Department of Energy (DOE) utilized a similar retrospective analysis, with the assistance of the NRC, examining the impacts on energy-producing and energy-using industries of R&D programs executed by the DOE laboratories over the time period 1978-2000. The report summarizing the findings of the assessment makes the case in economic terms for an ROI that by itself could justify funding the research, while recognizing that societal impacts are far more difficult to measure and are not readily quantifiable.[7]

Companies, or even laboratories themselves, may commission histories. Sometimes a popular book describes developments in technical organization; examples are developments in the Bell Laboratories[8] or the General Electric laboratories.[9] Occasionally the history of an

---

[5] Office of the Director of Defense Research and Engineering (DDRE), 1969. *Project Hindsight: Final Report.* Office of the DDRE, Washington, D.C.

[6] J. Lyons, R. Chait, and D. Long, 2006. *Critical Technology Events in the Development of Selected Army Weapons Systems: Project Hindsight Revisited*. National Defense University, Washington, D.C.

[7] National Research Council, 2001. *Energy Research at DOE: Was It Worth It? Energy Efficiency and Fossil Energy Research 1978 to 2000.* National Academy Press, Washington, D.C.

[8] J. Gertner, 2012. *The Idea Factory: Bell Labs and the Great Age of American Innovation*. Penguin Press, New York, N.Y.

[9] B. Gorowitz, 1999. *The General Electric Story: A Heritage of Innovation 1876–1999*. Schenectady Museum, Schenectady, N.Y.

organization may appear in a biography of one of the founders; recently a biography of a founder of Apple described the accomplishments of that company.[10]

SOME QUESTIONS TO CONSIDER DURING ASSESSMENT

The following is a series of questions that a manager—either the organization's director or a responsible remember of the parent organization—can ask when considering carrying out an assessment of a research organization. The body of this report addresses these questions in detail and suggests some best practices. A set of these questions is presented here and can serve as guidelines—a kind of "tool kit"—for anyone considering performing or sponsoring an assessment of an R&D organization.

Assessing Management

Answers to the following questions will be useful in the assessment of organizational management:

- Does the organization's management understand its mission and its relationship to that of its parent? Does the vision statement of the organization align with that of the parent organization?
- Is there a long-range plan for implementing the strategy by specific technical programs?
- Does the organization have an explicit strategy for its work and for securing the necessary resources?
- Do the program plans reflect a model for balance—that is, amount of basic versus applied and development research, and short-, medium-, and long-term work?
- Does the organization have a clear champion within the parent organization?
- Does management have an aggressive recruiting plan with well-defined criteria for new hires? Is there a set of practices for retaining, promoting, and recognizing the staff?
- Does the organization have a process for forecasting likely future technical developments in areas appropriate to its mission?
- Does the organization's management have discretionary authority to invest in new programs on its own initiative? Does management solicit ideas from the staff for new work?
- Does management regularly assess facilities and equipment for adequacy? Does it have a fiscal plan for updating or replacing laboratory equipment?
- Is there a process for regularly reviewing the organization's research portfolio for its alignment with the mission?
- What is the management climate, and how does one assess it? Is there enough flexibility to work across organizational lines?
- How does the structure of the organization support its mission?

[10] W. Isaacson, 2011. *Steve Jobs*. Simon & Schuster, New York, N.Y.
.

- How much collaboration is there with outside organizations? How many staff exchanges are there?
- Does management have a well-defined process and criteria for determining what work is performed in-house versus what work is sponsored via grants, contracts, or other mechanisms with external entities?
- Does the management support a culture of creativity, diversity, and entrepreneurship?

Assessing the Quality of Scientific and Technical Work

Answers to the following questions will be useful in the assessment of the quality of an organization's technical work:

- Does the assessment include the quality of the staff, equipment, and facilities?
- Does the assessment include the nature of the research portfolio as to alignment with the mission and the balance in regard to basic, applied, and development work and short-, intermediate, and long-term research?
- Does the organization have a set of indicators that can serve as parameters when the time frame precludes immediate assessment? Does the organization benchmark itself against premier organizations?
- Who is the expected audience for the assessment?
- Is the review done by technical peers?
- What are the criteria for ensuring the credibility and validity of the assessment?
- What is the scope of the assessment? Does it include proposals for new work? Does it include assessment of completed work—internal review and authority to release a report, publications, patents, invited lectures, awards, and the like?
- Who designs and manages the assessment?

Assessing Relevance and Impact

Addressing the following questions will be useful in the assessment of an organization's relevance and impact:

- Does the organization have a process for identifying its stakeholders and customers?
- Does it have a regular process for reviewing its programs and plans with its stakeholders?
- Does the organization have a process for learning of its customers' current and likely future needs and expectations for the organization?
- Does the organization have an explicit process for tracking the utilization of its results (e.g., is transition to the next R&D stage actively managed and measured)?
- Does it have a formal program for recording the history of its work from concept to final utility or impact?
- Does the organization have a program to conduct retrospective studies of its earlier work?

1

Introduction

Research and development (R&D) organizations are operated by government, business, academe, and independent institutes. The success of their parent organizations is closely tied to the success of these R&D organizations. In this report, *organization* refers to an organization that performs research and/or development activities (often a laboratory), and *parent* refers to the superordinate organization of which the R&D organization is a part. Where the organization under discussion is formally labeled a laboratory, it is referred to as such. The questions arise: How does one know whether the organization and its programs are achieving excellence in the best interests of its parent organization? Does it have an appropriate research staff, facilities, and equipment? Is it doing the right things at high levels of quality, relevance, and timeliness? Does it lead to successful new concepts, products, or processes that support its mission and the interests of its parent?

How does one assess a research organization to answer these questions? This report offers assessment guidelines to senior management of R&D organizations and of their parent organizations and other stakeholders. The report lists the major principles of assessment, noting that details will vary from one organization to another. The report provides sufficient information to inform the design of assessments, but it does not attempt to prescribe precisely how to perform assessments.

CALL FOR THE STUDY BY THE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

The National Institute of Standards and Technology (NIST), a major, multiprogram government organization, promotes U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve the quality of life for all Americans. Over its history there has always been at NIST the belief that a strong peer assessment of its programs is a critical aspect of ensuring that NIST is effectively fulfilling its mission. This assessment has traditionally been provided by the National Research Council (NRC) of the National Academies. Since 1959, the NRC has provided assessments of the technical quality of the scientific and technical work performed at NIST (formerly the National Bureau of Standards) laboratories.

In October 1, 2010, NIST implemented its first major reorganization in over two decades. The primary objective of the reorganization was to sharpen the focus of NIST's programs on their respective missions and to optimize their ability to deliver both the cutting-edge research and the related services critical to the U.S. economy. The most notable impact of the NIST reorganization is its effect on the NIST laboratory programs: The reorganization shifted the NIST laboratories from being activity or discipline-based organizations to being mission- and outcome-oriented organizations. The reorganization condensed the number of NIST laboratories and focused them along the distinct NIST mission lines of metrology, technology, and the provision of unique user facilities. This reorganization prompted the need for a reevaluation of the NIST processes for organizational assessment, to ensure that NIST is positioned to have the benefit of relevant feedback relating to all of its programs and mission areas.

NIST's Visiting Committee on Advanced Technology (VCAT), an external advisory board, highlighted in 2010 the need for NIST to reevaluate its assessment and review processes. The VCAT *2010 Annual Report* contained the following recommendation relevant to program assessment:[1]

> Following the reorganization of NIST into mission-focused laboratories, NIST should develop a comprehensive assessment program that includes effective peer review of scientific quality, [and] customer satisfaction for measurement services and for effectiveness in meeting the needs of the particular measurement program.

STATEMENT OF TASK

In 2011, NIST charged the National Research Council to conduct a study of best practices in assessing R&D organizations and to prepare a report providing NIST with information on peer-review and performance-evaluation systems. The NRC formed the Panel for Review of Best Practices in Assessment of Research and Development Organizations to carry out this study. The Panel was charged by the Director of NIST to consider means of assessing specific organizational aspects in a manner that satisfies the requirements of NIST to perform effective assessments but also identifies assessment methods that can be applied selectively to other R&D organizations. The organizational aspects are:

- Technical merit and quality of the science and engineering work,
- The adequacy of the resources available to support high-quality work,
- The effectiveness of the agency's delivery of the services and products required to fulfill its goals and mission and to address the needs of its customers,
- The degree to which the agency's current and planned R&D portfolio supports its mission,
- The elements of technical management that affect the quality of the work,
- The extent to which the agency is accomplishing the impact it intends, and
- The agency's flexibility to respond to changing economic, political, social, and technological contexts.

In this report, an R&D organization is readily understood to consist of a group of scientists, engineers, and support staff, with appropriate facilities and equipment, working to accomplish some stated mission goal(s). An R&D organization could be the Department of Defense (DOD) laboratories system, or it could be a single laboratory, or even a directorate, division, or group within a laboratory. It could be the entire array of Department of Energy (DOE) National Laboratories or any one of these. It could be a laboratory within an industrial organization. It could be an organized research center at a university, or a research center involving collaboration among several universities and industries. Even more broadly, an R&D

[1] Visiting Committee on Advanced Technology, 2010. *2010 Annual Report of the Visiting Committee on Advanced Technology of the National Institute of Standards and Technology*. U.S. Department of Commerce, Washington, D.C., p. 7.

organization might include such organizations as the National Science Foundation or the DOE funding offices.

THE PANEL'S STUDY PROCESS

To address the study charge, the NRC appointed 12 expert volunteers to the panel. These volunteers, whose biographical sketches appear in Appendix A, represent expertise across research, development, and management activities in the following organizational sectors: the federal government, industry, academia, and national laboratories. Several members of the panel had previously engaged in activities as members of NRC committees that assessed R&D activities at NIST and at other federal agencies.

The panel gathered inputs from relevant literature and from formal discussions at panel meetings with representatives of the following organizations: NIST, the Defense Laboratory Office, the Office of Science and Technology Policy, the Office of the Director of National Intelligence, and the IBM Corporation. On March 19, 2012, the panel steered a workshop at the Keck Center of the National Academies in Washington, D.C., at which representatives from the following organizations provided formal presentations on assessment practices: Johns Hopkins University Applied Physics Laboratory, the Association of Public and Land Grant Universities, former congressional staff, Sandia National Laboratories, The Dow Chemical Company, and Microsoft Research; a presentation was also made by the former Assistant Secretary of the Army for Research, Development, and Acquisition. The workshop also garnered inputs through small group discussions with approximately 100 representatives from organizations of the federal government, industry, academia, and national laboratories. A summary of the workshop was published in September 2012.[2]

The panel conducted four meetings at which it considered the literature reviewed, the presentations made and the discussions conducted, and information derived from the extensive personal experience and expertise of the panel members, and the panel drafted this report summarizing their findings. The report underwent rigorous review by a committee of experts appointed by the National Research Council. When that review committee expressed satisfaction with the responses of the Panel to their review comments, and when the report was subsequently approved by the National Research Council, it was made available to the public.

ORGANIZATION OF THE REPORT

This report is organized to reflect the elements in the panel's statement of task. The body of the report consists of the Summary and six chapters. This first chapter describes the formation and study process of the panel. Chapter 2 presents issues that formed the framework within which the panel examined assessment practices. Chapters 3, 4, and 5 present assessment considerations applicable to the three foci of assessment considered fundamental by the panel: management practices, the technical quality of the R&D, and the impact of the R&D. Chapter 6 presents a summary of guidelines for consideration during an assessment.

The report also includes 12 appendixes. In addition to the biographical sketches of the panel members (Appendix A), they present the following: a summary of a workshop presentation highlighting the importance of assessing the alignment between an organization's vision and its

[2] National Research Council, 2012. *Best Practices in Assessment of Research and Development Organizations—Summary of a Workshop*. The National Academies Press, Washington, D.C.

people (Appendix B); a discussion of considerations pertaining to the validation of assessments (Appendix C); an example of the effective application of peer advice during the planning phase of a research project at the Army Research Laboratory (ARL; Appendix D); a discussion of relevant statutes and requirements documents that may be involved in some types of assessments (Appendix E); a description of the process applied by the ARL to assess the relevance of its R&D (Appendix F); a discussion of stakeholder relationships for laboratories at the DOD, DOE, and NIST (Appendix G); a list of questions pertaining to the assessment of leadership and management (Appendix H); examples of recent cross-organizational assessments (Appendix I); a summary of assessment processes at NIST, ARL, and Sandia National Laboratories (Appendix J); a summary of assessment processes at selected other government organizations involving peer review of technical quality (Appendix K); and a set of assessment metrics and criteria applied by NRC panels that review the ARL.

2

Framework for Examination of Assessment Processes

CONTEXTUAL CONSIDERATIONS

As a prerequisite for effective assessment, the context within which an assessment of an R&D organization is conducted will be clearly elucidated before the assessment strategy is developed and applied. There is no single formula that works for all organizations, and there is no single way to compare one organization to another.[1,2] Assessments can have many purposes, and it is crucial to identify clearly at the outset the purposes of a particular assessment. A self-assessment can be considerably different from one commissioned by a stakeholder and from one performed by an external team of assessors, even though the basic approaches may be similar. Different types of R&D organizations require different approaches to assessment, although there are common basic paradigms to guide the process. Effective assessments will also reflect cognizance of the time horizon identified for the various characteristics being evaluated.

Metrics and quantitative analyses can be valuable tools in assessments of the technical quality and preparedness of an organization, whereas qualitative findings, informed by quantitative metrics, are more valuable for organizational decision makers.[3]

Elements of an R&D Organization

The context of an assessment includes an articulation of the following elements of the organization: vision, mission, strategic plan, operations plan, and resources.

The vision statement of an R&D organization defines what the organization wants to be and to become and/or the impact that it seeks to have on its stakeholders from a long-term perspective. The vision statement may be inspirational or even passionate. It is usually influenced by the organization's interpretation of its capabilities for impacting its stakeholders, customers, clients, and funding sources.

A mission statement defines the purpose of the R&D organization. It may serve as a guide for action by the organization, a definition of its goals, a high-level statement of the path by which the organization may reach its goals, and a decision-making guide. A mission statement may be influenced by the organization's principal stakeholders, customers, clients, and resources. It may also define how an organization provides value to its principal stakeholders, customers, and clients.

An R&D organization's strategic plan outlines what the organization does, for whom the organization executes its work, and how the organization plans to excel in executing its work. A strategic plan for any organization is best viewed as a dynamic process defined within a time

[1] J. Turner, 2010. *Best Practices in Merit Review*. Association of Public Land Grant Universities, Washington, D.C.
[2] M. Kennerly and A. Neely, 2002. A framework of the factors affecting the evolution of performance measurement systems. *International Journal of Operations and Production Management* 22 (11):1222-1245.
[3] J. Stephen Rottler, 2012. "Assessing Sandia Research." Presentation at the National Research Council's Workshop on Best Practices in Assessment of Research and Development Organizations, March 19, 2012, Washington, D.C.

frame typically identified as 3 to 5 to 10 years. A meaningful strategic plan is consistent with the mission statement of the organization. An effective strategic plan defines in detail the path by which the organization will attain its goals.

An operations plan includes the specific functions required to execute the strategic plan. The operations plan may focus on administrative, managerial, and production processes within the organization that enhance efficiency and effectiveness. An effective operations plan responds to gaps existing between resources and needs, including personnel and facilities. An effective operations plan is situation-sensitive. The operations plan is linked to the business of the parent organization.

The resources of an R&D organization include its personnel, facilities, finances, real estate, and the distribution and interrelationship of these elements. Resources drive the operations plan that binds the strategic plan that enables the mission within the vision.

Appendix B presents a summary of the presentation by Dr. John Sommerer, Head, Space Sector, and Johns Hopkins University Gilman Scholar, Johns Hopkins University Applied Physics Laboratory, delivered March 19, 2012, at the National Research Council's Workshop on Best Practices in Assessment of Research and Development organizations.[4] The presentation highlighted the importance of alignment between an organization's vision and its people, especially in an R&D laboratory.

So that an assessment can be appropriately tailored to achieve its purpose for a given organization, it is important to consider the factors relating to different types of R&D organizations, detailed below.

Types of Organization

In setting the context for an assessment, it is helpful to identify four types of R&D organizations (detailed discussion of the different forms of impact expected of these organizational types is presented in Chapter 5). Although no organization is a perfect match with a single type (e.g., federally funded research and development centers [FFRDCs]), the following general statements can be made:

1. *Mission-specific organizations (e.g., National Institute of Standards and Technology [NIST], Army Research Laboratory [ARL])*: The mission is clearly defined by the stakeholder, who is often responsible for commissioning external assessments that supplement the organization's own self-assessment practices.
2. *Industrial organizations (e.g., IBM, Microsoft Research, Dow Chemical Company), research institutes, and contract organizations:* The mission is clearly defined by the organization, and assessment is typically internal.
3. *Product-driven organizations (e.g., National laboratories [e.g., Sandia National Laboratories, Los Alamos National Laboratory):* The key missions are defined by a stakeholder, with considerable discretion available to the organization's management to define or seek new mission space. These R&D organizations are typically subject to all types of assessments.

[4] National Research Council, 2012. *Best Practices in Assessment of Research and Development Organizations: Summary of a Workshop*. The National Academies Press, Washington, D.C.

4. *University research organizations:* Basic research is typically part of the core mission; assessment is typically self-commissioned, conducted by external peers and focused on the longer term.

Types of Assessment

Assessments can be grouped in three general categories:

1. Self-assessment typically looks at the effectiveness of the organization's effort to improve the quality of its workforce and facilities, its preparedness to respond to current and future mission needs, and impact, as measured, for example, by return on investment.
2. Organization-commissioned external assessment is set by the organization; it usually looks at the quality of the work being performed and the strategies for maintaining and developing new core capabilities, as well as the impact of the organization on the broader community.
3. Independent external assessments are commissioned by and report to a stakeholder. The stakeholder sets the context, and frequently the assessment is focused on the impact and the return on investment.

TIMESCALES AND MULTIDIMENSIONALITY OF R&D ORGANIZATIONS

Research and development constitute a multidimensional process. An effective assessment includes consideration of all of these dimensions in order to provide a complete and comprehensive approach that enables valid, meaningful, and useful assessment of R&D organizations.

The timescale for the assessment may be short (1 to 3 years), medium (3 to 7 years), or long (7 years or longer). Typically, assessments involving shorter timescales focus more on the research process than on the research results. The nature of an R&D organization may reflect the sector within which it operates: mission-specific (generally government), industrial, national laboratories, or academic. The R&D performed in these different sectors may be done for quite different reasons, and assessment criteria may be different for these four settings.

The stages of R&D may be characterized as basic research, applied research, advanced technology development, preproduction, and, at times, production and product fielding. As illustrated in Figure 2-1, different considerations may apply to the assessment of different stages of R&D. Characteristics of the organization include quality, relevance, productivity, and impact as well as the characteristics of its management.

Assessment measures and criteria may be qualitative, quantitative, or anecdotal. Although the traditional demand has been for quantitative metrics for assessment, many characteristics of an organization are not well suited to yielding countable indicators. An assessment may be conducted at the levels of project/task, program, organizational element, laboratory, or overall organization. Collaborations within the organization and extramural collaborations may also be considered. The audience for the assessment findings may include

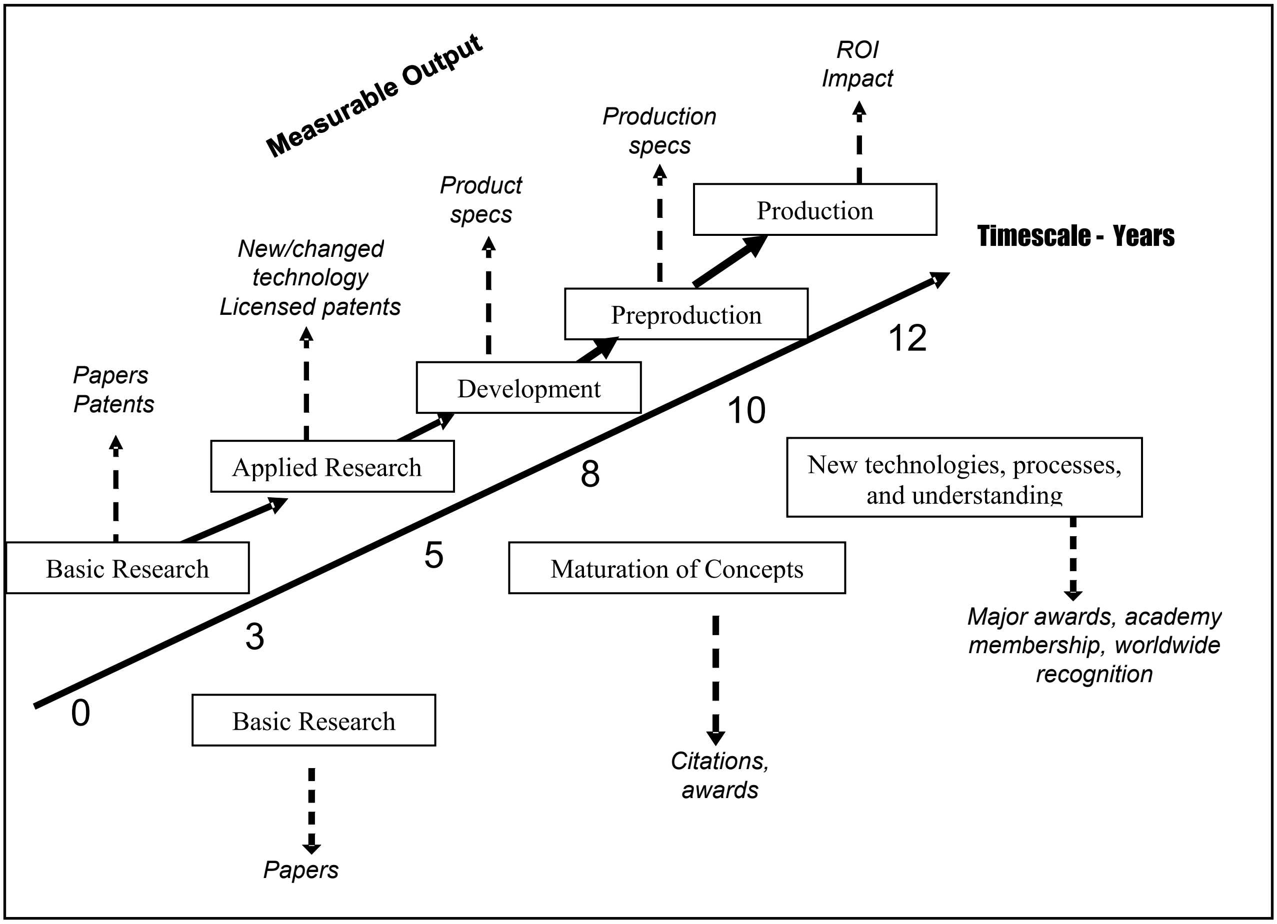

FIGURE 2-1 Simplified representation of the flow of the research and development (R&D) process. Applied activities are presented above the timescale, and basic activities are presented below the timescale. At different times during the R&D process, different assessment methods and measures may be applied. Notes: (a) In practice, the R&D process is not linear, but, rather, involves numerous feedback loops; (b) in practice, the R&D process is continuous; (c) distinctions between basic and applied activities are often blurred; and (d) timescales may vary considerably across different types of organization.

customers, stakeholders, users of an organization's products or services, internal management, the scientific and technical community, the general society, or other interested groups.

Figure 2-1 provides for both applied activities (above the timescale) and basic activities (below the timescale) a graphic representation of a simplified (linear) model of the R&D process (in practice, feedback loops connect the stages in the process). Figure 2-1 illustrates (without proposing an extensive set of metrics) that at each stage in the R&D process, different metrics can be considered for an assessment. The timescale indicated depends on the technical area considered.

For all timescales the timeline for innovation spans a continuum of R&D projects, from conception to implementation and finally to sunset. Human and capital resource requirements, and methods for the assignment of such resources, will likely vary greatly over this timeline. Similarly, the time lag from investment to return will vary as a function of program complexity, levels of funding and commitment, and other factors. An effective assessment of a given organization includes exploration of programs at varying states within their life cycles, because the processes and the outcomes of those processes at each stage are likely to differ. It is important that the efficacy of program selection at the front end of the R&D process be considered on the basis of long-term outcomes, but also that the efficacy of program de-selection be considered, which appropriately occurs at the early phase of projects showing little likelihood of success. Learning also comes from the commercialization of previous research. It is important to cover this landscape thoroughly, because needs and processes will evolve, along with the manner of assessment.[5] An effective assessment examines all elements of the organization, with time-appropriate quantitative and qualitative metrics and criteria, in terms of whether they are being integrated effectively. For each timescale appropriate guidelines will be defined in order to assess the quality, impact, and management of the activities of the organization. This includes an overall evaluation of the research portfolio of the organization.

The assessment of the human and capital resources, workforce development, and quality and relevance of the portfolio will vary for the different timescales defined above.

Short-term assessments (nominally 1-3 years) include factors such as the quality of personnel, customer interactions, and cross-organizational interactions. Quantitative metrics for this timescale include publications, patents, recruiting quality, retention, awards, extramural funding, and partnerships.

Midterm assessments (nominally 3-7 years) include personnel, new product introductions, publications, patents, citations, awards, funded programs, and customer satisfaction. Questions to be addressed include program implementation, project follow-through, financial metrics, and personnel growth and advancement.

Long-term assessments (nominally longer than 7 years) require a retrospective review of the major successes and failures of the organization. They include, where possible, the impact of projects that were launched at least a decade earlier. Questions to be addressed for long-term evaluations include whether programs transitioned from the organization are considered best in class, whether programs have met their financial commitments, whether assets have become or have led to the creation of a self-sustaining organization, and whether client partnerships are maintained.

FOCI OF ASSESSMENT

A comprehensive assessment addresses three key factors: (1) management practices (including the maintaining of preparedness to address future challenges), (2) technical quality of the R&D work and products, and (3) impact of the R&D and its products. Figure 2-2 illustrates sample assessment indicators for interactions between technical quality and management (referred to as “preparedness” in Figure 2-2 based on the management function of keeping the organization prepared to respond to changing demands) for four organizational types (academic, mission-driven, industrial, and product-driven). Preparedness is defined as the actions taken by the organization to identify and maintain the resources and strategies necessary to respond flexibly to future challenges.

[5] U.S. Department of Health and Human Services Office of the Inspector General, 2012. *FY 2012 Online Performance Appendix*. U.S. Department of Health and Human Services, Washington, D.C.

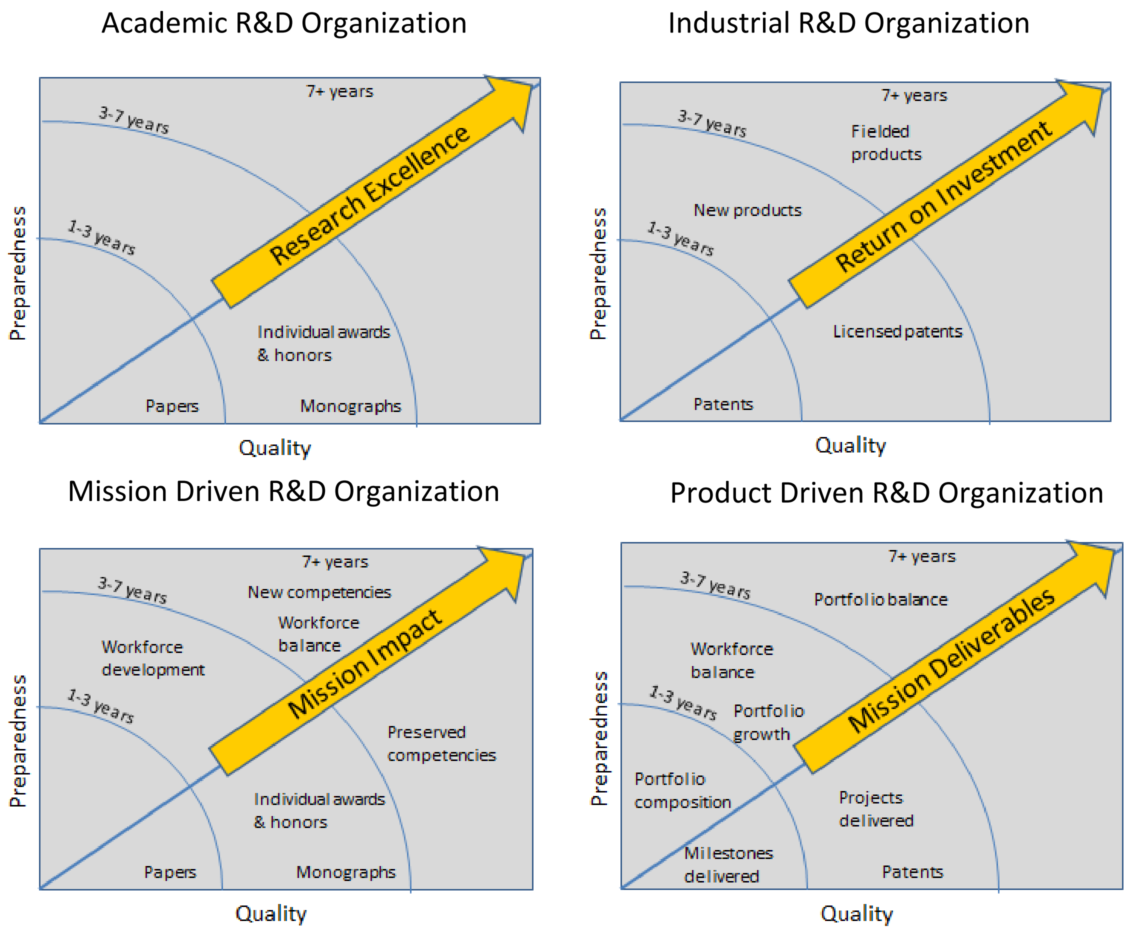

FIGURE 2-2 Quality and management preparedness are the key axes that drive an assessment of impact, relevance, or return on investment (impact is always dependent on context). Notes: (a) Metrics, outputs, and outcomes included in the figure are illustrative, not intended as an exhaustive set; (b) the time frames identified for outputs and outcomes are approximate.

PHASES OF ASSESSMENT

A research institution may be examined in three phases of the R&D: (1) in the planning stage, (2) during ongoing research, and (3) when the work is completed. In the planning stage, prior to the initiation of research, an organization develops goals, selects strategies and tactics intended to reach those goals, identifies personnel and/or workforce development pathways, and lists metrics for evaluating progress. Planning is done in the context of the organization's mission and may entail everything from broad capability development to specific, targeted development efforts. Appendix D provides an example of the effective application of peer advice during the planning phase of a research project at the Army Research Laboratory (ARL). The example describes how a set of experts selected from a pool of individuals familiar with the ARL contributed to effective refinement of a request for proposal for two significant ARL programs involving consortia that would examine multiscale modeling of materials relevant to Army R&D.

Ongoing research is the most common subject of review by processes external to the performing organization. The technical content of the R&D being performed is usually the subject of most attention, but an effective assessment also considers all of the context elements identified in the planning stage framework. Retrospective analyses of programs may be made at various times following the completion of research and/or development activities. Many of the same metrics used to evaluate work in progress are useful in examining an R&D program after its completion. Publications and patent awards to individuals and groups may be indicative of quality, particularly shortly after completion. Evidence of technology transition within the larger research, development, testing and evaluation, and fielding or commercialization system is another metric, although documentation of this metric is more readily obtained after time has passed. Reviews of economic impact and/or increased capability in a military system are examples of retrospective analysis that requires formal study by professionals, sometimes many years following the technical work of the organization. Such studies may be expensive and are most likely to be done on a case-by-case basis that tends to emphasize successes. Nevertheless, such anecdotal reviews after sufficient time has elapsed continue to represent the best evidence for later judgments of the effectiveness of the program.

Findings from assessments from each of the phases provide information to organizational decision makers, who are responsible for maintaining the preparedness of the organization to identify and respond to ongoing and future challenges.

After assessment findings are communicated to the organization, it is important to validate those findings—to assess the assessment itself. Appendix C provides a discussion of considerations pertaining to the validation of assessments, including discussion of various types of validity and reliability of assessment findings, factors relating to the efficiency of an assessment, and evaluation of the impacts of an assessment.

WHAT CONSTITUTES AN EFFECTIVE R&D ORGANIZATION?

Any assessment implicitly assumes what the attributes are that characterize an effective R&D organization. The attributes listed below were suggested in a report to the Secretary of Defense as part of the Base Realignment and Closure decision-making process pertaining to the Defense laboratories in 1991.[6] The list does not address the accomplishments of the past or offer

[6] Federal Advisory Commission on Consolidation and Conversion of Defense Research and Development Laboratories, 1991. *Report to the Secretary of Defense*. U.S. Department of Defense, Washington, D.C.

a forecast of potential results; however, a positive assessment against these attributes was then considered the best possible indicator of the eventual impact and relevance of the organization and the R&D that it performed.

- A clear and substantive mission,
- A critical mass of assigned work,
- A highly competent and dedicated workforce,
- An inspired, empowered, highly qualified leadership,
- State-of-the-art facilities and equipment,
- An effective two-way partnership with customers,
- A strong foundation in research,
- Management authority and flexibility, and
- A strong linkage to universities, industry, research institutes, and government organizations.

Attributes of effective organizations are also proposed in a National Research Council report assessing Department of Defense basic research,[7] the NIST Baldridge criteria for performance excellence,[8] and reports by Jordan and Binkley that suggest numerous attributes of effective R&D organizations.[9,10]

SUMMARY OF ASSESSMENT FRAMEWORK CONSIDERATIONS

An effective assessment will include an evaluation of an organization's management, the quality of its research and development activities, the impact of the effort, and associated interrelationships. Both qualitative and quantitative measures are required to assess management, quality, and impact. The context of the assessment is critically important in designing and carrying out an assessment and will differ depending on the mission and type of organization. Assessments are generally self-assessments, commissioned external assessments requested by the organizations themselves, or independent external assessments commissioned by and reported to stakeholders. Effective assessment of a research organization requires measurements over multiple timescales.

---

[7] National Research Council, 2005. *Assessment of Department of Defense Basic Research*. The National Academies Press, Washington, D.C.
[8] National Institute of Standards and Technology, 2011. *NIST Baldrige Program: 2011–2012 Criteria for Performance Excellence*. National Institute of Standards and Technology, Gaithersburg, Md.
[9] G. Jordan and J. Binkley, 1999. *Attributes of a Research Environment That Contribute to Excellent Research and Development*. Sandia Report SAND99-8519. Sandia National Laboratories, Albuquerque, N. Mex.
[10] G. Jordan, L. Streit, and J. Binkley, 2003. Assessing and improving the effectiveness of national research laboratories. *IEEE Transactions on Engineering Management* 50(2).

3

Assessing Management

Management of a research and development (R&D) organization is principally charged with two key tasks: establishing the vision and strategic plan for the organization and ensuring the preparedness of the organization to meet current commitments and future opportunities.[1] In addition to supporting these management tasks, assessments may, for some organizations, serve the additional function of fulfilling statutory and other requirements. Appendix E provides a discussion of such requirements for U.S. government agencies and U.S. government funds, including the Government Performance and Results Act of 1993 (GPRA), P.L. 103-62; the GPRA Modernization Act of 2010 (GPRAMA), P.L. 111-352; and the Office of Science and Technology Policy (OSTP) and Office of Management and Budget (OMB) *FY—Administration Research and Development Budget Priorities* document, which generally specifies that R&D programs will be assessed for quality, relevance, and performance, and indicates that budget decisions will be made based on these assessments.

.

CUSTOMERS AND STAKEHOLDERS

As part of an effective assessment process, the organization identifies its stakeholders and customers and ensures that their needs are addressed to their satisfaction. Definitions of "customer" and "stakeholder" vary. A customer is often viewed as someone, within or outside the organization, who purchases products from the organization or its elements. A stakeholder may be viewed as an entity that can impact the organization's vision, mission, plans, or resources. Customers may differ from or be a subset of stakeholders. Definitions are arguable, but the point remains that an effective assessment determines whether the organization has developed a clear and meaningful identification of its set of customers and stakeholders and the means for identifying and satisfying their needs.

Planning its research portfolio requires that the organization obtain inputs from stakeholders and customers. Reaching these individuals is important but in some cases may be difficult—especially the task of reaching customers. Sometimes a means of aggregating customers is available. For industry, trade associations can be asked to survey their members. Likewise, some scientific or engineering societies may have formal means of collecting information about the needs of their members. In another example, the state organizations concerned with accuracy in weights and measures are brought together in the National Conference on Weights and Measures (NCWM). The NCWM holds regular meetings during which problems facing the community are brought into focus. For organizations with very broad scope, such aggregation of customers and stakeholders is essential. For example, farmers are aggregated in various ways to interact with state and national technical organizations. One of these, the U.S. Department of Agriculture (USDA) Extension Service, applies very good

[1] RAND Corporation. 2012. *Improving Army Basic Research: Report of an Expert Panel on the Future of Army Laboratories*. RAND Arroyo Center, Santa Monica, Calif.

processes for learning about problems in the field and for transitioning research findings to the farmer. In some other government organizations, the customers are focused narrowly and are easy to reach more directly. For the Department of Defense (DOD) the warfighters are represented by the training schools and the developers and keepers of military doctrine. The military organizations can communicate with these groups and can actually rotate their staff to and from these customer organizations. Holding regular planning meetings with its customers helps to ensure that the organization's programs are relevant.

Effective processes for addressing customer and stakeholder satisfaction include the following:

- A systematic process to identify all stakeholders and make their identities part of the public record;
- The identification of individual(s) within the organization to be assigned the responsibility for interaction with each of the stakeholders or with organizations representing the stakeholders (e.g., industrial consortia or associations);
- A well-defined, open process for sharing information with stakeholders at all stages of R&D, including planning, execution, and delivery of results; and
- On a regular basis, internal review by management of this system of stakeholder interaction, and at intervals, review by outsiders using standing or ad hoc committees of expert advisers.

Appendix F describes the history of the process applied by the Army Research Laboratory (ARL) to assess the relevance of its R&D. This case study demonstrates an approach to identifying and responding to the needs of an organization's stakeholders. Appendix G provides a discussion of stakeholder relationships for laboratories at the DOD, the Department of Energy (DOE), and NIST.

RESOURCE MANAGEMENT

An effective assessment is conducted in the context of the organization's mission. In order to satisfy its mission, an organization will seek to be prepared to handle its current and future workload, and this will have a successful combination of the following elements:

- *R&D portfolio*—a collection of projects that are most likely to lead to the successful accomplishment of the organization's mission;
- *Resources*—a workforce with an appropriate skill balance; the needed physical plant and equipment; and sufficient funding to enable the mission;
- *Organizational leadership and management structure appropriate to the mission*; and
- *Planning for the future*—the preparedness needed to ensure that the required resources will be in place as the mission evolves.

These elements are properly considered in context. Academic research focuses on generating new knowledge with relatively few mission objectives, whereas government and industrial research organizations have fairly clearly defined missions. An effective management assessment also recognizes externally imposed limitations, including but not limited to regulatory and budgetary restrictions.

Portfolio

At all stages of R&D, it is important that the institution construct and manage its portfolio to maximize the probability of success. In basic research, it is important that the portfolio cover those areas that are likely to be important to the ultimate mission and that the assessors look at the portfolio and indicate whether there are areas that may be missing and whether there are areas covered that may not be very relevant. (These are sometimes very subjective evaluations.) In the product development portfolio, the areas are often well specified, but it is important to consider whether or not the correct technologies are being evaluated for possible use in a future product. As explained above, portfolios can be assessed during the planning phase, ongoing research phase, and retrospectively.

Three definable elements to consider in assessing the quality of a technical portfolio are (1) current projects and their relevance to the mission; (2) anticipation of opportunities; and (3) alignment of the planned future portfolio to mission, opportunities, and budget.

Generally an R&D organization has some way of cataloging its research portfolio. At the extreme of little specificity is funding per group with descriptions of the group responsibilities and recent accomplishments. This format is most common at the more basic or fundamental end of the R&D spectrum. At the more applied extreme are examples in some industrial organizations in which each project is specified in great detail, including timelines and anticipated return on investment.

In many respects surveying the research portfolio of a single organization (or a unit in that organization) is a straightforward process. Individual projects are grouped under programs and evaluated in concert with stakeholders' needs and expectations. This process is often done through organized meetings. It may be a continuing process that includes formal oversight from those stakeholders as well as outside reviewers and consultants. Industrial organizations may rely heavily on metrics involving financial return, whereas government organizations may focus more on delivering needed value to stakeholders, consistent with mission statements.

When more than one organization is involved in the portfolio being examined, because of differences in the missions, history, and/or policies, a more complex situation of portfolio management occurs. In that case, a project may have different objectives within the context of each organization.

Workforce and Physical Resources Management

An effective assessment considers the adequacy of both the workforce and the physical facilities and equipment available to address the mission. Because the skills and capabilities of the workforce and the equipment and facilities available to them are key factors affecting the quality of the R&D work performed, workforce and physical resources are discussed in more detail in the Chapter 4, which discusses assessing the quality of the work. It is noted here that it is the responsibility of management to ensure that effective assessments of the available and needed workforce and physical resources are performed and that the results of these assessments are addressed with an eye toward meeting the requirements of the organization's vision, mission, stakeholders, and customers.[2,3]

Organizational Leadership and Management

As stated above, understanding the context within which an organization conducts its R&D is fundamentally important during its assessment. Effective management structure will be appropriate to the nature of the work. Basic research typically requires a very flat management structure, significant individual freedom in selection of research directions, and a management very receptive to suggestions (although projects involving very large experimental resources such as accelerators may require more structure). Product development typically requires a more hierarchical structure in order to ensure mission progress, although a fast-to-market product requires a streamlined structure.

A research organization is an open system with input (the external environment), throughput (the organization and its components), and output (performance by individuals, groups [units], and the total organization). A feedback loop connects output with input and provides a continuous process.

Organizations perform many functions, activities, and events that define their existence. What happens on a daily basis determines what the organizations are. These many aspects of organizations can be quantitatively and interactively overwhelming. Therefore, a conceptual framework that summarizes and simplifies this organizational complexity can be helpful for more effective leadership and management. One such framework is that proposed by the Burke-Litwin model.[4,5]

The following, based on the Burke-Litwin model, is an abbreviated set of questions proposed for consideration in an evaluation of leadership and management; the mechanisms for addressing them can vary: Do senior leaders monitor the external environment (consumers, stakeholders, technology, scientific community), gather relevant information, and share it across the organization? Does the research organization have a clear statement of purpose that shows a clear link to the external environment? Does the organization have a clear process for executing the mission? What long-range planning and forecasting are done to address future needs (technology, staff, facilities, equipment)? Are employee satisfaction factors addressed so that the employees focus on their work and good performers remain and attract others? Are those selected for leadership chosen according to clear and evidence-based criteria such as technical and managerial competence, self-awareness, relevant experience, learning agility, vision, and energy? Are creativity and performance recognized and rewarded? Are people, vision, and mission aligned? Does the organizational structure encourage communication? Are organizational members provided with the resources they need? Are managers and other staff appropriately involved in decision making? Are there clear, reliable, and valid metrics for

---

[2] National Research Council, 2010. *Capabilities for the Future: An Assessment of NASA Laboratories for Basic Research*. The National Academies Press, Washington, D.C.
[3] Y. Saad, 2012. "Review of IBM's Technical Review Path." Presentation to the National Research Council's Panel for Review of Best Practices in Assessment of Research and Development Organizations, March 20, Washington, D.C.
[4] W.W. Burke and G.H. Litwin, 1992. A causal model of organizational performance and change. *Journal of Management* 18 (3):523-545.
[5] W.W. Burke, 2010. *Organization Change: Theory and Practice.* Sage Publications, Thousand Oaks, Calif.

evaluating individual performance, work unit performance, managerial performance, and performance of the organization overall?

A more detailed list of questions, also based on the Burke-Litwin model, is presented in Appendix H. Many of the questions presented here and in Appendix H may be addressed in a self-evaluation process by a management that is interested in continuous improvement. If the purpose of the review is internal—that is, requested by management to assist it in an assessment process, it can be handled by self-assessment or by invited external peers. However, for an externally mandated evaluation of all aspects of the organization, some questions have to be investigated by external reviewers.

Some of the questions listed here and in Appendix H need to allow for anonymous input from staff. This can be done in several ways, including the use of "skip-level" meetings, which are meetings between external reviewers and non-management staff in the absence of managers, during which anonymity is assured to the staff. These meetings can be valuable for identifying potential problems that may not have come to the attention of management. Some assessments organized by the National Research Council have used skip-level meetings, and it has been found that they are most successful when a few external reviewers meet with relatively small groups of employees at similar points in their careers (a small number of staff, so that there are opportunities for most to contribute) and at varying levels of leadership. It is important to communicate to the staff involved that any issues identified at such a meeting might be discussed with management, but that no individual attendee will be identified. It is helpful for the reviewers to have suitable open-ended questions to ask the group. For example: What aspects of the job make you eager to come in each day—and what aspects of the job make you dread coming in? It is important that the reviewers present attempt to prevent staff input from being monopolized by a single individual, and if one individual seems to focus on a small set of issues, it is important to try to determine if complaints are widely shared. It is also important that reviewers avoid leading questions that are designed to elicit specific responses. Information derived from these meetings is best used for the development of hypotheses for further investigation rather than as data from which conclusions can be immediately drawn, given that there are issues associated with sample size and selection.

Other means of garnering input from staff include the following: a parallel organization exercise (a temporary arrangement whereby formal authority is suspended and people can speak openly without fear of retribution), the purpose being to address issues not adequately addressed within the formal structure; and an informal relations survey (an organizational-network analysis consisting of a computer-generated map of the informal organization, identifying individuals to whom others reach out regardless of position or rank).

Planning for the Future

It is important that management continually plan for the future so that the organization is always in a position to fulfill its mission. This means that the right resources and leadership will be in place when needed. The demographics of the workforce will be tracked so that there will be leaders in place as retirements and departures occur, and people with new skill sets will be recruited and trained so that the staff is ready to deal with new technologies. An effective assessment considers the adequacy of this planning as well as plans for the necessary physical infrastructure to support the new skill sets and technologies.

SUMMARY OF FINDINGS

The key components of the management of an organization considered in an effective assessment are the R&D portfolio, resources, the organizational leadership and management structure, and planning for the future.

Research is characterized by the fact that the result of any particular approach is not known until the research has been performed. Hence it is important that a research program—even one with a specific mission—examine several alternatives in a portfolio of projects. Similarly, in development it is usually not known which technology will lead to the best (most cost-effective, efficient) solution, and so it is important to consider a portfolio of technologies. For an effective assessment, resources will be evaluated to determine if they are adequate for the proposed or ongoing program, but the evaluation will also recognize the external resource limitations (budgetary, regulatory, personnel). To be effective, research leadership and structure will be adapted to the environment in which it operates—no one model suffices for all styles of research organizations. The future always brings improved science and technology, and so if the parent organization is to remain competitive, it is important that its R&D units continuously upgrade the skill sets of its workforce and match these with adequate research facilities.

4

Assessing Technical Quality

It is important to consider such factors, discussed above, as type of R&D organization, type of assessment, and timescale of the assessment, so that the assessment of technical quality can be appropriately tailored to achieve its purpose for the given organization.

CHARACTERISTICS OF QUALITY

Quality is the characteristic of an R&D organization best suited to quantitative assessment and metrics. This characteristic encompasses both the workforce and the work being performed, as well as the adequacy of resources provided. Because no single statistic captures quality, it is important to define an appropriate set of tools that measure both the quality of the workforce and the quality of the output of the research. The aspects of quality listed here interact with and are also part of the assessment of the management practices, discussed in Chapter 3.

Quality of Personnel

The importance of the quality and expertise of people in an organization cannot be overemphasized.[1] Many organizations have determined that the quality of the workforce is the most reliable predictor of future R&D performance, independent of mission drivers or impact. An assessment of the quality of the workforce is a fundamental best practice common to almost all assessments. This assessment can be both quantitative and qualitative, and benchmarking can be useful. A productive skill-set balance of the workforce will match the mission of the organization and be of an appropriate quality. It is advisable to evaluate management's plans for recruiting the type of person most suited to the job.

Of special significance for many R&D organizations is the presence in any particular division, laboratory, or project of people with deep and creative technical capabilities.[2] It is important to assess the organization's policies and actions aimed at steadily building on the sets of capabilities associated with individuals possessing both breadth of experience across multiple projects and depth in one or more systems and disciplines. An effective organization enables its staff to capture new skills required for a given set of tasks at hand, while over the long term building the network required to make team members effective participants in global efforts to achieve the overall goals of the organization. To facilitate the creation of such capabilities, a diversity of personnel and work experience is vital. An effective assessment of an organization

[1] J. Lyons, 2012. *Reflections on Over Fifty Years in Research and Development: Some Lessons Learned*. National Defense University, Washington, D.C.
[2] Y. Saad, 2012. "Review of IBM's Technical Review Path." Presentation to the National Research Council's Panel for Review of Best Practices in Assessment of Research and Development Organizations, March 20, Washington, D.C.

includes consideration of a diverse workforce whose contributions may affect and advance the R&D mission of the organization.[3,4,5,6]

The success of a strategy that builds and strengthens this workforce can be assessed by examining movement of personnel both within and outside a given organization. Over time this strategy will provide for a diversity of experience with the broadest possible base of professional interactions. Thus, a key element of any meaningful assessment will be the agility with which an organization acts, at all stages of a program, to train its team and to provide a high level of exposure to diverse programs and organizations working with related or overlapping technical agendas and expertise.

At the heart of looking forward to the next generation of scientific and technological opportunities are the organization's scientists and engineers. Their knowledge of cutting-edge research will be an essential starting point in all such forward-looking efforts. To be effective, they will have flexibility to attend scientific and technical professional meetings and to participate in the international community of scholars. They will be encouraged to think about next-generation efforts, and they will be rewarded for that effort by some strategy that brings resources to bear on the most promising ideas and allows some to pursue high-risk, high-payoff efforts. More difficult but equally important is a parallel peer-review process that continually evaluates such highly speculative programs and aids in determining when available data indicate a clear likelihood of failure and suggest reallocation of assets.

Organization directors' discretionary funding, internal allocations for basic research, and outside ("other agency") funding sources are all key factors that enable the process of fostering a high-quality technical staff. This approach, enriched by interaction with academia, has proven to continue to yield dividends in new and unanticipated discoveries, often without guidance by a well-defined and planned timetable for discovery.

Quality of R&D

The desired outputs of the R&D organization differ by type of organization, but they can typically be represented by measurable quantities such as publications (and their quality), patents, copyrights, and peer awards. The absolute value may not be as important as the trends in such quantities. An effort to benchmark these metrics across similar organizations can make the absolute value more meaningful.

Preparedness

Preparedness is defined as the actions taken by the organization to identify and maintain the resources and strategies necessary to respond flexibly to future challenges.

---

[3] C. Herring, 2009. Does diversity pay? Race, gender, and the business case for diversity. *American Sociological Review* 74:208-224.

[4] O.C. Richard, 2000. Racial diversity, business strategy, and firm performance: A resource-based view. *Academy of Management Journal* 43:164-177.

[5] O. Richard, A. McMillan, K. Chadwick, and S. Dwyer, 2003. Employing an innovation strategy in racially diverse workforces: Effects on firm performance. *Group and Organization Management* 28:107-126.

[6] O.C. Richard, B.P.S. Murthi, and K. Ismail, 2007. The impact of racial diversity on intermediate and long-term performance: The moderating role of environmental context. *Strategic Management Journal* 28:1213-1233.

Capabilities

A core capability is defined as an area of sustained investment by the organization. It is frequently measured by comparison with peer organizations. Once core capabilities are identified, typically as part of any strategic planning exercise, the maintenance of those capabilities will be one essential component of an assessment of an organization's strategy. The development of new capabilities, when applicable to an organization, is also part of a preparedness assessment.

Research Infrastructure

The quality of the facilities and equipment, including buildings as well as capital equipment, is also an important part of any assessment of quality, since the lack of this infrastructure can stand in the way of even the highest-quality workforce. Benchmarking is an effective method here. Assessment of the plans for future infrastructure, including the physical plant, capital equipment, and other factors, is part of the assessment of preparedness. Assessing the current state of infrastructure is best done as part of an overall assessment of near-term quality; the planned future state of the infrastructure is properly assessed as part of preparedness for long-term quality.

It is important that an evaluation of the quality of research management be done in the context of the possible, considering whether management is doing the best that it can with the hand it has been dealt (and is making its best case for a better hand). For example, an assessment of the facility infrastructure necessarily reflects budget realities—if there are strong budget constraints, it is appropriate to ask whether management is investing in the best combination of new and upgraded facilities. If the assessment is also intended for an audience that includes budgeting authorities, it is appropriate to assess the potential value of additional infrastructure funding.

Technical facilities encompass the physical space of the organization and how it is occupied. Does the workforce have what it needs to carry out the research program as identified or planned? An evaluation of facilities independent of a clear understanding of program content is just as imprudent as an evaluation of program content without recognition of appropriate facilities needs. Metrics for the quality of the research infrastructure are not readily generalized, but it is important that the assessment panels be presented data to assist their evaluation and not merely asked to "eyeball" a site while looking at posters of technical accomplishments.

Ultimately, a good assessment will help management to judge whether the quality, cost, and capability of facilities are matched to the technical need. For example, space in an old facility may be acceptable, as long as it is at least adequate for optimum equipment utilization. Some of the required elements are adequate power and appropriate power backup, and environmental control (e.g., temperature, pressure, vibration) that is consistent with the requirements of the equipment needed.

How much equipment the organization has is another metric that may be examined. It is not possible to specify a generalizable fraction of available funds that should be allocated for infrastructure maintenance and upgrade. To determine this requires the evaluation of context, including available funds and R&D priorities. "Home-built" equipment is a potentially useful measure of creativity that can be properly identified and documented for the assessment team.

The successful capture of material intellectual property from custom efforts is a possible indicator of quality.

HOW TO ASSESS QUALITY

An effective assessment includes appropriate qualitative and quantitative measures for all aspects of the organizational research activities. For an effective assessment, research activities will not be viewed in isolation, but as part of the entire research portfolio of the organization, and in the context of the priorities within that research portfolio. An assessment will often include the utilization of panels of domain expertise, but the use of such domain expertise is not sufficient to ensure that the assessment is appropriate and adequate.

R&D organizations are frequently organized according to scientific or technical disciplines, and when their projects and programs do not involve multidisciplinary collaborations, they are amenable to being assessed by discipline-oriented panels of peers. This was the historical precedent created by the National Research Council (NRC) in concert with the National Bureau of Standards (NBS; now the National Institute of Standards and Technology [NIST]). This same process was adopted by the NRC as it established an assessment process for the Army Research Laboratory (ARL) and other federal organizations. This type of process also describes the assessments of the Air Force Research Laboratory (AFRL) made by its external Scientific Advisory Board and is similar to that used by many other organizations. However, as technology and related programs become more complex, one increasingly finds a matrix-like organizational structure developing in organizations. Even if formal management and personnel policy do not reflect this matrix, technical programs crossing disciplinary organizational lines do.

Such cross-organizational programs will also be reviewed in a comprehensive assessment.. The process and frequency of cross-organizational program assessments will differ in most respects from one another, and an effective assessment will not be found in a process focused exclusively on visits to discipline-organized subsets of an organization. Assessment of a cross-disciplinary program may benefit significantly from establishment of a peer group drawn from a cadre of experts already fully engaged in disciplinary assessment of relevant programs. Appendix I presents examples of recent cross-organizational assessments that may clarify some of the issues faced in such assessments. The examples presented in Appendix I are recent reviews of the manufacturing-related programs at NIST and the autonomous systems program at ARL—each of which involved projects whose participants were drawn from multiple laboratories.

The Role of Peer Review

Peer review, coupled with quantitative and qualitative metrics (see the section below), offers an opportunity to gain a better understanding of and then to assess an R&D organization. A well-selected team of experts can produce valuable insights with respect to the overall quality of the R&D strategy and its execution within an organization. Detailed critiques and insightful suggestions from experts permit checks and balances among contrasting points of view. Including individuals with expertise in emerging areas at the margin of the main focus of the assessment can help to identity new or missed opportunities.

Peer review is an accepted process that is understood by all, and it provides scientific accountability. It can also identify links to the external scientific and technical community and to relevant R&D performed elsewhere. Peer review can also provide advice on decision making, particularly with regard to resource allocation for the array of R&D directions that an organization may be considering.

External reviewers bring to an assessment their individual perspectives, which may constitute biases or conflicts of interest. There is generally more credibility for independent external assessments. A fully independent assessment will be arranged by and managed by an independent contractor. It is essential that candidates for membership on assessment panels be required to present any biases or potential conflicts of interest to the contractor so that the appointment decision can take such potential conflicts into consideration.

In any case, external peer review is essential for evaluating any R&D organization. Peer review is the most valuable and credible best practice an organization can employ to assess its quality and, in many cases, its impact.[7,8,9,10,11]

Appendix J summarizes assessment processes at NIST, ARL, and Sandia National Laboratories. Appendix K summarizes assessment processes at some other government laboratories: within the DOD (Army, Air Force, and Navy); the National Institutes of Health; and the Department of Energy. Each of the processes described in Appendixes J and K involves peer review of technical quality.

Metrics

Before considering the best metrics to use and how to use them, it is necessary to determine what is to be measured and why.[12,13] A report of the National Research Council committee that examined metrics for the U.S. Environmental Protection Agency suggested that metrics must be meaningful to the recipients of the results of the assessment, simple and understandable, integrated into an overall assessment, aligned with the goals of the organization, both process- and outcome-oriented, accurate, consistent, cost-effective, and timely.[14] In evaluating metrics for R&D, it is important to avoid simply counting (i.e., numbers of papers, patents, etc.) and instead to identify how the measures will be mapped to decisions that are expected to rely on the results of the assessment. It is also helpful to identify and distinguish between leading and lagging indicators among metrics. Investments in R&D may provide for a number of outputs such as those identified in the following list (a detailed discussion of possible metrics is also provided by Geisler[15]):

---

[7] National Research Council, 1995. *On Peer Review in NASA Life Sciences Programs.* National Academy Press, Washington, D.C.
[8] National Research Council, 1998. *Peer Review in Environmental Technology Development Programs.* National Academy Press, Washington, D.C.
[9] National Research Council, 1999. *Evaluating Federal Research Programs: Research and the Government Performance and Results Act.* National Academy Press, Washington, D.C.
[10] National Research Council, 2007. *Assessment of the Results of External Independent Reviews for U.S. Department of Energy Projects*. The National Academies Press, Washington, D.C.
[11] U.S. Office of Science and Technology Policy (OSTP), 1983. *Report of the White House Science Council* ("Packard Report"). OSTP Report Number NP-3902794. OSTP, Washington, D.C.
[12] The National Research Council, 2005. *Thinking Strategically: The Appropriate Use of Metrics for the Climate Change Science Program*. The National Academies Press, Washington, D.C.
[13] R. Behn, 2003. Why measure performance? Different purposes require different measures. *Public Administration Review* 63(5).
[14] National Research Council, 2003. *The Measure of STAR: Review of the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) Research Grants Program*. The National Academies Press, Washington, D.C.
[15] E. Geisler, 2000. *The Metrics of Science and Technology.* Quorum Books, Westport, Conn.

- Publications, patents, reports, and the citations garnered;
- Technical assistance provided to end users, customers, and stakeholders;
- Invited presentations (e.g., at conferences and workshops);
- Training and mentoring of personnel;
- New and improved products, materials, and processes;
- Patents leading to new products;
- Development of test and evaluation protocols, codes, and standards;
- Technology transfer;
- Maintained competencies;
- New competencies;
- Cost savings (e.g., in materials or processes);
- Increased productivity;
- Safety practices and culture;
- Effectiveness of management structure and strategy; and
- Recognition of the R&D organization as best, among the best, or unique.

Bibliometrics

Bibliometrics are methods to quantitatively analyze scientific and technological literature. Bibliometrics can be used in the evaluation of individuals, groups, or institutions as a whole. The collection of the data and its analysis are straightforward. Bibliometric measures allow for the quantitative assessment of R&D outputs by simple counts of papers, citations, and patents. This process allows for clear assessments of core journals and their relative impact, including their journal impact factors. Citation analysis (Science Citation Index) can be used to help determine the role that individual scientists, their groups, and their institutions have in the evolution of new ideas and their technological development. With proper analysis, bibliometrics can identify trends and emerging concepts in science and technology even across diverse scientific disciplines. Since bibliometrics are accepted by the general scientific community, they can be used as a reasonable representation of the outputs of a research organization.

In using bibliometrics, it is important to recognize that the analysis does not include all published articles and that other types of written outputs such as technical reports are not covered. Also, they do not account for work in progress. Citations are not a measure of quality; a high citation count may simply be a measure of those papers that are in concert with others and that are not truly important. It is also hard to validate cross-disciplinary research in publications, owing to different structures and procedures for different disciplines. New, evolving, and mature areas all have different publication and citation rates. There is no standard for the validation of counts of papers and citations as they relate to quality. For example, much can be written about a mistake.

Another factor to consider is that bias can be an issue in assessing the merit of paper and publication counts. For example, for papers that discuss research near the boundaries of their discipline publication and citation in top journals may be more difficult to achieve. Also bibliometric databases may not adequately cover conference proceedings, which are increasingly important in some fields such as computer science. Given the caveats above, absolute bibliometrics can be less than useful. Assessing the trend in such metrics can, however, provide a lagging indicator of research quality.

The visibility and global impact of research investments may also be measured by means of the number of presentations given at meetings both internal and external to an organization. Of particular significance is the number of invited presentations. This measure has value in the assessment of both the organization and individual effort. Although presentations internal to an organization are not searchable through bibliometric databases, those given at symposia, workshops, and conferences are available in databases. As with the other bibliometric data described above, it is important to consider this metric in context and to avoid its use in isolation.

Patents

Patents are another measure of the outcome of investments, and they may be viewed as a measure of the potential market applications resulting from R&D. Patents are also legal documents and may be viewed both as a measure of scientific productivity and as a measure of intermediate outputs. Patents are a measure applied across the science and technology organization. Because the format of patents is uniform, comparisons can be made across diverse research organizations and even between countries. Because patents contain information on inventions that have resulted from science and technology activity, it is possible to reconstruct levels of investment. Science and technology can lead to a high quantity and quality of patents, which may correlate with an improved knowledge base, improved pool of skills, the protection of promising ideas, new products, and improved innovation activity.

Data related to patents are easily quantifiable; patent databases are large and easily searched, and the data can be manipulated and readily cross-correlated. Also, the citations are readily available. Patents provide important information about the actual work and level of effort that led to the patent. The potential impact of a patent suggests that it can be considered a reasonable measure of output. Successful patents are generally considered to be a measure of technological performance. As such, they provide indicators of the knowledge base and the quality of the research that led to them.

In making assessments based on patents, there are a number of points to consider. Whether or not to file a patent is decided by the individual organization, and there are marked differences in the propensity of organizations to file for patents; a more meaningful metric, particularly for long-term assessment, is the quantity and quality of licensed patents. Also, not all of the research investment of an organization will lead to a patent, and thus patents alone, licensed or not, cannot be used as a stand-alone metric of quality. Further, there is not a one-to-one correspondence between a patent and a product, and such correlations are not easily derived from a database. Although there may not be a clear method for modeling patents and their relationship to R&D performance, the patent metric can be a useful tool when considered in the proper context.

Cultural Metrics

Some metrics associated with the quality of an R&D effort relate to the culture of the organization. Some factors that may allow an assessment of the cultural environment include the importance placed on the training and mentoring of personnel and the commitment to safety practices and culture. A healthy organization requires buy-in and commitment to the importance of these elements at all levels, from the most senior manager to the individual executing a given program.[16] Some elements of the assessment of cultural factors will be quantitative—for example, numbers of courses or degree programs associated with training, safety initiatives, accidents, and other indicators relating to safety—although the following important metric is largely qualitative in nature: Does the cultural environment provide meaningful support?

Other considerations related to organizational culture may include: To what extent are organizational members clear about organizational policies and processes, explicit and implicit? To what extent do organizational members agree that "this is a great place to work"? To what extent are organizational members treated with respect and dignity? Are differences among people respected and encouraged, or is the expectation one of bias and prejudice? Is conflict surfaced and managed, or is it avoided?

As is readily gleaned from the above discussion, indicators may be either quantitative or qualitative in nature and may be found in various data sources, including employee surveys. For instance, quantitative measures include number of publications and presentations, citations, and new products and processes. Return on investment and performance outputs can be important metrics.

Other outputs not as readily associated with quantitative measures might also have significance. Examples include impact on customer satisfaction, contributions to the pool of innovations, global recognition, the effectiveness of organizational leadership, communication among various entities within the organization and with relevant stakeholders, and the ability to transition research from invention/innovation to later stages of development.

Appendix L provides a set of assessment metrics and criteria applied by NRC panels that review the ARL. This set of metrics and criteria is not presented as a prescription, but, rather, as an example of a tailored set developed to meet the perceived assessment needs of one organization. The assessment items identified fell into the following categories: relevance to the wider scientific and technical community, impact to customers, formulation of the goals and plans for projects, methodology applied to the research and development activities, adequacy of supporting capabilities and resources, and responsiveness to the findings from previous assessments.

Benchmarking

One commonly used assessment approach is to compare one R&D organization with one or more others judged to be at the top level of performance. This is usually done with metrics that are normalized to account for size differences. Thus one may cite the number of archival publications for each technical professional. Using percentages accounts for size differences, e.g., the percentage of doctorates among the professional population. It is important that comparisons made are among R&D organizations operating in similar contexts.[17] For example, comparing an engineering research organization with an academic department provides little meaningful information, because the two operate in different contexts. A problem with benchmarking with metrics is that such assessments do not reveal the effectiveness of the organizations. A first- class organization may reside in a parent that fails to capitalize on the

---

[16] B. Jaruzelski, J. Loehr, and R. Holman, 2011. *Why Culture Is Key*. Booz & Company, New York, N.Y.
[17] National Research Council, 2000. *Experiments in International Benchmarking of U.S. Research Fields*. National Academy Press, Washington, D.C.

organization's breakthroughs. Nonetheless, benchmarking can be a useful addition to the assessment tool kit.

SUMMARY OF FINDINGS

Different aspects of the assessment of technical quality are suited to appropriate quantitative and qualitative metrics. The quality of the research activities is best viewed in the context of the entire portfolio of the organization, including suitable assessment of cross-organizational programs.

Peer review, coupled with quantitative and qualitative metrics, is a critical part of an effective assessment of the R&D organization.

Quantitative metrics can play a key role in assessment, but it is important to determine first what will be measured and why, and to avoid counting the numbers without a good rationale. Typical bibliometric measures are publications and presentations. Patents are another quantifiable metric. For each of these, the total number is not nearly as important as the quality and the contribution (and commonly accepted measures such as number of publication citations can be a poor surrogate for quality). Other indicators, generally qualitative, are associated with the culture of the organization; these include training and mentoring, safety initiatives, and the effectiveness of management.

Quantitative metrics can usually be associated with such additional aspects as return on investment, the development of new products and processes, and internal productivity and/or cost savings. Other indicators that are more qualitative include technical assistance, customer satisfaction, communication (both internally and with stakeholders), and global recognition (including benchmarking). The transitioning of research into products has both quantitative and qualitative aspects.

5

Assessing Impact

Measuring the impact of R&D activities is the most subjective aspect of an assessment. An insightful definition of impact was posed by William Banholzer in his presentation at the National Research Council's workshop on best practices in assessment of research and development organizations: "What would not have happened if you did not exist, and how much would society have missed?"[1]

An assessment of impact may be designed by considering the following questions:

1. Does a survey of program outcomes for the time period being assessed (near term, midterm, and far term) indicate expected levels of success in the context of the assigned program objectives?
   a. Was the stakeholder, sponsor, or customer satisfied with the output delivered?
   b. Did the outcome advance the field significantly?
   c. Was there a sustainable advantage associated with the outcome?
   d. Was there external recognition of the outcome as fundamentally differentiating in the field?
2. Does a survey of programs launched within the relevant time period indicate an ability to consistently lead in the definition of the next steps required to continue to advance the field within the scope of the mission—for example, what is the track record of the organization in providing forecasts to the sponsor, stakeholder, or customer?
3. Has the organization consistently identified foundational discontinuities in the trajectory of the field of interest, opening unanticipated new fields of study and opportunities?

Providing answers to these questions has been challenging for most organizations. The answers are best sought primarily through analysis of completed R&D, sometimes by examining the recent past and at other times by conducting retrospective analyses of more distant history.

R&D organizations are part of a system or process that leads to products that an organization's management, stakeholders, and customers will ultimately use as the basis for judgments about the worth of the R&D. However, the links between the R&D and the final impact, perhaps well understood by those performing the R&D, are often not understood by some stakeholders and customers, who may not be technical experts. It is the task of the R&D organization to make the case about those links between R&D and impact, and various strategies have been used. For convenience, a single word, *transition*, is used here to characterize the process and the consequence of delivering the product of any accomplishment of the R&D organization to any one of its stakeholders and customers.

For basic research, the claim is often made that the distance between the research and the end-item product is so great that one should judge the research as an entity on its own. To

[1] National Research Council, 2012. *Best Practices in Assessment of Research and Development Organizations—Summary of a Workshop*. The National Academies Press, Washington, D.C., p. 10.

identify transitions from basic research, measures are collected on papers published, citations by others, invitations to speak about the work, patents received, and awards conferred. These metrics for assessing research quality are useful for assessing the quality of R&D, as discussed in Chapter 4, but they are often of little interest to those stakeholders and customers whom the organization wishes to satisfy. Organizations that conduct R&D recognize this, and so their reports to stakeholders frequently include anecdotal stories about how research funded many years earlier has found its way into application today.

In many R&D organizations transitions from basic to applied research are frequent, and are often the subject of management decisions. Within an R&D organization additional indicators for examining transitions from basic research are available, but they are less frequently tracked than the obvious outputs in papers and technical talks. Some related questions are the following: Has the technical advance in a particular project opened the door for more applied work? Does the applied work justify additional funding and possibly new hires? These questions may apply whether the basic work was done in-house or extramurally. These questions are regularly asked and decisions are made, but organized records of such transitions are not commonly compiled or recognized as appropriate indicators for judging the basic work.

With respect to assessing the impact of applied research or product development, the situation becomes far more complex. Industry will appropriately focus on its bottom line, and predicted return on investment for the R&D investment can be calibrated against actual sales. Feedback from both failures and successes may be communicated to stakeholders and used to modify future investments. Government organizations rarely have such a direct metric and must search for more information and a structure in order to communicate to their myriad stakeholders.

Four general types of R&D organizations are identified above in Chapter 2: mission-specific, industrial and contract organizations, product-driven organizations (e.g., national laboratories), and universities.

The question of impact is different for each type of organization. Mission-specific organizations in the government and industrial organizations with clearly defined missions are generally considered the underpinning of an organization that includes centers for development, testing and evaluation, and maintenance. Each organization includes management processes designed for transitioning the product of its efforts to the next stage of development and/or application. In many instances these process include ensuring that adequate resources are available so that a transition can indeed occur. Formal agreements often ensure handoffs of responsibility upon completion of the organization's contributions. These processes offer excellent opportunities for examining the recent past as well as for aiding in the scholarly task of examining the more distant past.

In the NIST laboratories, for example, there are frequently tangible consequences of the R&D work that can be monitored, including standard reference materials sold, data used, and standards developed and promulgated by standards-making bodies. In addition, NIST participates, as do other government labs, in cooperative research and development agreements (CRADAs) that may be seen as a potential database of relevant transitions. Within such CRADAs, the R&D organization and its partners (other government organizations, universities, and industries) carry out collaborative efforts designed explicitly to transition the results of more basic research toward application. The challenge with respect to providing information supporting assessments is to identify these transitions in a manner that reveals the implications for future decisions and to make the data accessible and comprehensive enough to allow comparisons over time and within the organization.

For product-driven national laboratories, the impact of R&D activities has more variation in its characteristics. An organization typically has carved out one or more core mission spaces, and the impact of those is measured by the degree to which these missions are being carried out. In emerging mission areas, the impact is measured by the growth of external investment, and in future mission areas impact is measured by the successful establishment of the workforce and competencies that will be needed to address these areas.

In its simplest form, an industrial organization could measure its impact in terms of return on investment. This form of assessment is a best practice, but not a universal one, and involves caveats—many things affect earnings, such as price, volume, and cost of the existing product, and the time lag between investment and earning can be as long as a decade.[2]

For academic research groups, a common best practice is to assess the quality of the organization by benchmarking it against other laboratories generally recognized to be successful. This is a qualitative but widespread practice.

TIMESCALES OF IMPACT

In attempts to project the future impact of current or proposed programs, the R&D organization is hampered by the fact that it may take many years or even decades before the full impacts of current programs are realized. When that eventually does come to pass, there have usually been so many different organizations involved in developing, engineering, producing, and fielding the end item that its identity with the research organization is often lost. Regardless of how the story is formulated, the stakeholders' confidence in the organization and its management will be bolstered by demonstration that the decision making and processes of the present are comparable to, or better than, those of the past that led to measurable impacts. To tell this story properly, many organizations have had recourse to looking backward and tracing the consequences of R&D events long past.

An example of this approach for learning about impact is Project Hindsight.[3] Carried out by the Department of Defense in the mid- to late 1960s, Project Hindsight was a study of the development of 22 different weapons systems drawn from across the military services. It involved more than 200 personnel over a period of approximately 6 years. For years afterward the observations and conclusions of Project Hindsight guided military R&D planning and decision making. In 2004, recognizing that much had changed in the intervening years, the U.S. Army commissioned a new study, Project Hindsight Revisited.[4] The Air Force Research Laboratory (AFRL) identified research transitions over a 50-year period at the AFRL,[5] and Berry

[2] W. Banholzer and L. Vosejpka, 2011. Risk taking and effective R&D management. *Annual Review of Chemical and Biomolecular Engineering* 2:8.1-8.16.
[3] Office of the Director of Defense Research and Engineering (DDRE), 1969. *Project Hindsight: Final Report.* Office of the DDRE, Washington, D.C.
[4] J. Lyons, R. Chait, and D. Long, 2006. *Critical Technology Events in the Development of Selected Army Weapons Systems: Project Hindsight Revisited*. National Defense University, Washington, D.C.
[5] Air Force Office of Scientific Research, 2002. AFOSR at 50: Five decades of research that helped change the world. *Research Highlights,* March/April.

and Loeb identified breakthrough Air Force capabilities spawned by basic research.[6] Coffey and coauthors describe methods and challenges in tracing the development of selected technologies from basic research through significant exploitation of the research.[7]

The Department of Energy (DOE) utilized a similar retrospective analysis, with the assistance of the NRC, examining the impacts on energy-producing and energy-using industries of R&D programs executed by the DOE laboratories over the time period 1978-2000. The report summarizing the findings of the assessment makes the case in economic terms for a return on investment that by itself could justify funding the research, while recognizing that societal impacts are far more difficult to measure and are not readily quantifiable.[8]

An exemplary example of after-the-fact analysis of impact is the economic impact analyses performed by NIST. Stimulated in part by the concerns of many in Congress that a program such as the NIST Advanced Technology Program (ATP) was inappropriate for government and that the R&D carried out under its aegis would be better left to industry, NIST began the ATP program in the late 1980s with a coordinated plan to measure economic impact. Using outside expertise and a well-articulated process, this impact analysis was applied to all funded activities and its record made available to all interested parties. A summary of the first 10 years of ATP funding, published in 2003, provides methodologies and results gained from their application to many industrial sectors. NIST has gone on to apply such organized studies to many of its laboratory-based efforts, and the historic record of accomplishment provides a strong justification for future work in the areas whose impacts have been assessed.[9]

SUMMARY OF FINDINGS

Measuring the impact of R&D activities is the most subjective aspect of assessment and is ill-suited to quantitative measures. Transitions are measurable at every level of R&D (see Figure 2-1). The impact of research can only be measured after the fact. Near-term impacts, including many transitions during R&D phases, require looking back at the recent past but can be monitored for most of the R&D effort. Long-term impacts require deeper historical probes and are more likely to be assessed for only a few notable examples. In both cases, organized processes for gathering and analyzing the data require management attention and designated leadership. Developing the data and presenting them in a manner that makes the data useful to the intended audience is a job for professionals. The R&D organization will benefit from having an appropriately supported historian and internal report requirements to ensure the utility of the process.

---

[6] W. Berry and C. Loeb, 2007. *Breakthrough Air Force Capabilities Spawned by Basic Research.* National Defense University, Washington, D.C.
[7] T. Coffey, J. Dahlburg, and E. Zimet, 2005. *The S&T Innovation Conundrum*. National Defense University, Washington, D.C.
[8] National Research Council, 2001. *Energy Research at DOE: Was It Worth It? Energy Efficiency and Fossil Energy Research 1978 to 2000.* National Academy Press, Washington, D.C.
[9] R. Ruegg and I. Feller, 2003. *Evaluating Public R&D Investment Models, Methods, and Findings from ATP's First Decade*. National Institute of Standards and Technology, Gaithersburg, Md.

6

Guidelines for Consideration During Assessment

This chapter presents a series of questions that a manager—either the organization's director or a responsible remember of the parent organization—can ask when considering carrying out an assessment of a research organization. The preceding chapters of this report addressed these questions and suggested some best practices.

ASSESSING MANAGEMENT

Answers to the following questions will be useful in the assessment of organizational management:

- Does the organization's management understand its mission and its relationship to that of its parent? Does the vision statement of the organization align with that of the parent organization?
- Is there a long-range plan for implementing the strategy by specific technical programs?
- Does the organization have an explicit strategy for its work and for securing the necessary resources?
- Do the program plans reflect a model for balance—that is, amount of basic versus applied and development research, and short-, medium-, and long-term work?
- Does the organization have a clear champion within the parent organization?
- Does management have an aggressive recruiting plan with well-defined criteria for new hires? Is there a set of practices for retaining, promoting, and recognizing the staff?
- Does the organization have a process for forecasting likely future technical developments in areas appropriate to its mission?
- Does the organization's management have discretionary authority to invest in new programs on its own initiative? Does management solicit ideas from the staff for new work?
- Does management regularly assess facilities and equipment for adequacy? Does it have a fiscal plan for updating or replacing laboratory equipment?
- Is there a process for regularly reviewing the organization's research portfolio for its alignment with the mission?
- What is the management climate, and how does one assess it? Is there enough flexibility to work across organizational lines?
- How does the structure of the organization support its mission?
- How much collaboration is there with outside organizations? How many staff exchanges are there?
- Does management have a well-defined process and criteria for determining what work is performed in-house versus what work is sponsored via grants, contracts, or other mechanisms with external entities?

- Does the management support a culture of creativity, diversity, and entrepreneurship?

ASSESSING THE QUALITY OF SCIENTIFIC AND TECHNICAL WORK

Answers to the following questions will be useful in the assessment of the quality of an organization's scientific and technical work:

- Does the assessment include the quality of the staff, equipment, and facilities?
- Does the assessment include the nature of the research portfolio as to alignment with the mission and the balance in regard to basic, applied, and development work and short-, intermediate, and long-term research?
- Does the organization have a set of indicators that can serve as parameters when the time frame precludes immediate assessment? Does the organization benchmark itself against premier organizations?
- Who is the expected audience for the assessment?
- Is the review done by technical peers?
- What are the criteria for ensuring the credibility and validity of the assessment?
- What is the scope of the assessment? Does it include proposals for new work? Does it include assessment of completed work—internal review and authority to release a report, publications, patents, invited lectures, awards, and the like?
- Who designs and manages the assessment?

ASSESSING RELEVANCE AND IMPACT

Addressing the following questions will be useful in the assessment of an organization's relevance and impact:

- Does the organization have a process for identifying its stakeholders and customers?
- Does it have a regular process for reviewing its programs and plans with its stakeholders?
- Does the organization have a process for learning of its customers' current and likely future needs and expectations for the organization?
- Does the organization have an explicit process for tracking the utilization of its results (e.g., is transition to the next R&D stage actively managed and measured)?
- Does it have a formal program for recording the history of its work from concept to final utility or impact?
- Does the organization have a program to conduct retrospective studies of its earlier work?

APPENDIXES

Appendix A

Biographical Sketches of the Panel for Review of Best Practices in Research and Development Organizations

JOHN W. LYONS (NAE), *Chair,* is a Distinguished Research Fellow at the Center for Technology and National Security Policy at the National Defense University. In 1993, he was appointed the first permanent director of the Army Research Laboratory (ARL). At ARL he managed a broad array of science and technology programs: electronics, information science and technology, armor/armaments, soldier systems, air and ground vehicle technology, and survivability/lethality analysis. In 1990, Dr. Lyons was appointed Director of the National Institute of Standards and Technology. He has received the Department of Commerce Gold Medal and the Department of the Army's Decoration for Exceptional Civilian Service. He has published 4 books and more than 60 papers, and he holds a dozen patents. Dr. Lyons also has served on many boards and commissions. He received his A.B. degree from Harvard University and his A.M. and Ph.D. degrees from Washington University, all in chemistry.

EDWARD A. BROWN is a Principal Staff Member in the Center for Integrated Intelligence Systems of the MITRE Corporation, where he is concentrating on innovative techniques for assisting member organizations of the intelligence community (IC) in the management of their science and technology (S&T) programs. His expertise spans the broad area of managing government S&T enterprises. He went to MITRE after a 33-year career as a government employee within the Army's research and development (R&D) community. One of his final assignments as a civil servant was as a member of the Director of Central Intelligence's Strategic S&T Management Task Force, which was chartered to develop new techniques for managing the IC's S&T enterprise. Dr. Brown is now assisting the IC to implement the results of the task force work in his current position with MITRE. He has supported S&T management improvement efforts in a variety of government agencies and served for 4 years on the Army Laboratory Assessment Group reporting to the Deputy Assistant Secretary of the Army for Research and Technology. Before arriving at MITRE, he was the Director for Special Projects at the U.S. Army Research Laboratory (ARL). In that position he supported the ARL Director in administering and coordinating activities relevant to the management of both the laboratory and its technical program. He was responsible for much of ARL's groundbreaking work in performance measurement and business planning as it applies to R&D organizations. For his work in innovative R&D management, Dr. Brown was awarded the Army's Superior Civilian Service Award. Dr. Brown received his bachelor's degree from Washington and Lee University and his master's and doctoral degrees from New York University, all in physics.

W. WARNER BURKE is Edward Lee Thorndike Professor of Psychology and Education Program Coordinator, Graduate Programs in Social-Organizational Psychology, and Chair of the Department of Organization and Leadership, Teachers College, Columbia University. He is currently engaged in teaching, research, and consulting. He teaches leadership, organizational dynamics and theory, and organization change and consultation. His research focuses on multirater feedback, leadership, and organization change. Dr. Burke's consulting experience has been with a variety of organizations in business-industry, education, government, religion, medical systems, and professional services firms, and he has served as senior adviser to the strategy and organization change practice of IBM Global Business Services. Prior to his move to Teachers College, Dr. Burke was professor of management and chair of the Department of Management at Clark University. Previously he had been an independent consultant as well as serving in various other capacities. Dr. Burke is the author of more than 150 articles and book chapters on organization development, training, change and organizational psychology, and conference planning; he has contributed as an author, co-author, editor, and/or co-editor of 19 books. His most recent (2011) book, published by Sage, is *Organization Change: Theory and Practice, Third Edition.* He received his B.A. from Furman University and his M.A. and Ph.D. from the University of Texas at Austin.

ROSS B. COROTIS (NAE), PE, is Denver Business Challenge Professor of Engineering at the University of Colorado at Boulder. He has research interests in the application of probabilistic concepts and decision perceptions for civil engineering problems, and in particular their application to societal trade-offs for hazards in the built infrastructure. His current research emphasizes the coordinated roles of engineering and social science with respect to framing and communicating societal investments for long-term risks and resiliency. Dr. Corotis was on the faculty at Northwestern University for 11 years; established the Department of Civil Engineering at the Johns Hopkins University, where he was also the Associate Dean of Engineering; and was the Dean of the College of Engineering and Applied Science at the University of Colorado at Boulder. He has numerous research, teaching, and service awards; chaired several committees on structural safety for the American Society of Civil Engineers (ASCE) and the American Concrete Institute (ACI); served as editor of the international journal *Structural Safety* and the ASCE *Journal of Engineering Mechanics*; and chaired the Executive Committee of the International Association for Structural Safety and Reliability. For the National Research Council, he served on the Building Research Board and the steering committee of the Disasters Roundtable, and he chaired the Panel on Assessment of the NIST Building and Fire Research Laboratory. He is the founding chair of the Committee on NIST Technical Programs and past Chair of the Civil Engineering Section of the National Academy of Engineering. Dr. Corotis is a registered professional engineer in Illinois, Maryland, and Colorado; a registered structural engineer in Illinois; and a Distinguished Member of ASCE. He is the author of more than 200 publications. He received his B.S., M.S., and Ph.D. degrees in civil engineering from the Massachusetts Institute of Technology (MIT).

WILLIAM W. CRAIG is the Director of Laboratory Directed Research and Development in the Institutional Science and Technology Office at the Lawrence Livermore National Laboratory (LLNL) and Payload Manager for the NuSTAR Small Explorer Mission at the University of California, Berkeley. He is also the Aerospace Program Manager at the University of California Space Sciences Laboratory, Berkeley. Dr. Craig previously served in the following positions: Deputy Director, Institutional Science and Technology Office, LLNL; Chief Scientist, Physics and Advanced Technologies Directorate, LLNL; Technical Advisor in the Domestic Nuclear Detection Office, U.S. Department of Homeland Security; Group Leader at the Kavli Institute for Particle Astrophysics and Cosmology and Hansen Experimental Physics Laboratory, at Stanford University; and in a number of other positions at LLNL and at the Columbia Astrophysics

Laboratory, Columbia University. Dr. Craig received his B.A. and M.S. degrees in physics and his Ph.D. in astrophysics from the University of California, Berkeley.

C. WILLIAM GEAR (NAE) is Senior Scientist, Chemical Engineering, at Princeton University and President Emeritus of the NEC Research Institute. Dr. Gear's National Academy of Engineering (NAE) citation is for "seminal work in methods and software for solving classes of differential equations and differential-algebraic equations of significance in applications." His primary interest is scientific computation, particularly involving differential equations, and even more specifically, stiff equations and differential-algebraic equations. More recently he has become interested in numerical techniques applied to computer vision. Dr. Gear received his B.A. and M.A. degrees from Cambridge University and his M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign, all in mathematics.

WESLEY L. HARRIS (NAE) is the Charles Stark Draper Professor of Aeronautics and Astronautics and the Director of the Lean Sustainment Initiative at the Massachusetts Institute of Technology (MIT). Before his appointment as Associate Provost, Dr. Harris served as head of MIT's Department of Aeronautics and Astronautics from 2003 to 2008. From 1972 to 1985, he taught and held several administrative positions at MIT. Dr. Harris served as Dean of the School of Engineering at the University of Connecticut from 1985 to 1990, and as Vice President and Chief Administrative Officer of the University of Tennessee Space Institute from 1990 to 1993. As NASA's associate administrator for aeronautics from 1993 to 1995, he was responsible for all programs, facilities, and personnel in aeronautics at NASA. He earned his B.S. in aerospace engineering at the University of Virginia and his M.A. and Ph.D. in aerospace and mechanical sciences at Princeton University, on whose board of trustees he later served.

ELENI KOUSVELARI is a Senior Scientist at the Biological and Materials Sciences Center, Sandia National Laboratories. She is an expert in the direction and organization of bioengineering and translational research. Before joining Sandia National Laboratories, she was the Associate Director for Biotechnology and Innovation at the National Institute of Dental and Craniofacial Research (NIDCR) at the National Institutes of Health (NIH). Before that, she held a number of positions at the NIDCR, including Acting Director of the Center for Biotechnology and Innovation, Acting Program Director and Program Director for a variety of programs, and Chief of the Cellular and Molecular Biology, Physiology and Biotechnology Branch and the Biomaterials, Biomimetics and Tissue Engineering Branch. Before her service at NIH, Dr. Kousvelari held a number of positions at the School of Dentistry at Temple University, the School of Dental Medicine at the University of Connecticut, and the School of Graduate Dentistry at Boston University. She has received numerous awards and is a member of the American Association for the Advancement of Science, International Association for Dental Research, American Society of Cell Biology, Society of Biomaterials, and American Dental Association. Dr. Kousvelari received her D.D.S. degree from the Athens University Medical and Dental School, her M.Sc./C.A.G.S.P. degree in prosthodontics from the Boston University School of Graduate Dentistry, and her D.Sc. in oral biology from the Boston University School of Graduate Dentistry.

BERNARD S. MEYERSON (NAE) is the Vice President for Innovation at IBM, and he leads IBM's Global University Relations Function within IBM's Corporate Headquarters organization.

He is also responsible for the IBM Academy, a self-governed organization of 800 executives and senior technical leaders from across IBM, having been appointed to this position in December 2005. In 1980, Dr. Meyerson joined IBM Research as a staff member, leading the development of silicon, germanium, and other high-performance technologies over a period of 10 years. In 1992, he was appointed an IBM Fellow by IBM's Chairman, and in 2003 he assumed operational responsibility for IBM's global semiconductor research and development efforts. In that role Dr. Meyerson led the world's largest semiconductor development consortium—members being IBM, Sony, Toshiba, AMD, Samsung, Chartered Semiconductor, and Infineon. He has received numerous awards for his work. Dr. Meyerson was cited as Inventor of the Year by the New York State Legislature in 1998 and was recognized as United States Distinguished Inventor of the Year by the U.S. IP Law Association and the Patent and Trademark Office in 1999. He was most recently recognized in May 2008 as Inventor of the Year by the New York State Intellectual Property Law Association. He has published more than 180 papers and owns more than 40 patents. Dr. Meyerson has a Ph.D. in physics from the City College of New York.

ELSA REICHMANIS (NAE) is a professor in the Department of Chemical and Biomolecular Engineering at the Georgia Institute of Technology. Her National Academy of Engineering (NAE) citation is for "the discovery, development, and engineering leadership of new families of lithographic materials and processes that enable VLSI [very large scale integration] manufacturing." Her research is at the interface of chemistry, materials science, optics, electronics, and engineering, spanning the range from fundamental concept to technology development and implementation. Her research is focused on organic and polymer materials design for electronic and photonic applications. She is experienced in leading cross-cultural, multidisciplinary research teams and in generating value for intellectual property through patent and technology license agreements. Dr. Reichmanis has published extensively; has organized national and international workshops, symposia, and conferences; and has mentored students and post-doctoral fellows and taught courses. She has received numerous awards and has more than 150 publications, more than 15 patents, and 5 books to her credit. Dr. Reichmanis received her B.S. in chemistry and her Ph.D. in organic chemistry from Syracuse University.

JOEL M. SCHNUR is a professor in the College of Science at George Mason University (GMU). Dr. Schnur retired from the Naval Research Laboratory (NRL) in 2008. His role at GMU is to stimulate new science of "impact" across department lines in GMU's College of Science and to initiate collaborations in the College of Engineering. As Director of the Center for Bio/Molecular Science and Engineering at NRL, Dr. Schnur provided scientific direction and management in the areas of complex bio/molecular systems with the aim of modifying structures in ways that will lead to the development of useful devices, techniques, and systems of use for the Navy and the Department of Defense. Dr. Schnur's research interests focus on understanding the relationship between the structure of molecules and observed macroscopic phenomena. This interest has led to his publications in the areas of critical phenomena, liquid crystals, picosecond spectroscopy, high-pressure and shock-related phenomena, self-assembly of biologically derived microstructures, and, recently, bio-based power sources bioinformatics, systems biology, and genomics. Dr. Schnur has more than 150 publications and issued patents, which have led to more than 3,000 citations; 20 of his more than 40 patents have produced or are currently producing royalties. He received his A.B. in chemistry from Rutgers University and his M.S. and Ph.D. in physical chemistry from Georgetown University.

LYLE H. SCHWARTZ (NAE) retired from government service in 2004, after 18 years as a member of the Senior Executive Service. In his last position, as Director of the Air Force Office of Scientific Research (AFOSR), he guided the management of the entire basic research investment for the U.S. Air Force. He led a staff of more than 200 scientists, engineers, and support people in Arlington, Virginia, and two foreign technology offices, in London and Tokyo. As Director, he was charged with maintaining the technological superiority of the Air Force. Each year, AFOSR selects, sponsors, and manages revolutionary basic research relevant to Air Force needs. The investment of AFOSR in basic research programs is distributed across 300 academic institutions, 145 industry contracts, and more than 150 research efforts within the Air Force Research Laboratory. Prior to becoming AFOSR's Director, Dr. Schwartz directed the AFOSR's Aerospace and Materials Sciences Directorate. From 1984 to 1997, he served as Director of the Materials Science and Engineering Laboratory at the National Institute of Standards and Technology. In that position, he managed programs in both structural and functional materials, with research emphasis ranging from basic to applied. From 1989 to 1997, he led the multiagency materials research coordination committee for the Office of Science and Technology Policy, and was responsible for the development of the Presidential Initiative on Advanced Materials and Processing launched in 1991. Previously, he taught and served as Director of the Materials Research Center at Northwestern University. He has written more than 85 technical papers and is a co-author of two textbooks in materials science and engineering. He received his B.S. in science engineering and Ph.D. in materials science from Northwestern University.

Appendix B

Importance of Alignment Between an Organization's Vision and Its People

At a National Research Council workshop on best practices in assessment of research and development organizations (a data-gathering workshop conducted under the auspices of the Panel for Review of Best Practices in Assessment of Research and Development Organizations) on March 19, 2012, Dr. John Sommerer, Head, Space Sector, and Johns Hopkins University Gilman Scholar, Johns Hopkins University Applied Physics Laboratory, highlighted the importance of assessing the alignment between an organization's vision and its people. The following is taken from the summary of his presentation titled "Assessing R&D Organizations: Perspectives on a Venn Diagram," provided in the NRC workshop summary.[1]

Alignment between an organization's vision and its people is addressed by asking the following questions, and any assessment, even of technical quality, needs this context: Does the parent stakeholder have a strategy articulated with clear milestones so that it can be internalized by the organization? Does the organization have a supportive strategy? Is there a clearly articulated vision of what the parent/organization is trying to achieve according to some milestones? Are all of these elements in synchrony? Are these strategies mutually supportive and updated? Are they good or bad strategies? Within this alignment, is the organization looking for first-mover advantages or second-mover advantages? What developments does the organization consider important to capture?

Vision is addressed by asking the following questions: Does the organization know what it wants to become (in 1-, 5-, 10-year frameworks)? What expertise is it trying to achieve? Acknowledging that strategy is about what one is going to do and *not* do, where does the organization choose to be a leader as opposed to a follower of fast developments? Does the organization have expertise in areas in which it desires to be a leader, and less in areas in which it desires to be a follower? Are the synergies nurtured? Are there exit strategies? Are there realistic stretch goals? Are there sufficient resources? A vision without resources is a hallucination.

The component of people is addressed by the following considerations: Human capital is fundamental. Innovation requires free energy—that is, giving researchers some latitude and discretion in their work. There is no hope for the future of an organization without free energy. Peer reviews, which measure competence, have been well defined, but it is more difficult to measure motivation and external engagement. There is a need for external engagement globally in order to innovate. An assessment of human capital includes asking: Are the people in the organization trying to become better?

The intersection of people with alignment is addressed by the following questions: Do the people know the strategy of the organization and its parent? Are there mechanisms by which the people can contribute to the strategy? Can they interact with the organization's customers? Are the leaders administrators or role models? What are their credentials and qualifications? Do they have a strategy to support the people? Does the organization assess and mentor the people? Does

---

[1] National Research Council, 2012. *Best Practices in Assessment of Research and Development Organizations—Summary of a Workshop*. The National Academies Press, Washington, D.C.

the organization have the will to release people who should not be there? Does the organization have a strategy and the resources for engagement with the external world and for encouraging such engagement? Is innovation welcomed, supported, protected? External engagement must be focused on the broad global community.

The intersection of vision with alignment is addressed by the following questions: Is there updating of the vision in response to changing external factors? Is there a process of self-assessment? Is there a list of lessons learned, and are they really learned, not just recorded? Is the self-assessment diligent, and does it have integrity? Is the assessment updated in acknowledgment of new strategies? There is need for both bottom-up and top-down assessment.

The intersection of people with vision is addressed by the following questions: Do the people know the vision? Can the people contribute to the vision? Does the R&D organization have a strategy and appropriate resources for engagement with the larger technical community, the commercial sector, and the global community? Is innovation welcomed, supported, and protected?

The intersection of vision, people, and alignment is addressed by examination of the organization's agility, flexibility, and adaptability in the face of changing pressures, budgets, and external contexts. This intersection needs to be consciously worked by staff and leadership, and it must be internalized.

Appendix C

Validating the Assessment

An organizational assessment results, fundamentally, in a set of predictions based on sampling of the characteristics of an organization. The predictions may be about the current characteristics of the organization's wider activities, staff, or processes based on the set examined; or they may be about the organization's future relevance or impact based on observed trends. This report identifies guidelines that may be considered and possible measurement methods applicable to the key characteristics of a research and development (R&D) organization. Some measures and criteria may be quantitative, and others may be qualitative, including anecdotal evidence. Just as an organization's activities can be assessed, so too can the assessment itself be assessed with respect to the validity of its measurement of quality, preparedness (management), and impact.

DEFINITION OF VALIDITY

Validity is the extent to which an assessment measures what it claims to measure. It is vital for an assessment to be valid in order for the results to be applied and interpreted accurately. Validity is not determined by a single statistic, but by a set of parameters that demonstrate the relationship between the assessment and that which it is intended to measure. There are four types of validity—content validity, criterion-related validity, construct validity, and face validity.

Content Validity

Content validity signifies that the items constituting an assessment represent the entire range of possible items that the assessment is intended to address. Individual assessment questions may be drawn from a large pool of items that cover a broad range of topics. For example, to achieve adequate content validity the project assessed will be shown to represent by some clearly defined strategy the wider pool of projects to which the conclusions of the assessment are also intended to apply; similarly with respect to surveys of an organization's customers.

In some instances when an assessment measures a characteristic that is difficult to define, expert judges may rate the relevance of items under consideration for the assessment. Items that are rated as strongly relevant by multiple judges may be included in the final assessment.

Criterion-related Validity

An assessment is said to have criterion-related validity when it has demonstrated its effectiveness in predicting criteria or indicators of the characteristics it intends to assess. There are different types of criterion validity: concurrent validity and predictive validity.

Concurrent validity is examined when the criterion measures are obtained at the same time as the assessment. This indicates the extent to which an assessment's measures accurately estimate the organization or project's current state with respect to the criterion. For example, on an assessment that measures current levels of customer satisfaction, the assessment would be said to have concurrent validity if it measured the current levels of satisfaction experienced by the organization's customers. Predictive validity refers to the extent to which the predictions yielded by an assessment turn out to be correct at some specified time in the future. For example, if an assessment yields the prediction that a certain avenue of research will yield a certain outcome, and that avenue is pursued, the accomplishment of the predicted outcome enhances the predictive validity of the assessment.

Construct Validity

An assessment has construct validity if the measures on the items assessed correlate well with measures of the same items performed by other assessment methods. For example, if quantitative measures of research productivity (e.g., papers published) correlate well with subjective measures (e.g., expert rating of the productivity of the research), this supports the construct validity of the assessment.

Face Validity

Face validity is the extent to which the participants in the assessment agree that it appears to be designed to measure what is intended to be measured. For example, if an assessment survey contains many questions perceived as irrelevant by the participants, its face validity will be low.

RELIABILITY OF THE ASSESSMENT

The validity of an assessment instrument is reliant on its reliability. Examples of reliability include inter-rater reliability, test-retest reliability, and parallel-forms reliability. Inter-rater reliability is the extent to which multiple raters of a given item agree. For example, if there is consensus among the members of a peer review committee, this indicates good inter-rater reliability. Test-retest reliability is the extent of agreement among repeated assessments of an item that has not changed between the assessments. Parallel-forms reliability is gauged by comparing two different assessments, created using different versions of the same assessment items and then randomly dividing the items into two separate tests. The two forms of the assessment would then be administered together, and the correlation of their results would indicate the parallel-forms reliability.

EFFICIENCY AND IMPACT OF THE ASSESSMENT

Efficiency and impact are also key aspects of an effective assessment. Factors related to the efficient conduct of an assessment include its cost in terms of money and time, burdens perceived by those being assessed, and timeliness of reported findings. Factors relating to the impact of an assessment include the extent to which the recipients of the assessment implement the advice provided in the assessment, the extent to which the assessment findings are distributed to those who should receive them, and the content of the feedback from those who receive the findings.

Appendix D

Example of Peer Advice During the Planning Phase of R&D

This appendix describes a successful example of an assessment of program content during the program planning phase. This assessment occurred when the Army Research Laboratory (ARL) involved invited experts during the formulation of a request for proposal for two multiscale modeling programs.

The ARL wished to initiate a complex, multi-organizational attack on problems deemed critical to future Army needs in areas requiring advanced materials. Influenced by developments in this field as reflected in the Defense Advanced Research Project Agency's Accelerated Insertion of Materials (DARPA AIM) program of the early 2000s and many subsequent efforts by other agencies, the ARL elected to focus on the development of computational tools to assist the effort. This multiscale modeling approach had recently been codified in a report of the National Research Council (NRC) on integrated computational materials engineering[1] and supported by the Office of Science and Technology Policy (OSTP) through its Materials Genome Initiative.

The ARL scoped its needs in two areas, electronics and high-speed deformation, and developed a conceptual framework for addressing these areas. Broadly speaking, the program envisioned was to include both extramural and intramural entities working in partnership. The funding of the extramural entities would arise from a competition stimulated through the conventional process of the release of a program announcement. The intramural programs would span the areas of structural materials and functional materials and would include extensive computational expertise. The centers of excellence for these three areas are located in three different directorates in the ARL, ensuring that a cross-organizational effort would be required.

Of great significance to the consideration of best practices was the inclusion of extramural assessment in the development of the program announcement of funding opportunity. Members of the NRC panels that conduct the usual peer assessments of the ARL were invited to participate in the public forum that preceded finalization of the call for proposals. This group of invited visitors offered, as individuals rather than as representatives of the NRC panels, observations that were deemed valuable by ARL management and that led directly to modification of the program scope and characterization in the final call for proposals. The selection of experts from a known and highly credible set of peer reviewers was a key factor in the successful planning activity described here—the experts were already familiar with those elements of the intramural research that would be impacted by the planned program; they knew the overall mission, structure, and operational procedures of the ARL and could assess how the new proposed activities might fit. Because these individuals were already well known and respected by ARL senior management, their advice perhaps carried with it more credibility than might have been accorded an ad hoc group of strangers, no matter what their credentials. To protect against conflict of interest, no member of the advisory group nor their colleagues were permitted to submit proposals for the multiscale modeling work.

---

[1] National Research Council, 2008. *Integrated Computational Materials Engineering: A Transformational Discipline for Improved Competitiveness and National Security*. The National Academies Press, Washington, D.C.

Appendix E

Relevant Statutes and Requirements Documents for U.S. Government Laboratories and U.S. Government Research

Along with the various annual budgetary legislative acts that support and often control research and development (R&D) at government agencies and together with such legislation as the Stevenson-Wydler Technology Innovation Act of 1980 (Public Law No. 96-480) which directs the technology transfer processes with the private sector, the Government Performance and Results Act of 1993 (GPRA; P.L. 103-62) is a significant law. This act directs all government agencies (with very few exceptions, such as the Central Intelligence Agency) to enunciate in quantifiable form the "outcomes" (as opposed to "outputs") of federal programs.

The GPRA was followed by the GPRA Modernization Act of 2010 (GPRAMA; P.L. 111-352), which stresses the Congress's intent and imposes additional requirements on federal agencies. The intent of this legislation was to provide evidence to the Congress, and so to the public, that there is true value resulting from the budgetary investment in federal programs. For many such programs, types of assessments involving the reporting of metrics for evaluating a program's outcomes in real time were doable, albeit somewhat difficult. However, for R&D, whose outcomes generally appear many years after funds have been expended, complying with GPRA's short-term assessments of outcomes is not possible.[1,2,3] Shorter-term outputs of R&D abound (papers, patents, citations, etc.), but outcomes in real time do not.

Another requirements document, specifically directed at federal R&D organizations, is the annual letter entitled "FY [-] Administration Research and Development Budget Priorities," which is jointly signed by the directors of the Office of Management and Budget and of the OSTP. It contains a listing of the top-priority R&D issues identified by the President of the United States and his senior science advisors. Until 2012, the letter's final paragraph, titled "Research and Development Investment Criteria," specified that all such programs will be assessed for quality, relevance, and performance, and indicates that budget decisions will be made based on these assessments.

---

[1] National Research Council, 2008. *Evaluating Research Efficiency in the U.S. Environmental Protection Agency*. The National Academies Press, Washington, D.C.

[2] G. Jordan and E. Malone, 2001. "Performance Assessment," Chapter 6 in *Management of Publicly Funded Science.* Department of Energy Office of Science. Available at http://www.science.doe.gov/sc-5/wren/benchmark.html.

[3] G. Jordan, L. Streit, and J. Binkley, 2003. Assessing and improving the effectiveness of national research laboratories. *IEEE Transactions on Engineering Management* 50(2).

Appendix F

The Army Research Laboratory's Process for Ensuring Relevance

For the Army Research Laboratory (ARL), the definition of *relevance* is how well a research and development (R&D) organization is performing its mission for the benefit of its stakeholders.[1,2] Thus, for an organization, or within the organization individual programs or even projects, to be relevant requires that their stakeholders be identified and the requirements or needs or desires of the stakeholders be made known. In the most fundamental sense this requires that the organization take several steps:

- *Identify its stakeholders and customers*: Does it know who it is supposed to be supporting?
- *Communicate with the stakeholders and customers:* Ask them what they want or need. Negotiate the technological terms to the extent necessary.
- *Direct an appropriate amount of effort toward fulfilling those needs*.
- *Continue to communicate with the stakeholders or customers as the work progresses:* This communication is needed to ensure that strategies and work are going in the right direction and that requirements have not changed.
- *Follow up after delivery of the final product:* This follow-up is needed to ensure that the stakeholders' or customers' needs have been fulfilled. If not, an examination is needed of ways to fix problems and develop improved processes for the future.

Based on the work of Edward B. Roberts at the Massachusetts Institute of Technology's Sloan School of Management, the ARL identified the following stakeholders. The Army's research, development, and engineering centers (RDECs), in particular, were considered immediate customers, whose missions were, in general terms, to develop materiel solutions to Army problems and to engineer these devices in preparation for production and fielding. The RDECs rely on the ARL to provide them with the basic and/or applied technology from which RDECs can carry out their development and engineering missions. Another significant group of ARL stakeholders consists of the various senior Army management organizations and funding activities. These include ARL's parent organizations, such as the Research, Development, and Engineering Command (RDECOM) and the Army Materiel Command (AMC), plus the senior Army management—the Army staff. The ARL relevance issues for these organizations revolve around how well the ARL is functioning as a support to the larger Army structure. The other major stakeholder is the end-item user—the soldier in the field, represented by Army program managers and program executive officers as well as the Army Training and Doctrine Command (TRADOC).

---

[1] E. Brown, 1997. *Measuring Performance at the Army Research Laboratory: The Performance Evaluation Construct.* Army Research Laboratory, Adelphi, Md.
[2] E. Brown, 1998. *Reinventing Government Research and Development: A Status Report on Management Initiatives and Reinvention Efforts at the Army Research Laboratory*. Report No. ARL-SR-57. Army Research Laboratory, Adelphi, Md.

In addition to these stakeholders identified above are such entities as Department of Defense (DOD) leadership, the Office of Science and Technology Policy (OSTP), the Office of Management and Budget (OMB), the Congress, and eventually the public.

When the ARL was established in 1993, the decision was made to develop processes by which it could show its relevance specifically to its stakeholders or customers and to the senior Army leadership. The chartering document required that several specific actions be taken with regard to its stakeholders and customers, the rationale being that concentrating most of the basic and applied materiel research in one organization with little to no development activity ran a very real risk of decoupling the newly formed corporate laboratory from the rest of the Army—that is, it ran the risk that ARL could lose its relevance. Thus, the ARL was required to direct at least 50 percent of its program resources toward stakeholder or customer requirements. This 50 percent requirement limited discretionary spending ability and thus provided a balance that did not stifle creativity but at the same time prevented an ivory tower effect from taking hold.

Meeting this requirement was accomplished by having the first ARL director meet with his counterparts in the RDECs; a memorandum of understanding (MOU) was executed in which each acknowledged this relationship. Then, for the directed tasks to be undertaken, ARL line managers were required to execute a Technical Program Annex (TPA) to the original MOU; the TPA was a one-page contract signed by the ARL line manager and one of his counterparts from a customer organization specifying the details of the work to be done. These details included a description of the specific task, the application of the end product by the customer, the amount of funds that the ARL would expend on this task, points of contact in both organizations, and any other information deemed necessary. The TPAs had to sum to at least half of ARL's base funding. This novel approach was deemed to be very unusual since, in effect, the ARL was asking its customers' permission to spend its own money. At the end of the year, the customer signatory on each TPA was surveyed to determine the level of satisfaction attained. If there were any complaints, or if the customer's needs were not satisfied, a senior ARL manager was required to contact the customer about how to correct the situation.

Another piece of this customer-relevance construct was a board of directors (BOD) created by that original charter. The BOD members were the directors of the RDECs that constituted ARL's customer base. The BOD would meet at least once a year at the ARL to review how well the laboratory had been serving their people. They reviewed the results of the year's TPA process as well as other processes. They then made comments, observations, suggestions, and recommendations as to the state of their relationship with the ARL and how, if necessary, improvements could be made. This entire process ensured that the laboratory would indeed be closely coupled and very relevant to its customer organizations.

In order to reach the senior leadership, the ARL conceived of a device called the Stakeholders' Advisory Board (SAB). The ARL director approached the commanding general (four-star) of the AMC, ARL's parent organization, who agreed to chair a group of senior Army staff. By the group's charter, the purpose of the SAB was to ensure that the ARL was closely coupled to the Army's vision of how it would fight in the future, ensure that the ARL would be responsive to and support the Army's senior leadership as it evolved the doctrine and requirements for the Army of the 21st century, and ensure that the ARL would be sustained as a valuable resource for providing the technological edge to shape the future Army.

The SAB's membership included the senior members of the Army staff—G1 (personnel administration), G2 (intelligence and security), G3 (operations), G4 (logistics), the deputy under secretary, and other members, all at the three-star or equivalent level. Reaching the soldier in the field personally was not possible. However, one member of the SAB was the deputy commander (three-star) of the TRADOC, which describes itself as the user's representative.

The SAB would meet at the ARL once a year for a half day, during which it would be briefed on ARL's activities of the past year and its projections and challenges for the upcoming year. The SAB would then be given a few short technology briefings on recent breakthroughs, a summary briefing from both the BOD and the Technical Assessment Board (TAB) of the National Academies' National Research Council in order to get a complete picture of all facets of the laboratory's activities, a tour of the facilities with demonstrations of technologies, and a working lunch at which time the chairman (the AMC commander) would solicit from each member opinions, critiques, suggestions, requests, and recommendations. These activities were captured as notes and action items, with the action items being scrupulously responded to over the next few months. Thus, the ARL was able to reach out and communicate its performance and achievements to assure this very senior group of stakeholders of its relevance.

This activity proved to be an enormously powerful management tool for ensuring the laboratory's relevance and performance as an Army asset. Demonstrating this close relationship with its customers and senior stakeholders served to drive ARL's performance. However, the SAB concept was abandoned after 2000. With the formation of RDECOM as a new command organization injected between the ARL and the AMC, it was no longer possible from a protocol standpoint to bypass the two-star RDECOM commander to the four-star AMC commander in order to continue the latter's chairing of the SAB, nor could the RDECOM commander summon all the three-star SAB members to assemble.

Appendix G

Examples of Stakeholders

DEPARTMENT OF DEFENSE R&D ORGANIZATIONS

Funding

The ultimate source of funding for the Department of Defense (DOD) laboratories is the Congress, but funding rarely comes by way of a single committee (usually there are as many as four committees, and sometimes more, involved in the process), and so there is rarely a single set of priorities and individuals to satisfy. The DOD organizations operate with direct appropriations as well as with funds from other entities that purchase services, such as the Defense Advanced Research Projects Agency, the Department of Homeland Security (DHS), and the intelligence community. In the Army, the research, development and engineering centers (RDECs) buy services from the Army Research Laboratory (ARL); in the case of the Naval Research Laboratory, considerable funding comes through the program management system at the Office of Naval Research; and in the case of the Air Force Research Laboratory (AFRL), program funds for basic research come from program managers at the Air Force Office of Scientific Research (AFOSR), which is itself a part of the AFRL. Each of these funding sources represents a stakeholder, and each must be satisfied in an appropriate manner or the funds will not flow to the named projects.

Recipient of the R&D

The customer receiving the benefits of products of research and development (R&D) carried out in a DOD laboratory is, ultimately, the warfighter. However, with the exception of the rapid-response actions taken by these organizations, usually in time of war, most products of the R&D find their way to the warfighter through a complex array of government and industrial developers. These too are customers and stakeholders, and each must be satisfied that the laboratory is responsive and is delivering with respect to its agreed-on part of the plan to transition items from research to fielded systems. Transition plans differ for the three services, depending on the specific technology; the plans change over time, as acquisition processes are continually reexamined and changed.

Oversight

Members of society, including taxpayers, are the ultimate stakeholders for the DOD laboratories. Their interests are represented within the government by such agencies as congressional committees, the Office of Management and Budget (using the Government Performance and Results Act of 1993 [GPRA], the Government Performance and Results Modernization Act of 2010 [GPRAMA] and the Program Assessment Rating Tool [PART]); the Occupational Safety and Health Administration (OSHA) to protect the workforce; the Environmental Protection Agency (EPA) to protect the environment; and others.

DEPARTMENT OF ENERGY LABORATORIES

Funding

The Department of Energy (DOE) laboratories system includes the national security laboratories at Sandia, Los Alamos, and Livermore and an array of laboratories that have specific or general missions ranging from primarily fundamental science to applied areas such as energy production and conservation. These DOE laboratories are government owned-contractor operated, and these contractors have local responsibility for laboratory management. Typically the funding for the DOE laboratories comes from more than one of the many offices within the DOE, and the program managers of these offices are major stakeholders. Significant funding also comes from other federal agencies. As with the DOD laboratories, Congress is the principal appropriator, and the views and desires of each of several subcommittees must be addressed.

Recipients of the R&D

The R&D of the DOE laboratories is transferred to society in many ways. The national security laboratories perform R&D aimed at protecting the nation's nuclear stockpile; they also perform R&D for customers such as the DOD and DHS. With the exception of classified material, dissemination of basic research follows standard publication and communications paths. Applied research can be done in collaboration with or can be directly transferred to some industry. For example, the work focused on energy efficiency has found its way to a vast array of manufacturing and energy-producing industries. In each instance, this means that a variety of customer groups must be identified and their needs addressed. Often, as when the R&D is a consequence of government-imposed regulation, the laboratories must work with other agencies (e.g., the EPA) as well as those affected industries and satisfy all of these customers.

Oversight

In addition to the usual oversight as identified above for the DOD, the DOE laboratories, often involved with safety and security issues, receive oversight from organizations within the agency charged with such issues.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST)

Funding

The National Institute of Standards and Technology (NIST) receives a large appropriation for intramural research and additional funding for extramural programs. The NIST laboratories also receive funding for new construction in a separate account. The report line for its intramural funding is much shorter than for many government laboratories, with a single authorization and appropriation committee in each house of Congress. The administrative report line is also quite short, as the director of the laboratory also serves as the Under Secretary of Commerce for Standards and Technology, with direct reporting to a cabinet-level appointee.

Recipients of the R&D

NIST, as the U.S. national measurement laboratory charged also with assisting U.S. commercial organizations to develop advanced technology, has an extremely wide responsibility. By providing the underlying technical capability that enables the entire system of standards and norms, NIST R&D supports a broad set of industries and government agencies. Direct customers may be the industries and government agencies themselves, or standards-setting organizations that act as intermediaries in the process. For each area of application, there is a different set of stakeholders, and the constant requirement for reexamining needs and addressing them through a broadly based program that includes basic research aimed at uncovering the next generation of measurement capabilities and technologies

Oversight

NIST is subject to the same oversight as that for other government laboratories. It is unique, however, in being the only government owned-government operated laboratory that operates a nuclear research reactor on its site. It is, therefore, subject to the oversight of the Nuclear Regulatory Commission and, of course, the citizenry and governments of neighboring communities in the Gaithersburg, Maryland, region.

Appendix H

Questions Pertaining to Assessment of Leadership and Management

Chapter 3, entitled, "Assessing Management," includes a discussion of the assessment of the leadership and management of a research and development (R&D) organization. It includes an abbreviated set of questions to be addressed in such an evaluation. This appendix presents a more detailed list of questions for assessing leadership and management. The questions are derived from the Burke-Litwin model of organizational leadership and management.[1]

1. External Environment: consumers, stakeholders, technology, scientific community, and others.
    a. Do senior leaders monitor the external environment and gather relevant information?
    b. To what extent do senior leaders distill this information into understandable if not simplified language and communicate it to all organizational members?
2. Mission
    a. Does the research organization have a clear statement of purpose and reason for being that show a clear link to the external environment?
    b. Does the mission statement consist of key components such as what the scope of the scientific fields is, who is being served, and what the research organization stands for—its philosophy, how it wants to be known in the scientific community and the world at large? The mission statement can effectively be considered a mini-business plan.
3. Strategy
    a. To what extent does the research organization have a clear and understandable process—a business model—for executing the mission? In other words, can any and all organization members define the organization's strategy in 25 words or less?
    b. Is management ensuring that researchers are working on important problems and that multiple approaches to solving those problems are being explored (when appropriate)?
    c. If the work is not characterized by having expertise that few if any other organizations have, is management making sure that there are appropriate interactions with external sources?
    d. Is the best use made of available capital funding (recognizing the constraints)?
    e. What long-range planning and forecasting are done for future needs in the areas of technology, staff, facilities, and equipment?
    f. Does the organization lead to successful new concepts, products, or processes that support the interests of its parent?
    g. How well has management of the organization identified its mission and vision as they are aligned with those of the parent?

---

[1] W.W. Burke and G.H. Litwin, 1992. A causal model of organizational performance and change. *Journal of Management* 18(3):523-545.

   h. How well has management of the organization identified its stakeholders and customers? How well does the organization stay in close contact with these groups, and how well does it respond to changing demands?
   i. How well does management identify the constraints on its flexibility, and how well does management work within those constraints and/or to overcome them?
4. Culture
   a. To what extent are organizational members clear about "how we do things here"?
   b. To what extent do organizational members agree that "this is a great place to work"?
   c. To what extent are organizational members treated with respect and dignity?
   d. Are differences among people respected and encouraged, or is the expectation one of bias and prejudice?
   e. Is conflict surfaced and managed, or is it avoided?
5. Leadership
   a. To what extent are senior leaders congruent and compatible with the mission of the organization? (Example: If the mission is more in the direction of basic science, leaders may need to be supportive if not laissez-faire as opposed to "hands-on" and highly directive—the scientists may need some freedom; if the mission is more in the direction of development and applied research, perhaps leaders need to be more declarative and directive.)
   b. To what extent is succession planning a common and accepted practice?
   c. To what extent are people who are selected for leadership chosen according to clear and evidence-based criteria such as self-awareness, relevant experience, learning agility (ability to learn from many different types of situations), vision, energy, and flexibility?
   d. To what extent is there a systematic process of developing people for leadership on the basis of different job experiences, reflection, mentoring, and coaching?
   e. To what extent does the leadership team inspire as opposed to simply managing? Does the team have the personal traits appropriate to this task—for example, are they compelling speakers, thoughtful listeners?
6. Management Practices
   a. To what extent do managers hold the people who report to them accountable for their various levels of performance?
   b. To what extent do managers make decisions in a timely manner?
   c. To what extent do managers provide clarity to those who report to them about goals and what is expected of them?
   d. To what extent does management provide recognition for high performance?
   e. To what extent do managers provide performance feedback to those who report to them?
   f. To what extent does management ensure that all staff receive adequate mentoring?
   g. Is creativity recognized?
   h. Does management provide channels for input from staff?

7. Structure
    a. To what extent is the research organization's structure (reporting and accountability system) congruent and in alignment with the strategy and mission?
8. Systems (Policies and Procedures)
    a. To what extent do members of the organization have the resources that they need to accomplish their work?
    b. To what extent is the reward structure supportive of good work on the part of all organization members?
    c. To what extent are organization members informed about matters that are pertinent to their work?
    d. Does staff evaluation use appropriate metrics?
9. Work Unit Climate: To what extent are organizational members at the work-unit level:
    a. Clear about what is expected of them?
    b. Informally recognized by their managers and peers for outstanding work?
    c. Mutually supported by one another?
    d. Communicated with in a timely fashion?
    e. Involved in decisions that directly affect them?
10. Motivation
    f. What is the level of job satisfaction of organizational members, particularly staff?
    g. How would one rate the degree of morale in the organization?
    h. To what extent are members of the organization engaged in and committed to their work responsibilities?
    i. Motivation will follow naturally if the following two issues are dealt with adequately in the organization:
        i. Task Requirements and individual skills and/or abilities: To what extent is there a good fit or match between organization members' talents and/or competencies and the job responsibilities that they hold?
        ii. Individual Needs and Values: To what degree is there congruence between individual members' beliefs and values and the culture of the organization?
11. Individual and Organizational Performance: What metrics and indicators are used for:
    j. Individual performance,
    k. Work-unit performance, and
    l. Performance of the organization as a whole?
    m. Are these appropriate?

Appendix I

Sample Crosscutting Assessments

As the nation's measurement laboratory, the National Institute of Standards and Technology (NIST) has always had the mission of working with industry to provide measurement techniques, data, and reference materials and to work with standards-setting organizations to provide the technical expertise required to assist them in the establishment of standards. Most of this activity has been carried out by small groups within the NIST laboratories, working with external partners in other government agencies, at universities, and in individual or collective groups of industrial organizations. The enormous array of these programs has always made it somewhat difficult for external assessments to be comprehensive.

Driven for the past several years by many external factors including the strong focus of the current administration on industrial competitiveness and manufacturing, NIST has been engaged in aligning its programs to fit these broad areas of government emphasis. Advanced manufacturing, as defined in the 2012 National Science and Technology Council (NSTC) document *A National Strategic Plan for Advanced Manufacturing,*[1] is impacted by already existing and newly initiated activities within NIST, with participation in one manner or another by all NIST organizational units. In 2011 the Director of NIST asked the National Research Council (NRC) of the National Academies to assemble a panel to assess the state of this vast enterprise within NIST.

NIST organized its program presentations for the NRC panel established to carry out the assessment to highlight three areas: Nanomanufacturing, Advanced Materials, and Smart Manufacturing. The NRC panel included members with expertise in the technical areas covered and, in many instances, having previous experience with NIST programs through direct interaction or in the course of service on an earlier NRC assessment. Although the presentations made to the panel were focused primarily on program scope and organization, some examples of actual past and current technical work were covered in sufficient detail to allow judgments to be made about program quality and the caliber of personnel.

It is too early to judge whether this exercise in comprehensive review of crosscutting program content will lead to significant impact on the NIST program. Nonetheless, the very fact of participation in such a review encouraged NIST presenters at all levels to think carefully about what they would say to outsiders. This internal, thoughtful assessment by capable people who care deeply about their programs and the reputation of the organization and of its staff can be the most important outcome of any assessment.

In a second example of a cross-organizational assessment, an NRC team of experts conducted a crosscutting review of the Army Research Laboratory's (ARL's) Autonomous Systems Program. Robotic systems have become ubiquitous in military operations, entering first as devices controlled from nearby by a single individual and now appearing as swarms controlled from a distance. With a vision toward fully autonomous systems with increasing capabilities, all of the military services and the Defense Advanced Research Projects Agency have launched extensive research and development (R&D) programs. Throughout the ARL directorates,

[1] National Science and Technology Council, 2012. *A National Strategic Plan for Advanced Manufacturing.* National Science and Technology Council, Washington, D.C.

programs have developed that focus primarily with respect to thrusts on materials, power sources on propulsion, protection, sensing, and human interfacing. Recognizing that results from these programs must ultimately be integrated to yield a final system, cross-organizational efforts have been launched, and the area of autonomous systems has been elevated to the level of senior-level visibility. This effort represents a very large component of the ARL science and technology program, yet it grew from many, rather separate, local programs and then was wrapped up for coordination. Many of these programs have been maturing for years. In some respects this ARL Autonomous Systems Program resembles in kind what might be found in the NIST Advanced Manufacturing Program several years from now. How is the Autonomous Systems Program at ARL progressing? What is the technical quality of the work being done? Are appropriate areas being covered? Are required cross-disciplinary issues being addressed? Those were the challenging questions posed by the ARL to the ARL Technical Advisory Board (ARLTAB) of the NRC.

The ARLTAB is charged with regular assessments of the technical quality of the work done at the ARL. Peer panels of experts visit the roughly disciplinary directorates of the ARL on an annual basis and view most of the ongoing program over a 2-year period. To address the one-time charge of assessing the Autonomous Systems Program, the NRC identified technical experts from the various existing panels and assembled them into a large panel with the range of skills covered by the program description. The panel heard overview talks presented by leaders of the various thrusts and integrated programs. They formed smaller groups and held simultaneous sessions, which were generally organized around posters. They had an opportunity to meet directly with scientists and engineers to get the pulse of current activity.

This challenging set of presentations and the format of the review created a first-time opportunity for program self-evaluation. Representatives of some parts of the program were learning more about other parts. This learning vehicle proved useful for researchers in a large, loosely connected set of R&D programs in separate buildings on two campuses. One important advantage in such a review is being able to assemble the panel from a large cadre of experts, each of whom has familiarity with the global mission of the laboratory and an in-depth knowledge of part of the laboratory from previous service on a disciplinary peer-review panel.

Both examples of cross-organizational programs cited above represented R&D within a large laboratory structured mainly along disciplinary lines. They spanned the time frame from program inception to maturity. Matrix management is not the overarching structure in the organizations cited, nor are most efforts cross-organizational. However, in each instance cited, important aspects of program assessment would not have occurred through a process organized according to disciplines. Also noteworthy, in each case cited, is that the crosscutting effort benefited significantly from the establishment of the peer group from a large cadre of experts who were already fully engaged in disciplinary assessment of elements of the cited programs. Panels organized for assessment of technical directorates are an ideal source of expertise for specialized, ad hoc panels targeted at cross-organizational issues.

Appendix J

Assessment at the National Institute of Standards and Technology, Army Research Laboratory, and Sandia National Laboratories

ASSESSMENT AT THE NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

Ongoing assessments have been provided for more than half a century by the National Academies' National Research Council (NRC) through contracts with the National Institute of Standards and Technology (NIST). The assessments are focused primarily on the technical and scientific quality of the scientific and technological research of the laboratories. Recognizing the broader mission of NIST, however, and in order to provide actionable feedback to management, the review panels have almost always received a multifaceted statement of task that typically has included adequacy of the facilities and/or budget and alignment with mission and/or desired impact.

The review process has been approached on a peer-to-peer basis, with the panel for each laboratory consisting of about 15 to 20 individuals, from academia, industry, and other scientific and engineering environments, divided into thematic subgroups of approximately 3 individuals aligned with laboratory divisions (generally about five or six divisions). NIST suggests membership for the panels, but the final selection is made by the NRC, reflecting expertise, balance, potential conflict of interest, diverse characteristics, and willingness to serve.

Each laboratory prepares and posts to a dedicated website review material and references to published papers and a set of projects for detailed review. Not all projects can be reviewed during each assessment, and so a negotiation determines which will be reviewed, primarily reflecting NIST management needs and concerns. Among the projects reviewed, each panel's goal is to portray an overall impression of the laboratory. Generally, the subgroups meet with their NIST counterparts for a day, attending presentations, taking tours, and engaging in interactive sessions with laboratory staff and managers. Then the panel as a whole meets for another day and a half, mostly in closed session, to deliberate on findings and define the contents of the assessment report. Observations are written by the subgroups and then blended into a cohesive report from the panel as a whole.

Panels as technical experts in the areas being reviewed generally depend on the experience, technical knowledge, and expertise of their members. The Director of NIST provides a charge to the assessment panels, identifying the general aspects of the laboratories' R&D that he wishes assessed (this may include, for example, the technical quality of the work, its impact, and the adequacy of supporting resources), but rigid metrics and criteria are not mandated for the review. The panels focus on evaluation of the quality of the research, the number of publications and quality of the journals, timeliness of the research, knowledge of relevant research being conducted elsewhere, and other metrics such as patents, presentations, professional committee service, extramural awards, and so on. Overall, however, the assessment is more accurately characterized as qualitative rather than quantitative.

ASSESSMENT AT THE ARMY RESEARCH LABORATORY

For nearly two decades, the Army Research Laboratory (ARL) has sponsored assessments by the National Research Council of the technical quality of the work at the laboratory's six directorates, which perform research and development (R&D) in the areas of computational and information sciences, human research and engineering, sensors and electron devices, vehicle technology, weapons and materials research, and survivability/lethality analysis. Accordingly, the NRC appoints six panels of experts that provide annual assessments of the scientific and technical quality of the work at the ARL. These panels provide biennial reports that summarize their findings and recommendations related to the quality and appropriateness of R&D for each of ARL's technical business areas. The reports are delivered to the Army sponsor and to Army and Department of Defense (DOD) stakeholders and are made available to the public.

The panels are charged to address the following items:

- Is the scientific quality of the research of comparable technical quality to that executed in leading federal, university, and/or industrial laboratories both nationally and internationally?
- Does the research program reflect a broad understanding of the underlying science and research conducted elsewhere?
- Does the research employ the appropriate laboratory equipment and/or numerical models?
- Are the qualifications of the research team compatible with the research challenge?
- Are the facilities and laboratory equipment state of the art?
- Does the research reflect an understanding of the Army's requirement for the research or the analysis?
- Are programs crafted to employ the appropriate mix of theory, computation, and experimentation?

In addition to the panel's addressing of the items listed above, the NRC selects from the panel membership individuals who form ad hoc groups to respond to specific issues identified by the Director of the ARL. The assessments are conducted by annual panel visits to the ARL facilities and review of supporting documentation describing technical projects and programs, equipment and facilities, and staff backgrounds and characteristics. The interactions between the panels and the ARL staff include dialogue during presentations, tours, demonstrations, poster sessions, and other means of presenting the ARL technical work, during which panel members seek clarification of facts and additional contextual information relevant to the assessments. Dialogue also occurs at the end of each review meeting, and panel chairs often engage in follow-up discussions with ARL staff. The purpose of these dialogues is to provide the opportunity for panel members to gather (and for ARL staff to provide) accurate, substantive, relevant information and to clarify contextual factors that may relate to the impressions under formulation by the panels; the dialogues are important so that subsequent findings, conclusions, and recommendations in the biennial report are based on accurate facts and adequate understanding.

As can be discerned from the examples of NIST and ARL assessments, there are various forms of peer review, each tailored to the specific type of organization. Considerable advantage to the process can be gained from the continuity of peer committee membership over time to allow for the assessment of change within the organization. Staggered appointments allow for some carryover of expertise in subsequent reviews.

ASSESSMENT AT THE SANDIA NATIONAL LABORATORIES

At a National Research Council workshop on best practices in assessment of research and development organizations on March 19, 2012 (a data-gathering workshop conducted under the auspices of the Panel on Review of Best Practices in Assessment of Research and Development Organizations), Dr. J. Stephen Rottler, Chief Technology Officer, Vice President, Science and Technology, at Sandia National Laboratories, described Sandia's assessment processes. The following is taken from the summary of his presentation, titled "*Assessing Sandia Research*," provided in the NRC workshop summary.[1] Jordan and coauthors also provide a detailed description of the assessment methods employed at Sandia.[2]

Sandia National Laboratories has undergone a continuous evolution in the assessment of quality, relevance, and impact, with quantitative assessment evolving into qualitative assessment that is informed by data. Organizations are complex systems, composed of interconnected parts. The properties of the whole organization are not necessarily perceived by looking at individual parts. Systems behave in nonlinear ways that are difficult to predict. Assessors must probe, watch behavior, probe, and watch behavior, iteratively, noting that the assessment impacts behaviors. There has been a need to shift from quantitative to qualitative assessment informed by data.

Organizations that traditionally have been stovepiped are increasingly evolving strategies and funding approaches that acknowledge the importance of multidisciplinary research organizations.

At Sandia, there are three assessment categories: (1) Self-assessments try to be objective, but they are inherently limited. All successful organizations have mature self-assessments that are objective and that promote responsive behaviors. (2) External peer reviews and visiting committees (external advisory boards) are used to examine quality, relevance, impact, and responsiveness to customers. (3) Benchmarking compares the organization being assessed to other organizations and is accomplished by formal assessments (through the visiting of other organizations) and less formal interactions as well.

Self-assessment at Sandia has become increasingly more formal and disciplined. Quarterly assessments are opportunities for leaders to examine with their teams whether their expectations about quality, relevance, and impact are being met. These assessments are performed at all levels of management.

Independent assessments are performed through a research advisory board that meets twice a year. The board is composed of senior individuals drawn from across academia and the public sector. The board is used in a broad sense to assess technical quality using external measures and comparison against other organizations. The assessment examines whether Sandia is meeting the criteria for its roles as fast follower or first researcher. It also examines the health

---

[1] National Research Council, 2012. *Best Practices in Assessment of Research and Development Organizations: Summary of a Workshop*. The National Academies Press, Washington, D.C.

[2] G. Jordan, P. Oelschaeger, A. Burns, R. Watkins, and T. Trucano, 2010. *Description of the Sandia National Laboratories Science, Technology and Engineering Metrics Process*. Sandia Report SAND2010-0388. Sandia National Laboratories, Albuquerque, N. Mex.

of the research environment and connections with internal and external customers. It elucidates what is working or what is getting in the way in terms of innovation. The board also meets with customers of the organization and examines the impacts of prior investments. It assesses whether investments have enabled the laboratories to continue fruitful work or to initiate new work. Laboratory Directed Research and Development (LDRD) funds are an important element of the laboratories. The laboratories' director is permitted to decide how the LDRD funds are allocated across projects consistent with the laboratories' mission. The National Nuclear Security Administration provides oversight for this program, which captures principal-investigator-generated ideas within the management context. The program includes five or six grand challenges projects; each of these larger projects has an assigned external advisory board. Historically, these larger projects have transitioned successfully to have impact within the laboratories or have achieved follow-on external funding—these impacts have been achieved with the help of the external advisory boards.

Assessments have been traditionally performed according to a balanced scorecard that guides the selection of data to support assessment decisions. Metrics are defined to assess three areas of measurement: value to customers, outputs, and inputs. Within each area, metrics are defined to support the assessment of what the organization is doing and how it is doing it. To assess value to customers, the value and impact in terms of leadership, stewardship, and mission satisfaction are addressed by examining measures of the effectiveness of strategic partnerships with industry and technical collaborators. To assess outputs, the excellence of scientific and technical advances is addressed within the context of management excellence, which involves measuring elements of the work environment and management assurance. To assess inputs, the capabilities of staff, technology infrastructure, and facilities are addressed by examining the science, technology, and engineering strategy through measurements of parameters indicative of the portfolio and the technical planning process.

The evolving assessment processes increasingly include the examination of qualitative factors informed by the quantitative data. The following elements are assessed: clarity, completeness, and alignment of the research strategy; alignment of the research with the organization's missions; the quality and innovation of the research; the vitality of the organization's scientists and engineers; and long- and short-term impacts of the research with respect to the organization's missions and to advancing the frontiers of science and engineering.

In summary, successful organizations and their assessors strive to be clear about the purposes of the assessment and its context, carefully decide what data to collect and what the assessment framework is, and link the assessment to the organization's concept of what makes a great organization.

Appendix K

Examples of Peer Review Conducted at Federal R&D Organizations

In addition to the reviews conducted at the National Institute of Standards and Technology (NIST), the Army Research Laboratory (ARL), and Sandia National Laboratories (SNL), described in Appendix J, peer reviews are conducted at other Department of Defense (DOD) laboratories and at laboratories at other federal agencies, either owned and operated by the government or contracted by the government to private entities. This summary of peer reviews conducted at other laboratories is taken from the report *Strengthening Technical Peer Review at the Army S&T Laboratories*.[1] Cozzens and coauthors also provide a summary of the assessment processes at NIST, ARL, the Department of Energy (DOE), Environmental Protection Agency (EPA), National Institutes of Health (NIH), Naval Research Laboratory (NRL), and the Agricultural Research Service (ARS), noting that relevance and quality are key assessment foci across these organizations.[2] A 1999 report of the U.S. General Accounting Office summarizes the peer-review practices for the following organizations, noting the significant variance across the organizations with respect to the amount and type of peer review applied: ARS, Forest Service, NIST, National Oceanic and Atmospheric Administration (NOAA), DOE, EPA, NIH, U.S. Geological Survey (USGS), NASA, National Science Foundation (NSF), and the Federal Aviation Administration (FAA).[3] Strauss and Loper define the assessment processes applied at the ARS, noting that the process has resulted in measureable improvements in the research.[4]

One type of review is that done by the National Academies' National Research Council (NRC). The NRC has boards that form and oversee study groups for all three military services. The boards may perform in-depth technical reviews.

The Defense Science Board (DSB) has reviewed the DOD laboratories—not in terms of specific technical work, but looking at DOD's policies for the laboratories, a higher level of review.

One aspect of external review is that the Federal Advisory Committee Act of 1972 (FACA; Public Law No. 92-463) may govern the operation. This means that meetings must be public, members of the public may make presentations to the committee, and members are appointed only with concurrence of agents of the President (typically the General Services Administration or agency heads). These requirements would have infringed on the traditional operations of the National Academies. After consideration, the FACA was amended to exempt the Academies from many of the requirements.

The services have traditionally convened their own advisory committees to look at technology issues—for example, the Army Science Board, the Air Force Science Advisory

---

[1] J. Lyons and R. Chait, 2009. *Strengthening Technical Peer Review at the Army S&T Laboratories.* National Defense University, Washington, D.C.
[2] S. Cozzens, B. Bozeman, and E. Brown, 2001. *Measuring and Ensuring Excellence in Government Laboratories: Practices in the United States.* Canadian Council of Science and Technology Advisors, Ottawa, Canada.
[3] U.S. General Accounting Office, 1999. *Peer Review Practices at Federal Science Agencies Vary.* GAO-RCED-99-99. U.S. General Accounting Office, Washington, D.C.
[4] M. Strauss and J. Loper, 2012. *Peer Review of Prospective Research Plans at the USDA Agricultural Research Service*. U.S. Department of Agriculture, Agricultural Research Service, Washington, D.C.

Board (AFSAB), the Naval Studies Board, and the Naval Research Advisory Committee. These groups differ in their operations. Some are involved in detailed technical studies; others restrict themselves to policy studies. All of them come under the FACA (with modified FACA rules applied to the Naval Studies Board). On occasion, ad hoc advisory committees may be set up, often by direction of the U.S. Congress. Congressional committees, the Government Accountability Office, and even the Library of Congress have also done policy studies of the laboratories.

Interviews with senior officials at a number of federal laboratories have revealed several different models for peer review. Some are formal review panels that come under the FACA; others are formal but are managed internally by the service or the laboratory being assessed.

An example of one managed by a service is the review by the Air Force Science Advisory Board. The AFSAB, a formal Federal Advisory Committee, reports to the Air Force Chief of Staff and the Secretary of the Air Force. Among other duties, the AFSAB reviews in depth the scientific work of the Air Force Research Laboratory (AFRL) on a 2-year cycle. The review covers the 6.1 through 6.3 work and also work sponsored by downstream users, such as program managers or executive officers. One panel has been created for each AFRL directorate. Panel members are selected by the AFSAB; all are external to the Air Force. Most are members of the AFSAB, but when necessary the AFSAB brings in consultants. At least one member of each panel must be a member of the National Academies. The AFSAB conducts reviews of each directorate, devoting a full week to each. The panels provide an exit briefing and a formal report. The reports are sometimes made available to the public, but some may be restricted (for official use only). The panels look at four factors for the programs: technical work, relevance for the near term (5 years), future impacts, and resources. When evaluating the 6.1 programs, the AFSAB does not look at near-term impacts. For the technical work, it evaluates technical innovation, technical rigor, productivity, and collaboration.

Two external groups have been established to look at various technical topics for the Navy: the Naval Research Advisory Committee and the National Research Council's Naval Studies Board created by the Chief of Naval Operations. Both groups consider the impact of technical developments on the future of naval forces. For a detailed technical review of its technical base research programs, the Naval Research Laboratory establishes and manages its own peer review. The NRL has seven focus areas: materials and chemistry, electronics, battlespace environment, undersea warfare, electromagnetic warfare, space research/space technology, and information technology. The technical review of the technical base covers about one-third of the total research program each year. The NRL selects members for the external review panels for each focus area. A panel typically has four to six members, drawn from academia and elsewhere, along with at least one member of the National Academies. The NRL asserts that these members are unbiased. The panels meet for from 2 to 4 days, with time for immersion in the laboratories and discussion with the staff. The panels give exit briefings to the NRL management. Subsequently, they submit a formal written report of about 10 to 15 pages, and the NRL responds to this report in writing. The panels evaluate the programs for scientific merit: they examine the research approach, the credentials of the staff, a project's relevance, equipment, and costs. Details of these differ for 6.1 and 6.2. programs. For a 6.1 program, the panel looks for work that seeks to expand the frontiers of known science; for a 6.2 program, the panel looks for whether the NRL is investigating and developing recent advances in science and technology. The panels have seven categories of evaluation containing metrics. For the customer-funded work, the criteria used are those of the customers.

The National Institutes of Health operates peer review by 19 formal advisory committees, termed boards of scientific counselors, one for each of the institutes. Their duties are described as follows: Boards of scientific counselors serve a dual function in providing expert scientific advice to scientific directors regarding particular employees and projects, and providing the NIH as a whole with an assessment of the overall quality of its intramural efforts. The Committee Management Office at NIH tracks these and many other NIH advisory committees. The office stays in contact with the FACA Office of the General Services Administration. The NIH process is the most formal type of quality review discovered in interviews for this study.

The National Security Laboratories of the Department of Energy—Los Alamos National Laboratory (LANL), Lawrence Livermore National Laboratory (LLNL), and Sandia National Laboratories (SNL)—are government-owned, contractor-operated facilities under the DOE. Interviews at these three laboratories revealed very similar processes for peer review. The Department of Energy's NNSA contracts for LANL and LLNL explicitly call for regular peer review; Sandia's contract contains similar language. LANL and LLNL are operated by a consortium of the University of California and the Bechtel Corporation, with additional partners from the Babcock and Wilcox Company and the Washington International Group. Two limited-liability corporations have been created to manage the contracts: Los Alamos National Security LLC and Lawrence Livermore National Security LLC. There is a joint board of governors for the two LLCs. The board has a Subcommittee on Science and Technology (the S&T Committee) that oversees and controls peer reviews of S&T at both laboratories. The S&T Committee makes the appointments to the peer-review panels upon nomination by the laboratories. The panels are independent and balanced. Each does both review and critique. The panels have 8 to 10 members, drawn from academia (about half), industry, and other laboratories; there are some University of California faculty, and some from other national laboratories. The operation of the panels does not come under the FACA.

Los Alamos National Laboratory conducts three kinds of review. The first is strictly concerned with the quality of scientific activity and is focused on capability areas rather than on scientific disciplines. The other reviews are on weapons design and customer programs. Typically, these capability areas are crosscuts from the discipline areas, such as weapons science and information science. The reviews cover eight capabilities per year; each capability area is reviewed every 3 years. The review panels meet for 3 or 4 days. (In earlier years these reviews were run strictly in-house and covered only scientific disciplines, not capabilities.) The panels also look at the adequacy of the laboratory infrastructure, the morale of the staff, and the research environment. The design reviews are done internally by DOE weapons design teams; the customer reviews look at quality, relevance, and performance against the mission. Customer reviews are set up by the laboratory subject to approval by the board of governors' S&T Committee.

Lawrence Livermore National Laboratory follows a similar assessment process. At LLNL, there are four principal disciplines and three major program areas. All are reviewed by peer panels. In addition, they conduct cross-cutting reviews of portfolios (for example, the National Ignition Facility) involving more than one of the seven areas. The panels consist of 10 to 15 members, drawn from academia, industry, and other laboratories. Members are selected by the LLNL directorates and vetted by the board of governors' S&T Committee. Selection factors include diversity on the panels and turnover of membership. The panels usually have a member from LANL and an observing auditor from NNSA. Panel reports contain both critique and advice. Verbal exit briefings are given to the head of the unit under review, as well as to senior managers and the Director of the LLNL. Reports of the panels are circulated within DOE but are restricted.

Sandia National Laboratories is operated under a contract between NNSA and Sandia Corporation, a wholly owned subsidiary of the Lockheed Martin Corporation. There are six Lockheed employees on the Sandia Corporation Board of Directors. The Corporation's S&T Committee is chaired by the chief technical officer of Lockheed. About half of the more than $2 billion budget at Sandia National Laboratories is for S&T. Peer review at Sandia is divided between S&T and nuclear weapons work. For S&T, the reviews are conducted under contract to the University of Texas (UT), which has two positions on the board of Sandia Corporation. Under supervision by the Sandia Corporation Board of Directors Subcommittee on S&T, the UT selects and convenes review panels. Both Sandia management and Sandia Corporation have input into the final selections. Each panel has 6 to 12 members from various disciplines, including physical sciences, computation, electronics, and materials. The UT draws some panel members from its faculty; other members are drawn from other sources. There is an external panel for each area of scientific and technical competence. Panels meet for multiple day sessions and cover individual projects and programs. They meet with staff members to assess morale and the research environment. They also may meet with groups of principal investigators at the program level, as well as with individual project leaders. The review results are reported to laboratory management and the board of directors' S&T Committee. Principal investigators receive the reports and must respond to panel critiques.

Nuclear weapons-related peer review at Sandia includes three kinds of internal reviews: design, management, and internal peer review (members from entities other than the design team under review). Reviews are assisted by a full-time office of assessment that reports directly to the laboratory director. The assessment staff is internal but separate from the program areas. There is also a standing panel for external independent review of these Sandia assessments. These assessments are a critical part of the regular certification of the U.S. nuclear weapons stockpile.

In 2010, the Congress mandated that the NNSA contract with the National Research Council to conduct a review of the quality of the science and engineering, as well as its management, at LANL, LLNL, and Sandia. That review is ongoing at the time of the publication of this report.

Entities that award grants—for example, the NSF, the Army Research Office (ARO), the ONR, and the AFOSR—do not operate laboratories and, therefore, conduct quality reviews in a different manner. The NSF is overseen by the National Science Board, which reports to the Congress. Each NSF directorate is monitored by a formal advisory committee that meets regularly to review performance. Periodically, under the auspices of the committee the grant folders are reviewed to ensure that procedures have been followed. Grant proposals are sent out to experts for evaluation; subsequently the folders are evaluated by NSF staff before the decision is made to award or not to award. For work in progress, grantees are visited on-site by NSF program managers. Quality is judged by these reviews, regular reports, and examination of publications. The ultimate indication of how well a grantee is doing is the renewal or termination of the grant. This is true for all of the granting agencies.

The ARO conducts two kinds of peer review concerning single investigator (SI) proposals for new work. One review evaluates technical merit. Typically the proposal is sent to external reviewers, mostly university faculty. The other review focuses on military relevance and is done by Army and DOD scientists and engineers. The SI grants are typically for 3 years. Two or more site visits are usually conducted during this time. Program managers audit grantees' presentations at scientific meetings and gauge audience reaction. The ARO receives formal annual reports and copies of all publications by its grantees. The ARO divisions are evaluated biennially by external boards of visitors, one for each division. The boards look at the overall portfolios, evaluating the strategic direction of the divisions and looking out for overlap with other programs in the DOD or elsewhere.

The AFOSR, a directorate within the AFRL, manages the 6.1 funds for the Air Force. Funding executed by AFRL internal research directorates is evaluated by the AFSAB during the biennial reviews of those directorates. The AFOSR as a whole is reviewed by the AFSAB every 2 years.

The Director of Defense Research and Engineering for a number of years conducted Technology Area Reviews and Assessments (TARA) that covered DOD basic research programs. Representatives of the service laboratories and operating commands participated in these reviews, along with outside experts. TARA reviews are no longer conducted.

Clearly there is no single, accepted best way to conduct peer review of federal S&T laboratories.

Appendix L

Metrics Applied by National Research Council Panels to Assessment of the Army Research Laboratory

Chapter 4, "Assessing Technical Quality," provides a discussion of metrics in the assessment of research and development (R&D) organizations. This appendix provides a set of assessment metrics and criteria applied by National Research Council (NRC) panels that review the Army Research Laboratory (ARL).

1. Community
    a. Papers in quality refereed journals and conference proceedings (and their citation index)
    b. Presentations and colloquia
    c. Participation in professional activities (society officers, conference committees, journal editors)
    d. Educational outreach (serving on graduate committees, teaching/lecturing, invited talks, mentoring students)
    e. Fellowships and awards (external and internal)
    f. Review panel participation (Army Research Office, National Science Foundation, Multidisciplinary University Research Initiative, etc.)
    g. Recruiting new talent into ARL
    h. Patents and intellectual property (IP) (and examples of how the patent or IP is used)
    i. Involvement in building an ARL-wide cross-directorate community
    j. Public recognition—for example, in the press and elsewhere for ARL research
2. Impact to Customers
    a. Documented transfer/transition of technology, concepts or program assistance from ARL to research, development, and engineering centers (RDECs) or RDEC contractors for both the long term and short term
    b. Direct funding from customers to support ARL activities
    c. Documented demand for ARL support or services (Is there competition for their support?)
    d. Customer involvement in directorate planning
    e. Participation in multidisciplinary, cross-directorate projects
    f. Surveys of customer base (direct information from customers on the value of ARL research)
3. Formulation of the Project's Goals and Plan
    a. Is there a clear tie to ARL Strategic Focus Areas, Strategic Plan, or other ARL need?
    b. Are tasks well defined to achieve objectives?
    c. Does the project plan clearly identify dependencies (i.e., successes depend on success of other activities within the project or outside developments)?

d. If the project is part of a wider activity, is the role of the investigators clear, and are the project tasks and objectives clearly linked to those of other related projects?
e. Are milestones identified, if they are appropriate? Do they appear feasible?
f. Are obstacles and challenges defined (technical, resources)?
g. Does the project represent an area in which application of ARL strengths is appropriate?

4. Methodology
    a. Are the hypotheses appropriately framed within the literature and theoretical context?
    b. Is there a clearly identified and appropriate process for performing required analyses, prototypes, models, simulations, tests, etc.?
    c. Are the methods (e.g., laboratory experiment, modeling/simulation, field testing, analysis) appropriate to the problems? Do these methods integrate?
    d. Is the choice of equipment/apparatus appropriate?
    e. Is the data collection and analysis methodology appropriate?
    f. Are conclusions supported by the results?
    g. Are proposed ideas for further study reasonable?
    h. Do the trade-offs between risk and potential gain appear reasonable?
    i. If the project demands technological or technical innovation, is that occurring?
    j. What stopping rules, if any, are being or should be applied?
5. Capabilities and Resources
    a. Are the qualifications and number of the staff (scientific, technical, administrative) appropriate to achieve success of the project?
    b. Is funding adequate to achieve success of the project?
    c. Is the state of the equipment and facilities adequate?
    d. If staff, funding, or equipment are not adequate, how might the project be triaged (what thrust should be emphasized, what sacrificed?) to best move toward its stated objectives?
    e. Does the laboratory sustain the technical capability to respond quickly to critical issues as they arise?
6. Responsiveness
    a. Have the issues and recommendations presented in the previous report been addressed?