APPLYING THE PRINCIPLES OF THE GOVERNMENT PERFORMANCE AND RESULTS ACT TO THE RESEARCH AND DEVELOPMENT FUNCTION

A Case Study
Submitted to the

Office of Management and Budget

by

THE ARMY RESEARCH LABORATORY

A GPRA Pilot Project for Performance Measurement

Dr. Edward A. Brown
Chief, Special Projects Office
GPRA Pilot Project Manager

June 5, 1996

I. INTRODUCTION: The Historical Context of ARL's Participation as a GPRA Pilot Project

The Army Research Laboratory (ARL) is a relatively new organization formed from the consolidation of seven older organizations (one dating back to 1816). These seven were formerly independent Army laboratories that were combined into a single "corporate laboratory" to provide the fundamental and applied research from which the future materiel for the nation's land warfare forces would spring. ARL has approximately 2600 full-time employees, about 1500 of which are degreed scientists or engineers. They are located at two principal campuses in Maryland (Adelphi and Aberdeen), as well as at several smaller sites around the country. The annual budget (mission funds plus reimbursable income) is about $450M. ARL is a subordinate element of the Army Materiel Command, a major command reporting to the Chief of Staff of the Army. ARL's mission and vision statements may be found at Appendix A.

When these seven former organizations were consolidated into ARL at the beginning of fiscal year 1993, the Acting Director at that time decided that beginning a new organization afforded the opportunity to introduce new management practices to the laboratory, so that it might function more in the manner of a private sector R&D organization. These new practices specifically included business planning and performance evaluation which, coincidentally, are the two basic components of the Government Performance and Results Act (GPRA).

Both planning and measuring from a management standpoint has long been studied for application to the world of R&D. Management literature is replete with writings going back many decades in both areas. However, no approaches, particularly in the area of performance evaluation, have ever been found to be completely satisfactory, especially when compared to the planning and measuring techniques used in other types of business functions. I have found this to be the case not only in government, but in the private sector as well, even though one would think that the "bottom line" is a suitable metric for all of a company's activities. Industry is struggling with the identical issues that GPRA lays before the in-house government research organizations.

The reasons for this difficulty are well known:

- o the likely outcomes of research are not usually calculable in advance;
- o the knowledge gained from research is not always of immediate value (i.e., there is a long time lag between inputs/outputs and outcomes), nor can that value always be determined or quantified far in advance;
- o results are more often serendipitous than predictable;
- o there is a high percentage of negative findings in research, which, while often very useful

themselves, are not valued as outcomes in the GPRA sense;

o the path by which the results of research are transitioned into products is not straightforward; and to speak the obvious truth,

o the unknown cannot be measured.

Even though the "D" of R&D is somewhat more amenable to planning and measuring than the "R," the difficulty of evaluating R&D is still daunting when compared to many production or service type functions.

Despite this background, and with many research managers carrying the scars of previous attempts at both planning and measuring in the research environment, ARL decided to try again to produce a business management system that could be tailored to the unique R&D environment and yet be logical, rational, meaningful, and above all, doable, and which would add value to the laboratory. ARL had invested almost two years of effort in the planning and measuring processes when GPRA came along; at that time, we decided that we were in as good a position as anyone to help the federal R&D community determine how it should be assessed by OMB and Congress. Therefore, ARL submitted a proposal to OMB to become a Phase I pilot project in March of 1994. OMB officially designated the laboratory a Pilot Project for Performance Measurement on 6 July 1994. This designation made ARL the only working R&D laboratory to be included in the GPRA process; as such, we have a responsibility to our research colleagues throughout the government for representing their interests as OMB and Congress work toward the government-wide implementation of the Act.

II. BUSINESS PLANNING AT ARL

A major initiative instituted at the time of ARL's activation was a fairly sophisticated business planning process. This was centered around a four-volume business plan, the production of which was tied to the annual budget cycle of the Defense Department:

Volume I, the Strategic Plan, looks out 10 to 15 years and, beginning with our mission and vision statements, assesses ARL's long-term goals, the environment, the users' long-term requirements, ARL's strengths and weaknesses, etc.

Volume II, the Long Range Plan, looks at the resourcing of the Strategic Plan over the next six years (the "POM" or Program Objective Memorandum period) in terms of funds, personnel, and facilities, based on projections of the DoD budget process.

Volume III, the Annual Performance Plan, takes the strategic goals from Volume I and the resources from Volume II, lays out the projects and resources for the coming year, and sets quantifiable metrics against which performance will be measured.

Volume IV, a corporate style Annual Report reports at the end of the year on the degree to which the proposed goals in the performance plan were accomplished.

When GPRA was enacted, it became clear that Volumes I, III, and IV were in alignment with the Act's requirements, with Volume II filling the role of a supporting document to Volume I. These plans are developed and updated at an annual cycle of quarterly meetings held by the ARL Director and his senior management team. The timing of these meetings and the resulting updated documents are tied to the annual budget cycle, so that the information is available for submission through Army headquarters to DoD in a timely fashion.

A. The ARL Strategic Plan

The strategic plan contains an introductory discussion about the ARL mission and vision, along with some broad statements about long-range policies. It continues with a brief discussion of the plan's relationship to other DoD and Army planning documents and processes. The last part of the introductory material is a discussion of Army requirements and needs in the near, mid, and far terms.

The heart of the document includes sections for each of ARL's primary mission areas (digitization and communications science, armor and armaments, soldier system technology, air and ground vehicles technology, and survivability and lethality analysis). For each of these, there is a statement of strategic intent followed by several long-term goals, each associated with a desired outcome. For example, for the Armor/Armaments mission area, the strategic intent states: "ARL will advance armor and armaments technologies to provide the 21st century soldier with the ability to overwhelm forces with minimum casualties and collateral damage and to survive against a range of current and future threats." One goal that follows from this is to "...create designs for advanced armors ... that offer enhanced combat survivability and significantly reduced weight," with the outcome of "Increased protection through lower cost/lower weight passive armor concepts ..."

This example illustrates the concept of the "planning thread," which starts from the organization's mission and vision statements and can be followed through the planning process to the individual scientist or engineer at the bench. In the example just cited, the goal of a lighter armor arises from the fact that the Army of the future will be primarily based in the United States rather than overseas and, thus, will need to be more readily deployable in time of crisis. The current main battle tank is too heavy for more than one at a time to be carried on an air transport. A significant reduction in the weight of the armor would allow two to be carried per plane, thus halving the number of transports and the amount of time required to deploy an armored unit overseas. Based on this strategic goal, we have a program to develop a composite armor material that will provide ballistic protection equal to current armors, but weigh 25 percent less. From such a strategic goal, we can then derive a host of short-term technical goals that can be expressed in the annual performance plans for the next

several years. Thus, the materials scientist working on the chemistry of composites can "follow the planning thread" to see where his work connects to the overall mission of ARL and to the Army.

This approach is obviously more amenable to requirements-directed work, that is, applied research or development. However, it can also be applied, at least in a limited fashion, to our basic research program. The assumption is that all research, no matter how basic, is undertaken for some purpose and with some goal in mind, no matter how far removed from an end item application.

The strategic plan is discussed and developed or modified every year in late January, at the second quarterly (or "Q2") meeting of the Director and his senior staff and Directorate heads (SES level). The timing allows for major changes in the organization's goals to be submitted to the following year's budget preparation. The effort is very much top management driven, with all the senior managers bringing to the table their experiences and contacts from the past year with headquarters, users, and customer personnel. The Director's staff then transforms the results of this meeting into the actual document. (The ARL Strategic Plan may be viewed at **http://www.arl.mil/EA/Publications.html.**)

B. The ARL Performance Plan and Performance Report

The performance plan is structured in three sections. The first section deals with management initiatives and goals. About a dozen goals are described for the coming fiscal year, dealing with things like the implementation of a new concept of operations for ARL, progress on two major construction projects, and other significant management undertakings. The second section covers the technical objectives. Here the heads of the 11 technical directorates each present their top five objectives to be achieved in the upcoming year's technical program. These objectives are only one or two sentences long and, to the greatest extent possible, state some quantifiable attribute or parameter to be attained (e.g., "use nanofabrication techniques to fabricate a 2x2 prototype spatial light modulator for optical image and signal processing. The modulator will exhibit picosecond switching speed with an average contrast ratio of 100 when switched with only 10 V."). Again, we note that this type of quantitative statement is certainly applicable to applied work, but only in a more limited sense to basic research tasks. The third section of the plan lays out the year's goals for the metrics that the Director wants to emphasize (metrics are discussed below).

An ARL-wide summary version of the performance plan was submitted to OMB in compliance with the GPRA pilot process. (The performance plan will soon be available for viewing at the web site shown above.) We also prepare a second version that contains an additional breakout of objectives and goals by directorate. Because of its greater size and emphasis on detail, this version is for internal use only.

The annual performance report is also structured in three parts. The first part is a brief discussion of the overall state of the lab and then a fairly extensive review of the past year's accomplishments,

technical and otherwise. The annual report correlates these accomplishments with each of the objectives set forth in the performance plan published at the beginning of the fiscal year. This presentation includes the quantitative goals that were met (or not met!). For example, in our FY93 performance plan (pre-GPRA) one goal stated "Demonstrate fully *automatic synthesis tool for creating VLSI chips*." The FY93 Annual Report published 14 months later reported the accomplishment: "Expanded the institutionalization process for electronic design automation methodology through the development of a *computer-aided design synthesis tool for VLSI circuits*."

The second part of the annual report contains a review and analysis of the lab's performance over the past year. Fiscal, personnel, and facilities data are presented graphically, with explanatory notes and discussion. This section presents the metrics, again in terms of the goals from the performance plan and the accomplishments for the year, along with appropriate discussion of the results and any variances from the plan. The final section of the report (published as a separate volume) is a listing of all the patents received and all the papers and reports published during the year. The annual report may also be found at **http://www.arl.mil/EA/Publications.html.**

Volume II of the business plan, the Long-Range Plan, is not submitted as part of the GPRA pilot process. It is a fairly weighty document for internal use only. It describes in great detail the fiscal and personnel resourcing of the strategic plan. It analyzes the finances of each directorate at the project level for the next six years and identifies potential problem areas requiring senior management's attention.

III. PERFORMANCE EVALUATION AT ARL

A second management innovation introduced just before ARL's activation was a methodology to evaluate the health and performance of the laboratory. R&D performance evaluation has been a problem that has evaded a satisfactory solution for many decades. Efforts in this area have fallen into one of several approaches:

- retrospective anecdotal reviews, which are usually impressive and meaningful to the technical community, but not to the senior agency leadership who want to be able to define specific progress toward goals in real time;
- peer reviews, which are, again, common practice in the technical community and are somewhat more acceptable to senior managers, although still not completely satisfying; and
- metrics, which usually means things easily countable like patents and papers; these are usually not at all satisfactory in evaluating the true worth of a technical program or organization. To make matters worse, from time to time, complex schemes have been proposed that take such "measures" and perform arcane algorithmic operations on them that often obscure whatever little meaning they may have originally had.

æ

In an attempt to cope with this long-standing problem, ARL developed its Performance Evaluation Construct, which we believe is a rational and logical approach to answering the questions of how well the lab is performing and the degree of its health. It is a semi-quantitative approach that requires the Director to take a broad view of many different factors, some numeric and some descriptive, and then to personally integrate them into a picture that he can present to a variety of audiences (including stakeholders) in whatever format is appropriate at the moment. We concluded that there are three questions or principal areas of interest to which the Construct responds:

o Relevance -- Does the work being performed respond to some bona fide requirement of a customer? That is, does anyone care about the work we are doing?

o Productivity -- On any given project, or for the lab as a whole, is progress being made toward some specified goal at an acceptable rate? That is, are we giving the customer some product of use to him and in a timely fashion?

o Quality -- Is the work being done at a level that could be called at or beyond the state of the art? That is, is ARL a world-class institution doing world-class work?

This philosophical framework of describing results in realistic and readily understandable terms demonstrates the relevance of the Construct to the evaluation requirements of GPRA. The Construct is built on three pillars, which, in various combinations and to various degrees, respond to those three areas of interest. These three pillars are peer review, metrics, and customer feedback. As indicated in Figure 1, each of these pillars is more or less useful for evaluating different aspects of a program depending on the precise nature of the work.

A. Peer Review: As noted before, contemporary or retrospective peer review is a generally accepted approach to performance evaluation throughout the world of science and technology. However, in and of itself, it has certain limitations, such as the breadth, depth, and independence of the reviewers. It very specifically responds to the area of quality and, to a more limited extent, productivity; it is not expected to speak to the area of relevance with any specificity. (Prospective peer reviews, often used as tools in grant selection, can respond to relevance if the reviewing body so desires and is appropriately constituted.) A panel of recognized experts in a scientific field is well-equipped to make judgements on the technical quality of the programs and the technical staff, and on the relationship of the facilities and equipment to the state of the art. These kinds of judgements are most important for the leadership of a laboratory if the programs are to be properly executed.

Taking a cue from the National Institute of Standards and Technology (NIST), ARL has contracted with the National Research Council (NRC) of the National Academies of Science and Engineering to assemble an ARL Technical Assessment Board (TAB). The TAB consists of about a dozen individuals of international reputation. Under this board are six panels of 8 to 10 people each, also of high repute within the technical community. These panels will review each of our primary mission areas. The Board, with its panels, will provide an appraisal of the scientific and technical efforts of

ARL. It is specifically enjoined from making judgements on the programmatic structure of our work, partly because it is not equipped to perform this function, and partly because there are numerous other channels through which ARL receives such guidance. Because of the size and diversity of the ARL technical program, the panels will review the total program in depth over a two- or three-year period, reporting annually to the TAB on the one third to one half of the program reviewed. Its product will be a descriptive assessment in the form of a written report that will be published by the NRC. The TAB will also meet annually with the Director to provide an informal report of its findings before the publication of the report, and to receive guidance on the Director's desires for special areas of emphasis in the next year's review. The initial contract with the NRC is for three years at approximately $650K per year. From these funds, the NRC will pay the expenses of all the members of the Board and the panels, supply a full-time Executive Director and appropriate clerical and support staff, and provide for the logistics and publication requirements of the Board. This approach has the advantages of providing the independence and stature required of a peer review, removing any possible conflicts of interest from ARL, and stamping the endeavor with the NRC's imprimatur.

B. Metrics: The things that are "countable" in the world of R&D, for the most part, speak only peripherally to the three areas of interest discussed above. Nevertheless, because of the additional information they can provide, specifically as indicators of the functional health of the organization, they are judged to be necessary in any evaluation process. However, it is imperative that they not be taken out of context nor given more importance than they warrant. For this reason ARL has eschewed the use of any sort of weighting schemes or algorithmic manipulations of these data. We are, however, considering attaching some sort of simple prioritization scheme (e.g., high, medium, low) to certain of the metrics to indicate to the workforce the significance placed on them by our Director. He is well able to personally integrate the values of the different metrics, and then to report them in whatever format might be appropriate for the particular audience of the moment. Overall aggregation into some sort of "score" for the lab does not seem to be productive.

Accepting these limitations, ARL assembled a set of 59 metrics for consideration. As might be expected, most of these were input metrics, with several output metrics; metrics for outcomes were, not unexpectedly, absent. These metrics naturally aligned themselves with the four elements of the ARL vision statement. While none of these groupings relate directly to the requirements of GPRA in terms of measuring outcomes, they do provide useful information about the health of the lab. Furthermore, depending on how some of these numbers are viewed, critical insights can be gained. For example, two of the metrics in the personnel area are number of staff members holding doctorates and the age distribution of the staff. While it is obvious that the first of these would be of concern to a laboratory, age data might seem of more questionable value. Yet, when one ARL directorate plotted doctorates against the age of its workforce, it found that they had no PhDs on the staff that were under 40 years old. This caused the head of that directorate to re-think his hiring and staff development strategies.

The most important of these metrics obviously concern the delivery of our product to our customers:

æ

specifically, items delivered and programs completed. For these metrics we went beyond a simple "met" or "not met" in our scoring. We used a Red-Yellow-Green methodology to allow us to indicate progress short of completion (Yellow). These categories are scored 0, 0.5 and 1 respectively so that, for example, a directorate's Top 5 Deliverables might be scored G, G, Y, Y, G = 1 + 1 + 0.5 + 0.5 + 1 = 4 out of 5 or 80 percent. Our Director set each directorate's goal for this metric at 85 percent purposely a little lower than perfection in order to allow the technical staff some room to "stretch" without fear of failure.

In general, 59 metrics are far, far too many to deal with regularly. The data can be collected and used in various formats, but to be useful a more compact set must be defined and then brought to senior management's attention. To do this, the Director decided on 15 of the 59 that were of special importance to him and which he wished to use as "levers" to move the organization in certain directions. These 15 were singled out on the basis of the Director's judgement. After a long and distinguished career of managing major R&D organizations, the Director had developed his vision of what a world-class laboratory should look like, and these particular metrics were part of that vision. These deserve some additional discussion.

In order for a metric system to be successful and have some meaning to, and impact on, an organization, it must be coupled to the compensation system of those responsible for meeting the goals. In ARL's case, this means the heads of the 11 technical directorates who report directly to the ARL Director. The Director has placed goals for these 15 metrics in the performance appraisals of these directorate heads, the values of which he negotiates with them individually, taking into account the nature of the work in the specific directorate. For instance, a directorate with a more basic research mission would be expected to have a higher goal for refereed journal articles and a lower goal for patents, while a more engineering-oriented directorate would have the opposite emphasis. Our Survivability/Lethality Analysis Directorate, which concentrates on systems testing for Program Managers, would be expected to have neither papers nor patents in any substantial quantity. Its metrics would include test reports delivered to customers.

Although it is true that the career of an SES-level manager will not rise or fall solely on whether his directorate makes its quota of patents or papers, placing these metrics in his performance standards does draw his attention to them and has had a noticeable effect on the overall performance of ARL in these selected areas; this was, in the end, the Director's intention. For instance, based on his years of experience as a director of several prestigious research organizations (most recently before ARL, he was the Director of NIST), he believed that a world-class research institution should have about 40 percent of its technical staff at the doctorate level. When he came to ARL, he found that this number was at 22 percent. Thus, he included this particular metric among the 15 for special attention.

At the end of the year, the statistics are gathered and reported to the Director, who takes them into

account as he prepares the performance appraisals of his senior managers. These figures are also rolled into an overall ARL picture. The effect of this process is demonstrated by the example of the percentage of doctorates on the technical staff, which has risen from 22 to 26 percent in two years. Especially in light of the various downsizing efforts (hiring restrictions, early retirements, buyouts) that the lab is undergoing, this is real progress toward the long-term goal of 40 percent.

Obviously, the long-term outcomes of ARL's programs will stem from its technical accomplishments. While these are usually considered in terms of shorter term outputs, they are among the more important of the metrics that we can collect. These include a variety of "countables" that are defined differently for the different directorates. However, they all boil down to counting how many of the specific technical goals set at the beginning of the fiscal year were actually accomplished. Keep in mind that the quality and value of these individual goals are not determined by the metrics, but rather by the other two pillars of the Construct -- peer review and customer feedback. However, the metrics provide a handy scorecard to keep track of how well the head of a directorate is doing in fulfilling his annual obligations.

Many of the other metrics are directed toward ascertaining the technical atmosphere and operational health of the laboratory. An example is a group of metrics called "esteem factors." The assumption is that in a world-class laboratory of over 2600 people, more than half of whom are scientists or engineers, there should be a certain number of fellows of national societies, adjunct professors at prestigious universities, winners of major awards, etc. Such goals are determined by comparison with other world-class laboratories. While such statistics may not be under the direct control of management, the lack of any such honors among the staff would certainly be an indication that the scientific atmosphere at the laboratory is somehow lacking and needs attention. In a similar manner, the various input metrics (fiscal performance, facilities, personnel data) provide indicators of organizational health to the Director.

The 44 metrics not regularly reported to the Director are tracked as part of normal business procedures by the various functional managers. They are not surfaced to senior management unless they begin to fall outside of some envelope and it is deemed that some form of executive attention is required.

The linkage between all these metrics and the outcomes of a laboratory's endeavors is certainly tenuous, but no more so than in industry. A great deal of work has been published in this area, much of it summarized in publications by the Industrial Research Institute. This body of work shows that although various metrics can be defined in financial terms, for R&D the collection and reliability of these metrics, and their relationship to the company's profits, are extremely problematic.

A complete listing of ARL's metrics, including their linkage to the ARL vision statement, is presented in Appendix B to this paper.

æ

C. Customer Feedback: The third pillar of the Construct applies the concept of a customer to the research effort. While this is certainly familiar in the world of development and production, it is somewhat foreign to the world of research. Nevertheless, if someone is being paid for some work, research or otherwise, the payer presumably expects a product from the payee. That payer is the customer.

ARL has developed this notion by the use of a model of the "stakeholders" for the research enterprise, developed by Professor Edward B. Roberts at the MIT Sloan School of Management. According to this model in industry there are three groups of stakeholders for the research effort: two within the firm and one without. The most obvious group within the firm is the development and/or manufacturing and/or marketing departments, who depend on the output of the research group for the raw material (ideas and technologies) that will become the basis for new product lines. The external stakeholder is, obviously, the end user or customer of the firm. However, the other internal stakeholder, which might not be as obvious, is the senior leadership of the firm, the Chief Executive, Chief Financial, or Chief Technical Officer, who needs the concepts under investigation now in the research arm of the firm to enable him to set the long-term strategic vectors for the company.

Translating this model to Army acquisition, we have customers for the immediate short-term products of our efforts: the Research, Development and Engineering Centers (RDECs) and the various system Program Managers and Program Executive Officers (PMs/PEOs). These customers require specific products from the research organizations to fit into their development programs. Like the private sector, ARL also has an ultimate end-item user -- the individual soldier in the fighting units. While we may not deliver a research product to this ultimate customer, we must be ever mindful of his real-world needs and requirements as we conduct our research on concepts for systems that might appear on a battlefield 10 or 20 years from now. (Part of maintaining this awareness of the soldier's needs is gained through the Greening program mentioned in the previous section.) Finally, again in parallel with industry, we have a senior leadership that also has to look farther into the future than a PM, for instance, who is concentrating on fielding a system in the next year or two. The Chief of Staff of the Army, for example, must shape the fundamental concepts of our fighting forces for the next several decades. He becomes a stakeholder in the laboratory, since the most advanced technologies being worked on today will shape not only the systems that appear on the future battlefield, but the combat doctrine and force structure as well.

A dramatic illustration of the impact of ARL's work on this last category of stakeholder occurred when we were visited by the then Chief of Staff, General Gordon R. Sullivan. As part of a tour around our facilities, we showed him a demonstration of a new fuzing concept that would allow first-round registration of artillery fire. We certainly realized the tactical value of such a technological innovation. However, General Sullivan immediately realized the much greater impact that this fuze could have. First-round registration of fires would dramatically reduce the requirement for ammunition, with a concurrent reduction in depots, trucks, logistics trains, ships, and even

ammunition plants. This could result in major potential savings in fiscal and personnel resources. It could also have a major impact on the force structure of the support troops required in the combat theater of operations. Such a potential from the technology base helps the senior leadership to restructure their long-term strategy to manage the Army.

For that segment of our customers to whom we deliver specific items, we use a series of feedback questionnaires, not unlike what industry uses, to sense the degree of their satisfaction. We have completed two annual cycles of questionnaires to our internal development customers: to the RDECs for whom we do work at no cost, and to other agencies that come to us on a reimbursable basis for specific tasks. These questionnaires (see Appendix C) are sent to the individuals in these other organizations who were responsible for specifying the product to be delivered. After the work is finished, each customer is asked to rate our performance on a 1-to-5 scale in terms of the quality, timeliness, utility, etc., of the deliverable. There is also a place for comments. Our Director has established a policy that any rating of less than 3 on any item, or any negative comment, must be responded to by the appropriate directorate head within one week. Aside from this feedback process itself, the questionnaire scores are also a part of the metrics process and are included as a basis for the senior leaders' performance appraisals.

We do not use questionnaires on the other two groups of stakeholders -- the end users and the senior Army leadership -- since we do not deliver a specific identifiable product directly to either of these groups. However, we are in the process of establishing a Stakeholders' Advisory Board (SAB) that will bring together, at the three-star level, the senior leadership and user representatives into a forum that will enable ARL to receive first-hand guidance and feedback from these other two very important segments of our stakeholders.

This survey process is obviously more applicable to the applied research that we do -- the requirements-driven R&D -- than to the more basic, opportunity-driven research. Peer review, on the other hand, is more applicable to our basic research work. Metrics are applicable with varying emphasis, depending on the metric and when it is applied. In other words, the three pillars of the Performance Evaluation Construct are given different weights when applied at different points in the R&D life cycle. An example of this variation is given in Appendix D.

The three pillars of the ARL Performance Evaluation Construct are related in varying ways to the area of interest for evaluation: quality, relevance, and productivity. Peer review, as realized by the Technical Assessment Board, provides feedback as to the quality of the technical program. The Stakeholders' Advisory Board will provide feedback on the programmatic and managerial performance of ARL, which applies to relevance and productivity. Our survey process also inquires into relevance and productivity from the standpoint of those customers who receive specific deliverables from us. Although there may be quantitative aspects to this information, neither it nor our system of metrics, is reduced to a single ARL "score." Rather, they are digested and integrated

by our Director for use in a variety of ways. This approach is in consonance with the basic principles of the National Performance Review, which aims to place the responsibility for managing with the senior managers and free them from mindless, bureaucratic rote. Figure 2 displays our implementation of the ARL Performance Evaluation Construct shown in Figure 1.

IV. ISSUES AND LESSONS LEARNED

A major issue presented by the GPRA to many government organizations is the difference in scale between agencies and how the requirements of the Act will be translated to the working level organizations within agencies. The Act says that the agencies responsible for reporting these various plans are defined as cabinet-level organizations. Thus, in our case, the reporting agency is the Department of Defense. Depending on how one counts the levels in the chain of command, ARL is between 5 and 10 levels down from the level where the strategic and performance plans will be drawn up. This means that, for all practical purposes, ARL will not be very visible in the aggregation of data at the DoD level. Some of the other civilian agencies, which are much smaller than DoD, may well have their R&D functions exposed in somewhat more detail at the OMB and Congressional level. If GPRA is to have value at the working level as well for the Congress, a great deal of individual initiative and "buy in" will be required by managers at all levels of government, not just at the cabinet level.

Along the way we have gathered other insights into the GPRA and related processes:

1. A new mindset is required, especially in the world of research. Thinking in terms of outcomes rather than inputs or outputs, effectiveness as well as efficiency, and customer satisfaction, is somewhat alien to the R&D business. However, for GPRA these concepts must be understood, translated into workable constructs for the R&D community, and embraced. Training may help, but in some cases a "religious conversion" will be required.

2. Senior management must buy into the process totally. This acceptance cannot be driven from the bottom up, since it is intimately connected with the very heart of an organization's business. If the senior leadership does not embrace this philosophy as part of their planning process, they will fail in their attempts to use planning and measuring techniques as a part of the management of their organizations.

3. Participation in this type of planning and measuring requires a significant investment in personnel, funds, and time. Despite the original protestations of those who testified before Congress for the establishment of the Act, it is very labor intensive at all levels of the organization. ARL dedicates many manyears of to this process, and has a number of full-time permanent positions established to perform the planning and measuring functions. If this effort were being expended solely to satisfy GPRA, it would have met with a great deal of resistance at ARL. However, since we began the

planning and measuring efforts independently for reasons of good business practice, GPRA has not become a major additional factor in the cost equation for us. But it should not be forgotten that, however motivated, this kind of management does not come cheap.

One might consider that an organization takes political or bureaucratic risks or "costs" by exposing itself to such in-depth scrutiny through publishing this array of planning and reporting documents. Even if such scrutiny were not required by statute, ARL believes that any organization that aspires to be world class must be able withstand, and learn from critiques by its peers and stakeholders. Otherwise, the entire endeavor is pointless.

As a side note, except for the National Research Council's role in our peer review system, ARL chose not to use contractors to support this effort, since planning is an inherently governmental function. We felt that no contractor could know as much about our own business as we do. It is also arguable whether we could get the same level of effort from a contractor for much less than we are expending ourselves. However, contractual support might be effectively used by some organizations for training in the mechanics of business planning, for developing data bases to handle the metrics, and perhaps for other similar support tasks.

4. A competent corporate information system is essential to managing the performance evaluation process. We have learned this the hard way, since we do not have one at this time, and doing the metrics work alone is enormously time-consuming.

5. The application of these techniques to basic research is more difficult than it is to applied research. Both planning and evaluating basic research (i.e., research that is not requirements-directed) is certainly recognized as very difficult for the R&D community. At ARL we have approached the problem by assuming that all work is done for some purpose and for somebody. In basic research, that purpose and person may not be immediately evident. It may be the Director himself, or it may be "second order": an output from some basic research may be an input to some applied research that does have an identifiable application and customer. Nevertheless, if a purpose and a customer can be defined, then the rest of the process follows more or less directly.

6. Metrics is a very contentious issue. In general, good business practice dictates that a half dozen meaningful metrics is all that should be used in the management of an organization. However, in R&D we have concluded that there simply are no meaningful metrics, in the sense that a manufacturing or service company would use them. We did not arrive at this conclusion capriciously. On the contrary, we undertook extensive face-to-face interviews with the senior leadership of many of our leading technology companies. In every case, they could proudly explain their metrics-related management processes in relation to their product development and production operations; when pressed to discuss their research function, however, they all admitted that they had no satisfactory approach, and that they relied heavily on the intuition of their Chief Technical Officers.

æ

Whenever we discuss our 59 metrics, we immediately receive criticism for their quantity and lack of relevance. Therefore, I emphasize again: these 59 "countables" are used only as indicators of the health of various functions within the laboratory. They serve only very limited purposes, but these purposes are often very important in the management of the R&D enterprise:

o certain of the metrics serve as indicators in narrow functional areas (such as fiscal performance) and never need to be surfaced to the laboratory's leadership unless they begin to move outside some predetermined bounds;

o certain metrics may be used by the Director as "levers" to move the organization in certain directions that he feels are necessary (such as the percentage of PhDs in a research organization); and finally

o certain stakeholders require data which, for reasons that may not be known to us, have significance to them.

This last purpose is, in a sense, the most difficult to understand, so we just accept it. Again and again, whenever we have reexamined the set of metrics and decided to drop one as being of no value, we will invariably receive a question from a stakeholder concerning that particular number. Therefore, our philosophy is: "If it's countable, we count it!" We just take care to assure that such data are not misused or misinterpreted.

7. Different stakeholders may have widely diverse views on what is important for ARL to accomplish. There are internal and external views, short-term views and long-term, emphasis on purely scientific output versus technical deliverables versus operational (fiscal, personnel, etc.) performance, and so on. The Performance Evaluation Construct was designed to provide the Director with the required flexibility to respond across these various interests.

8. The time lines for required submissions of documents to OMB have not been in synchronization with our own budget cycle. For instance, during our participation as a pilot project performance, plans have routinely been requested from six to eight months before the beginning of the fiscal year. At the agency level this may be appropriate. However, at the level of the individual organization, this is well before we know what the coming year's budget will be in even a crude sense, so the construction of a detailed performance plan is all but meaningless. In ARL's case, we actually cannot say with more than 75 to 80% certainty what our program will be until the fiscal year has begun. This is because major adjustments are made in our program well into the fiscal year at all levels, starting with the Congress and working down to our next higher headquarters. In FY95 we submitted a preliminary plan in September, which was already late according to OMB instructions. We then submitted a revised plan three months later. An analysis of the changes that occurred between the two submissions showed the following:

o Of the 47 technical objectives listed in the preliminary plan (which, although they were only the most significant goals of each of our directorates, were still only a fraction of our total program), 19 were dropped and 25 new ones were added. This reflected a 65 percent increase in the funds

æ

represented by just this subset of our technical program, and a complete rework of one entire directorate's program plan.

o Of the 19 metrics cited in the September version of the plan, 15 were changed by December, with the variances, both plus and minus, ranging from 1 to 100 percent.

o During October, one month into the fiscal year, $70M was placed on hold by DOD against possible Congressional actions. This amounted to fully 25 percent of ARL's program. By the end of the first quarter, $25M was still on hold.

o Other smaller special programs, such as Small Business Innovative Research (SBIR), are not known at all until the fiscal year begins. In FY95 this amounted to 6 percent of our program. Also, customer (reimbursable) funds are usually not identified, and never arrive until after the fiscal year begins. These funds represent, on average, a quarter of our program.

FY95 was not an anomaly. This has been standard operating procedure for ARL and its predecessor organizations for as far back as we can remember. Thus, OMB needs to realize that planning at the agency level is quite different from that at the organization level, and depending on what is required, appropriate accommodations must be made. On the other hand, if the agency level is all that will be dealt with by GPRA, this problem does not arise.

9. With only a year and a half left of the pilot program before the Act is implemented government wide, we are hearing hardly anything about a transition plan. As sole representative for the R&D function, ARL has, as described herein, put a great deal of effort into understanding the application of planning and measuring as applied to R&D. Yet there has been no discussion of how the knowledge gained from the pilots will be "harvested." We at ARL have been very active in mentoring other government R&D organizations through hosting symposia and making presentations, both formal and informal. We have been a leader in the deliberations of the interagency R&D Roundtable. However, this has all been ad hoc activity. The request for this case study has been the first we have heard from OMB, and even this was only voluntary. Is there to be no formalized transition process involving all the pilot projects?

Finally, this is still work in progress for ARL. We made a major advance in the latest version of our strategic plan by incorporating for the first time strategic *technical* goals, in addition to the more familiar broad management goals usually seen in such plans. In fact, this drastically revised plan was motivated in part by our receiving a new mission assignment from the Chief of Staff to develop the technology for the future Digitized Battlefield. That led to a whole new structure for our program and with it, a new strategic approach to our modus operandi. However, the all-important linkage from strategy to annual performance still needs improvement. While each of the volumes of the business plan is complete, consistent, and well-structured within itself, the natural flow or driving function from one to the other is not obvious. Also, strategic human resource planning, facilities master planning, etc., are not integrated into the overall strategic plan. In addition, the actual

æ

application of the plan is not optimum. Deployment down to the workforce and first-level supervisors is particularly spotty. Therefore, we will be undertaking a senior-level review of the planning process to see what improvements can be implemented.

Our set of metrics also needs continuing reevaluation in terms of their value to the process, their collectability, their amenability to validation, and most importantly, the message they send to the workforce. One must keep in mind the observation that "what gets measured, gets done." We need to assure ourselves that we are measuring the right things so that we will do the right things.

Part of the reason for this apparent lack of a total, smooth, and well-integrated deployment of our planning and measuring techniques may simply be the continuous fire-fighting mode that we in the government are so prone to operate in. To quote from "Winnie the Pooh" by A.A. Milne:

> "Here is Edward Bear, coming downstairs now, bump, bump, bump, on the back of his head, behind Christopher Robin. It is, as far as he knows, the only way of coming downstairs, but sometimes he feels that there really is another way, if only he could stop bumping for a moment and think of it."

V. CONCLUSION

Because of the scale question discussed above, the GPRA *per se* may not ultimately have a major impact on ARL's business operation. However, the planning and measuring that constitutes the GPRA process are still important to us and will continue, regardless of the specifics of the eventual implementation of the Act. We intend to continue to use these tools and improve upon them, simply because it is "good business" to do so.

Although it is difficult to do these things, in today's economic climate, they must nonetheless be done. A "report card", on the status of planning and measuring at ARL might yield something between a C and a B+ (depending on who is doing the grading and how generous they are inclined to be.) At this point, we cannot claim to have found all the answers. However, we do believe that we have figured out what the questions are.

æ

APPENDIX A.

The Mission and Vision Statements of the Army Research Laboratory

MISSION:

Execute fundamental and applied research to provide the Army the key technologies and the analytical support necessary to assure supremacy in future land warfare.

VISION:

- A laboratory preeminent in key areas of science and engineering relevant to land warfare.
- A staff widely recognized as outstanding.
- Seen by Army users as essential to their missions.
- An "intellectual crossroads" for the technical community.

æ

APPENDIX B.

ARL Metrics

For FY96, ARL's set of metrics comprises 59 different quantities, all of which are tracked by some part of the organization, but only 15 of which are tracked by the Director's staff and reported to the Director for use as a management tool. We originally assembled a list of metrics simply by gathering all those things that were "countable" and that appeared at the time to be of some relevance to the health of the organization. Most of them are input measures, and some are output measures. None is an outcome measure. Nevertheless, inputs and outputs are useful tools for the management of the laboratory, so long as their true significance is kept in perspective.

After spending two years working with these metrics and gaining experience with the system for collection, validation, and use, we have conducted a thorough review of these metrics. We subjected each one to scrutiny to determine its real meaning for the organization, assure that it is properly defined, ascertain what "message" it sends to the workforce, determine how best to collect and validate it, and determine of what use the information will be once collected. Most importantly, we wished to determine precisely who is the "owner" of each particular metric (i.e., who is responsible for performing the tasks represented by the metric and/or responsible for collecting the data,) and who is its "customer." As noted above, at this time the Director is the customer for only 15 out of the 59. The other 44 are of use to various functional managers as indicators of their operation's performance. They are surfaced to the senior management only if they start to move well beyond some pre-determined bound and are in need of executive attention.

Following is a listing of all 59 of the current metrics, along with a brief description or definition where

the meaning might not be readily obvious to a reader outside DoD. (An asterisk indicates one of those 15 that are reported to the Director and included in the performance appraisals of his senio r managers.) The metrics are grouped as they fall within the four elements of the ARL visio n statement thus forming a vital coupling in the overall linkage from the ARL mission to the vision to the strategic goals to the annual performance objectives.

Vision Element 1. *A laboratory preeminent in key areas of science, engineering, and analysis relevant to land warfare.*

Deliverables (outputs) -- these represent a quantity of work accomplished with no attempt to assess the quality of that work

1.* Percentage of "Top 10" tasks completed -- the ten most important annual tasks in a directorate as determined by the directorate head

2.* Percentage of STOs completed -- the Science and Technology Objectives, the attainment of which is determined to be the responsibility of ARL by Department of the Army Headquarters

Documentation: "Leaving Tracks" (outputs)

3.* Number of refereed journal articles/proceedings

4.* Number of ARL Technical Reports

5.* Number of Test Reports (for Survivability/Lethality Directorate only)

6.* Number of completed software packages (for Advanced Simulation/High Performance Computing and Information Sciences and Technology Directorates only)

7. Number of books/chapters in books published

8. Number of invention disclosures

9. Number of patents awarded

Infrastructure (inputs)

10. Facilities investment (in $M) during the fiscal year

11. Facilities replacement rate

12. Investment (in $M) in new capital equipment during the fiscal year

Vision Element 2. *A staff widely recognized as outstanding.*

Personnel Profile (inputs)

æ

13.* Percentage of PhDs among the Scientist and Engineer (S&E) portion of the workforce.

14. Ratio of technicians to S&Es

Training (inputs)

15. Percentage of employees with 40+ hours of training per year

16. Number of employees on long-term training -- defined as training (degree or non-degree) away from the individual's assignment during duty hours for a period exceeding 179 days

17. Number of PhD Candidates

Esteem Factors (outputs) (these provide an indication of the degree to which others see ARL as a "world class" organization)

18. Number of significant awards (from national or international bodies)

19. Number of invited presentations (at national or international conferences)

20. Number of prestigious posts (e.g., fellowships in national academies, professorships at leading universities)

Vision Element 3. *A laboratory seen by Army users as essential to their missions.*

Technology Transitions (outputs)

21. Number of significant technology transitions

Ratings from customer surveys (outputs)

22.* Customer Satisfaction Rating as ascertained via survey

23. Stakeholder (Users and senior leadership) Satisfaction Rating as ascertained via survey

Financial (inputs)

24.* Funding from customer sources (in $M) -- reimbursable funding

"Greening" the Workforce (inputs) -- An indication of the efforts made to acquaint entry and mid-level civilian staff members with the "real" Army, the customer for whom we work. Various Greening courses and assignments are conducted in which civilian employees, in addition to attending lectures, observe demonstrations of military hardware, fire small arms, drive military vehicles, and, in general, learn a little about what the life of a soldier is like.

25. Number of officer personnel

26. Number of enlisted personnel

æ

27. Percentage of employees completing the basic ARL "Greening" course
28. Number of FAST advisors -- individuals on two-year assignments as science advisors to the Commanders-in-Chief (CinCs) of the major defense commands as part of the Field Assistance in Science and Technology (FAST) program
29. Number of employees completing the FAST, Jr. training course -- training to participate in a short-term (several week) version of the FAST program

Vision Element 4. *An intellectual crossroads for the technical community.* (the following metrics give some indication of the degree of outreach of ARL to the rest of the technical community)

Guest Researchers out of ARL (inputs)

30.* Total number

31.* Total manyear equivalents

32. Average length of stay

33. Number staying more than three months

Guest Researchers into ARL (inputs)

34.* Total number

35.* Number of post-docs

36. Number from HBCU/MI (Historically Black Colleges and Universities/Minority Institutions)

37.* Total manyear equivalents

38. Average length of stay

39. Number staying more than three months

Advisors (inputs)

40. Number of National Research Council approved post-doc advisors

Cooperative R&D (outputs)

41. Number of Cooperative R&D Agreements (CRDAs) established during the fiscal year
42. Number of Patent Licensing Agreements (PLAs) established during the fiscal year
43. Income from CRDAs/PLAs

44. Number of Technical Project Officers/Associate Technical Project Officers (International agreements only)

Other

Financial (inputs)

45. Obligation rate
46. Disbursement rate
47.* Indirect overhead rate
48. Ratio of in-house to out-of-house work -- an indicator of balance between in-house technical strength and reliance on the private sector

Personnel (inputs)

49. Glidepath (S&Es; total) -- the annual manpower target during the period of personnel drawdown
50. Average age (S&Es; total)
51. Average grade (S&Es; total)
52. Average sick leave usage (S&Es; total)
53. Turnover rate (S&Es; total) -- the rate at which employees voluntarily leave the organization

EEO (inputs)

54. Female and minority distribution by grade
55. Female and minority promotion, training and awards statistics
56. Number of formal EEO complaints

Procurement (inputs)

57. Average small purchase cycle time
58. Percentage of HEI (Higher Educational Institution) contract funds going to HBCU/MIs
59. PALT (Procurement Acquisition Lead Time)

The following table shows the FY93 through 95 achieved goals of the Director's "special interest" metrics for each directorate and the total for ARL, along with the long term (FY99) ARL goals. This data is gathered annually by the Director's Staff from points of contact in each of the directorates.

APPENDIX D.

THE XM773 MULTIOPTION FUZE FOR ARTILLERY (MOFA)

An Application of the ARL Performance Evaluation Construct
Applied throughout the R&D Process

The XM773 MOFA is an electronic artillery fuze that can provide detonation signals to the round in a variety of options: contact burst, delayed burst, or proximity burst, with the burst height capable of being varied over a wide range of values. The fuze is currently one of two that have been selected to be fielded with the Army's Crusader self-propelled howitzer system. The Crusader is one of the Army's most important system development programs. It is almost ready to go into production and, when fielded, will be the Army's primary artillery weapon into the 21st century. The MOFA gives the Crusader the capability to automatically select the desired round function by an electronic signal communicated to the fuze while it is being automatically loaded into the breech of the 155-mm main gun.

The MOFA represents the absolute state of the art in fuzing and is without peer anywhere in the world. Its development began at the Army's Harry Diamond Laboratories (HDL) and continued when that organization became part of what is now the Army Research Laboratory (ARL). What follows is a short history of the program from its beginning as a fundamental research effort, through an applied research phase, and then into various stages of development to the present, when it is ready to leave R&D and enter production. Along the way, various methods were used to evaluate the program's performance; these are used as examples to demonstrate how the recently developed Performance Evaluation Construct could have been applied then, and can be applied to future programs at all stages in their development.

****

MOFA's roots go back to the 1968-1970 period, when basic research was done at HDL on wide-band linear frequency modulation. From this work, among other things, came the doctoral dissertation of the group's leader, which was entitled "Resolution in Frequency Modulated Radar." An arrangement with the Electronic Communication Laboratory of the University of Florida was entered into to continue the research on FM ranging theory. Over 100 reports and articles resulted from this collaboration. The work at this stage was supported under the DoD budget category 6.1 -- Basic Research. It received peer review by the HDL Scientific Advisory Committee, as well as the reviewers of the group leader's dissertation and the editorial reviewers of the several papers that emerged from the group during this period. Some of the metrics that were used by the laboratory's director to evaluate the program included the traditional indicators, such as the publication count, as well as the patents that resulted from the work, although there were not many at this early stage. There were no external customers at this stage; the only "customer" that could be claimed was the lab director. The Construct matrix at this point in the program looks like this (a "O" indicates no relationship or significance; a "+" indicates a relationship or significance of some importance; a "++" indicates a very strong relationship):

	Relevance	**Productivity**	**Quality**
Peer Review	O	+	++
Metrics	O	++	+
Customer Feedback	O	O	O

By 1974 the work had progressed under the DoD system to 6.2, which roughly translates as applied research. The earlier research had led to the concept of the first directional Doppler ranging system that possessed high electronic countermeasure resistance. Further work showed that the functioning distance (the burst height) of a fuze that would be built based on this technology could be remotely set by electronic induction techniques. Furthermore, applied research on antennas yielded patch antennas that were both wide band and small enough to fit on the fuze radome. During this period, the HDL Scientific Advisory Committee was still reviewing the program for its scientific excellence. There were few refereed journal articles, but many internal reports and increasing patent activity.

Early discussions were held with organizations such as Picatinny Arsenal that would have the developmental responsibility for any fuzing system that might eventually emerge from this early work. This begins to bring customer feedback into play. Also, metrics in the form of test results start to become important. At this point, the Construct matrix looks like this:

	Relevance	**Productivity**	**Quality**
Peer Review	O	+	++
Metrics	+	++	+
Customer Feedback	+	O	O

By the mid-1980s, the directional Doppler ranging system was refined and developed for the Medium Altitude Proximity/Time (MAP/T) fuze. The signal processor circuits were integrated into a custom integrated circuit. Concept studies were initiated for a MOFA-type device. Ballistic tests were performed on the MAP/T fuzes to assure that the user's requirements were all satisfied. Through the latter part of the 1980s into 1990, the work being done on MMIC (millimeter wave/microwave integrated circuit) technology by the Advanced Research Projects Agency (ARPA) was imported to the MOFA program to redesign the transceiver element. A technical demonstration program was initiated in 1990, and HDL built a limited number of fuzes to investigate technical performance and gather statistics. The program was now in 6.3 -- advanced development. As it moved into 6.4 -- engineering development -- in 1992, cost goals were set for eventual production, and development programs concentrated on re-engineering the earlier prototypes for production. By this point the customer, PM Crusader, was very heavily involved in reviews of the program, as the MOFA had been selected as one of two fuzes that will be procured as part of the overall howitzer system. Metrics relating to performance are now numerous. Ordinarily, at this stage, peer review would not be a major factor in a program. However, in this case, a blue ribbon panel composed of industrial and academic leaders was convened to review the program in February of 1995. They declared the MOFA to be a world-class design. The fuze is now on the verge of leaving the R&D process and entering the procurement system to be type classified, produced at large-scale rates, and fielded with the Crusader. The construct matrix, therefore, looks like this:

	Relevance	**Productivity**	**Quality**

Peer Review	+	O	+
Metrics	++	++	+
Customer Feedback	++	++	+

****

This example demonstrates how the Performance Evaluation Construct can be used throughout the life of a typical R&D program, from its inception as undirected fundamental research, through its evolution, to the final stages of development in preparation for production. The three matrices shown above are typical of what might have been used had this construct existed in the 1970s. They represent the actual tools used by first the HDL management, and later the ARL management to evaluate the program, and they indicate how R&D programs today might be looked at. Different programs in the future might have variations from these, but would probably be similar.