

**THE METRICS OF TECHNOLOGY EVALUATION:
WHERE WE STAND AND WHERE WE SHOULD GO FROM HERE**

by

ELIEZER GEISLER

Professor of Organizational Behavior
Stuart School of Business
Illinois Institute of Technology
565 West Adams Street
Chicago, Illinois 60661
Tel: (312) 906-6532 Fax: (312) 906-6549
Email: geisler@stuart.iit.edu

July, 1999

Presented at the 24th Annual Technology Transfer Society Meeting, July 15-17, 1999.

STUART WORKING PAPER 99-03

**THE METRICS OF TECHNOLOGY EVALUATION:
WHERE WE STAND AND WHERE WE SHOULD GO FROM HERE**

ABSTRACT

The transfer of technology is but one aspect of the evaluation of science and technology. The complexities of technology generation, transfer, and commercialization processes have always gravely taxed the way we measure and assess them. Recently government technology agencies have experienced the exigencies of the GPRA, only to encounter severe difficulties in the application of metrics that would satisfy the requirements of the GPRA and their own satisfaction that a solid evaluation is being conducted. This paper reviews the existing metrics for evaluation of science and technology. Among the categories of metrics, this paper reviews econometric methods, process methods, and bibliometric methods. The paper also reviews models of the innovation continuum, in an effort to link such models to the metrics categories. Based on the review of the state of the art, the author proposes future directions for the development and the application of metrics of technology evaluation. Some examples and "best practices" of industrial and government laboratories are also discussed.

INTRODUCTION

In 1872 the American essayist Ralph Waldo Emerson (1803-1882) had remarked in one of his lectures: "If a man can write a better book, preach a better sermon, or make a better mousetrap than his neighbor, though he build his house in the woods, the world will make a beaten path to his door." Although it had become a much quoted and popular statement in praise of inventors and inventions, Emerson's statement reflects only half of the conditions for use of a new technology. The mere possession of the technology is not a guarantee that it will be utilized and that the "world" will perform "make a beaten path" to the inventor's door. Technology is not judged by its existence alone, nor is its mere existence a sufficient condition for its successful usage.

We cannot evaluate technology unless and until we put it in the context of social and economic phenomena. Namely, until and unless we put a *human* dimension upon it. A better mousetrap, by itself, is useless, until and unless it is used and assessed by its users. So, technology is only evaluated as a tool, subjected to human needs, desires, preferences, and emotional as well as material parameters of existence.

Therefore, love or hate of technology per se is an empty concept. Since science is universal, hence shared and almost impossible to isolate or conceal, the technology that it generates will come into fruition regardless of any barriers imposed on its formation. For example, technological advances in genetics and bioengineering will emerge regardless of the debate concerning their ethical viability.

This paper is about the evaluation of science and technology and the metrics used in such evaluation. Science is lumped together with technology since there is a seamless flow along the innovation process, and most metrics discussed here are used to evaluate both science *and* technology.

The paper starts with a discussion of technology in its context, and the criteria for its evaluation. It then proceeds to describe the key metrics used in this evaluation. Strengths and weaknesses of each category of metrics are listed, and some best uses are illustrated. The paper ends with some conclusions and directions for the development and application of metrics for technology evaluation.¹

WHAT ARE WE EVALUATING?

We evaluate the process of moving downstream along the innovation continuum, from the research component to the final destination of a *useful* outcome and its impacts and benefits. The mere possession or existence of technology is not a guarantee for its actual use. We cannot understand technology by its existence alone, but only within the context of the uses: people, organizations, and society.

In its actual manifestation, technology ranges from just a piece of knowledge to a method or technique, all the way to a complex system of machinery and its inherent intelligence. The key is that technology is not defined and evaluated by what it is, but by criteria

outside itself—by its actual and potential users. This attribute creates a wide range of possible definitions and multiple interpretations of what technology consists of, and what are its impacts, outcomes, benefits, or harmful effects.

Another consequence of this attribute is the fact that evaluation of technology thus depends on the characteristics of the users—their ability to make assessments, to utilize the technology, and to minimize their preferences, predilections and subjectivity. Clearly, potential users generally cannot entirely distance themselves from subjectivity and idiosyncratic preferences—hence the multiple approaches to the evaluation of technology.

Consequently, technology is evaluated by users who are themselves limited in their capacity to fully integrate and utilize technology, and who are inherently biased due to their culture, unique needs, and a host of other factors, including even the mere understanding of the technology itself.

PARADOX OF TECHNOLOGY EVALUATION

A commonly discussed paradox of technology refers to the reality of technology diffusion and maturation. Once adopted, every technology is bound to be replaced. But the difficulties and the complexity of technological development and adoption/integration make replacement an activity mired in frustration, failure, and chance.

But the other less obvious paradox is that of technology evaluation. It concerns the gap between producer and users of technology. The more technology producers attempt to bridge the gap and to apply users' criteria and measures to their own reasoning and processes, the less likely they are to produce technology that will replace the existing technology.

To clarify, the more researchers, for example, attempt to subordinate their thinking to the commercial reasoning of the ultimate users of the technology, the less innovative the technology may thus become. Since the criteria for evaluation are generated by the users, producers of technology find themselves in a position where they must act in accordance with these criteria in order to survive. By doing so they may lose their distinctive edge of inventiveness, creativity, and novelty. Hence, in order to be successful (by the criteria of the users) producers of technology have to conform to the "rules of the game" of users. But conformity is an antidote to inventiveness, exploration, and risk taking—all of which are necessary conditions for humans to innovate.

PUSH-PULL APPROACHES

The basic relationship in the evaluation of technology is its link to commercialization, or otherwise, its ultimate use by social and economic actors. The link from generation to utilization may be viewed in the "pull" or "push" approaches.

The "pull" approach is also sometimes known as "demand-pull," in which the impetus for the technology comes from the user. The "push" approach describes the case where the generators of the technology put their effort in bringing the technology to the users.

These approaches and the distinction between them are important because of the criteria for evaluation that each approach seems to encompass. The "push" approach would be guided by the values and attitudes of the generators of the technology, whereas in the case of "pull," the criteria would be those of the users.

Who decides on the flow of technology? This question is at the heart of the push-pull framework for analysis of technological change, transfer, and utilization. In the process of innovation it has long been known that when the research activity simply "throws" its outputs "over the wall" to development, engineering, and other product-related functions in the company, the chances for successful transfer are diminished. When there is, however, a reciprocal wish to accept these outputs and the awareness on the part of the recipients that there is a need and potential uses for them—the chances for successful transfer are thus enhanced.

"Push" or "pull" alone cannot necessarily make the technology flow successfully towards utilization. As stated before, the mere existence of the technology will not always lead to successful usage. The graveyards of corporate technology laboratories are burgeoning with technological achievements that never made it to usable commercial applications. So, evaluation of technology in terms of its ultimate usefulness to the economy and to society depends, in part, on the source of the flow, or the motivation to make it happen. Both sides to this equation must be committed, if successful utilization is to be achieved. Total push may bear fruits in the short run, but without reciprocity from users it will ultimately fail. For example, Henry Ford's insistence on pushing his cars in one color, black, finally succumbed to the users' demand for a variety of colors. The need for reciprocity is magnified in the case of technology, as sometimes superior technologies in certain products are not preferred by customers.

EVALUATION AND METRICS: THE UNHOLY ALLIANCE

Technology has to be evaluated in conjunction with the scientific activity that generated it, on the one hand, and the impacts it produces, on the other hand. Technology as a stand-alone entity loses much of its power as a tool for evaluation. There is always the need for the context of where technology had originated, and where it is going.

In addition to this imperative, the seamless flow of the innovation process—from research to development to engineering and commercialization—creates even more powerful barriers to a strictly, isolated identification and evaluation of self-contained stages and activities in this process. In what I would call "balkanization" of the process of innovation, the available methods for evaluation are few, thus providing only a partial picture of the true phenomenon.

There is a conflicting discrepancy between the effort to evaluate science and technology, and the metrics available for the task. The objectives and the expectations from the

evaluation effort have by far outreached the pool of measures. So any attempt to match existing metrics to the evaluation scheme would almost always encounter gaps, challenges, and unanswered questions.

For example, technology cannot be evaluated without regard to its generation in science, and its impacts on the market. Yet, as measures tend to follow a "Balkanization effect," in which compartments are formed in the process, the effort to employ metrics is insufficient.

Much of what science and technology provide are an integral part of their evaluation, yet remain intangibles and very difficult to measure, other than by surrogates and similar techniques. The exhilaration of flying that pilots experience is an example of what we gain from technology. Similarly, the "eureka" feeling of extreme accomplishment in discovery and invention remains beyond the reach of common metrics. If science is our exploration for understanding, technology is the extension of what we can do.

Such achievements are sometimes as subjective as they are objective, and long term as they are immediate. Science and technology are embedded in our senses, in humanity's affinity for work as well as entertainment, for mastering the challenging environment, as well as for amusement and pleasure. For instance, many breakthrough innovations in computers and software in the 1990s were an almost direct outcome from games for children invented in the 1980s. The game called "Pacman" was the direct progenitor of the sophisticated computer graphics of the 1990s. So many of the inventors who were the driving force behind computer graphics and the internet throughout the 1990s had been young, sometimes *very* young. Playing with this technology was *fun*, pure and simple.

Science and technology are not a luxury, to be applied only by people, organizations, and countries who have enough stock resources to exploit this expensive "pas- time." Although science is no longer a preferred leisurely activity of inquisitive gentlemen and has become "big science" with its own dynamics—the exhilaration of creation and exploration is nevertheless very much alive in the scientific community.

Thus, the count of a publication or a patent, or a system that emerges from the flow of innovation, is barely enough of a metric to express all that S&T can do. There is always the human element of risk versus pleasure, failure versus fun, and the balancing of a calculated feeling of discovery, with the convulsion of energy and pleasure people feel when they ride the technological achievements of their doing.

WHAT ARE METRICS AND HOW TO CONSTRUCT THEM

Werner and Souder have conducted a comprehensive review of the literature from 1956-1998, and have concluded that "the choice of an appropriate R&D measurement metric depends on the user's needs for comprehensiveness of measurement, the type of R&D being measured, the available data, and the amount of effort the user can afford to allocate to it"³ (p. 34). Although these authors have recognized the difficulties in constructing a systematic classification of such metrics, they nevertheless categorized

them into quantitative-subjective, qualitative metrics, and integrated metrics. This classification may also be adequate for the metrics of science and technology (S&T).

But, what exactly is a metric? A metric is a system of measurement that includes the *item* being measured, the *unit* of measurement, and the *value* of the unit. If we combine this definition with Werner and Souder's classification scheme, metrics can be in the form of a single measure, a ratio, an index, or an integrated measure that combines several metrics, even with different attributes, such as objective and subjective. With such a broad range of possible modes, is there a "basic" metric that allows for combinations and manipulation to create more complex metrics? Such a "building-block" metric would be amenable to transformation by combining measures of diverse phenomena, such as peer review (subjective) and count of patent (objective).

The literature provides very few clues to the existence or to the definition of this basic metric. There seem to be overlapping usages and even divergent definitions that conjure to serve the unique needs and objectives of those who select and compose the metric. Standardization and coherence in rules of construction of intricate metrics are yet to be achieved.⁴

So, in the broad area of evaluation of R&D and S&T, the practice of creating and using metrics is in the form of a "menu." Evaluators avail themselves to a selection of metrics, measures, and instruments from which they pick and choose the combination that will address their objectives and their needs.

The six steps in [Figure 1](#) portray a generic view of how metrics are constructed. There seem to be two critical components to this process. The first is the conceptual guidance to the creation of the metrics, shown in steps 1 and 2. Without this framework, the metrics are simply data. The second is the available pool of measures and corresponding data. Kostoff had concisely discussed the problem of data collected into metrics without a guiding framework.⁵ He argued that "Every S&T metric, and associated data ...should have a decision focus; it should contribute to the answer of a question which in turn would be the basis of a recommendation for future action" (p. 12). Hence, when the data become more available and the means to assemble and to manipulate them become more sophisticated, there is a tendency among evaluators to concentrate on the data (steps 3-5) rather than on the constructs.

Are there fundamental measures of S&T that can be used as building blocks for S&T metrics? Apparently the answer is no, since basic counts of scientific articles or patents are in themselves insufficiently descriptive of the notions that S&T metrics should provide. In the absence of such basic elements from which more intricate metrics may be constructed, the reality for measuring S&T is an opportunistic selection of metrics from an existing menu of measures.

But, even as this process encapsulates selection of measures and metrics to suit the needs and objectives of the evaluator, there are several "preferred" measures and their more complex metrics. Such preferred measures include measures of processes and outcomes.⁶ This creates a reduced universe of available measures, so that their repeated use in research and in practice allows for comparisons over time and among organizations. There is seldom a need to "reinvent the wheel." Perhaps the "best" metrics are not always constructed, but they are consistently utilized, thus creating a "track record" that permits the identification of strengths and weaknesses of these metrics.⁷

In light of the effort to construct sets of metrics, as illustrated in Figure 1, the notion of a metric needs revisiting. First, since a metric may be composed of a single quantity, as well as a more intricate and complex set of measures (such as indexes and "macro" metric), a definition of a metric extends beyond simply a "measure."

Second, metrics of S&T are extracted from across the entire innovation continuum. They are designed to measure a variety of activities, events, and phenomena—some simple and short-lived, others highly complex and durable along an extended time frame. Third, the absence of a unique and single building-block increases the role of subjective reasons for the construction and selection of metrics.

Hence, a metric of S&T is more than a simple measure or a quantity, or an indication of some value. A refinement of the definition of a metric extends it to include: (a) the item measured (*what* we are measuring); (b) units of measurement (*how* we measure); and (c) the inherent value associated with the metric (*why* we measure, or what we intend to achieve by this measurement). So, for instance, the metric: *peer-review* includes the item measured (scientific outcomes), the unit of measurement (subjective assessment), and inherent value (performance and productivity of scientists, engineers, and S&T units).

A MODEL OF METRIC SELECTION

Due to the fact that loosely assembled sets of metrics and the pick-and-choose option are the norm in the literature and in the practice of S&T evaluation, the modeling of the selection of metrics is crucial to our understanding of how S&T should be assessed. A generalized model of the components of metric selection is shown in Figure 2.

A MODEL OF METRIC SELECTION

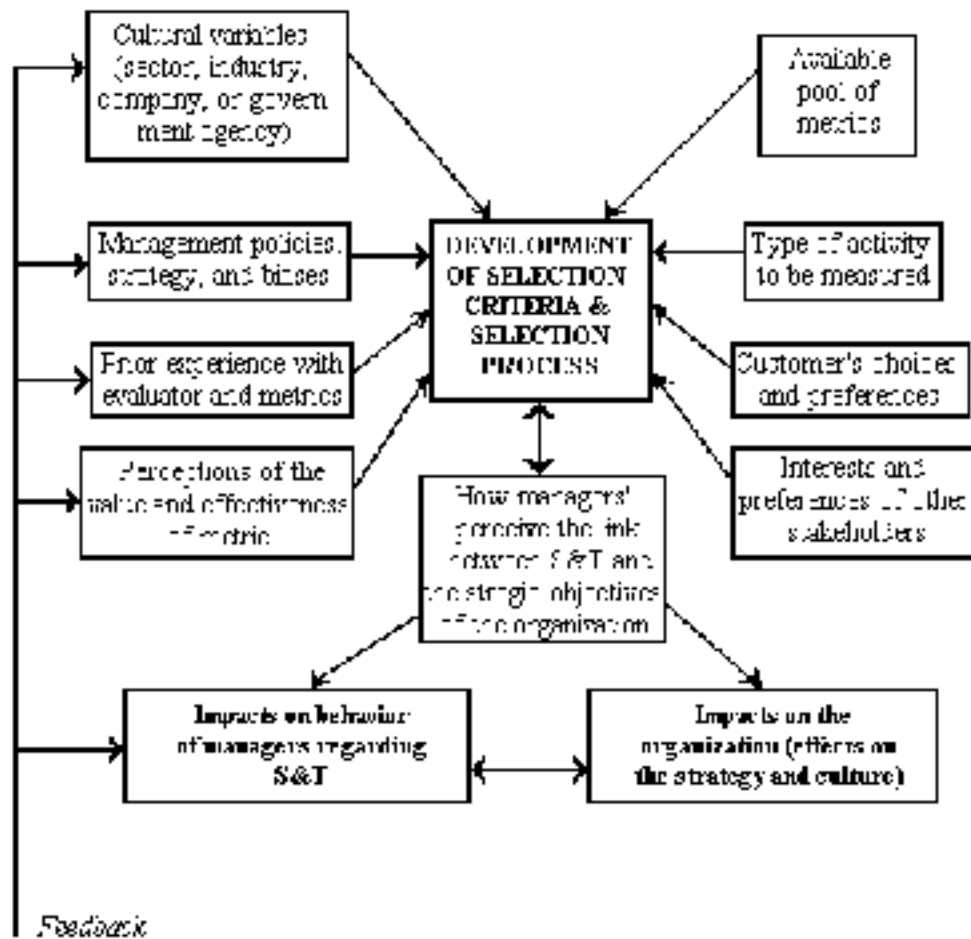
The selection process is influenced by three sets of variables: (1) the culture in the organization (including managerial preferences and perception); (2) the available pool of metrics and the type of activity being measured; and (3) influences of others, such as customers and other stakeholders (scientific community and business or government interests).

The outcome of the selection process may be categorized into two major impacts. First, impacts on the firm or the government agency or laboratory. In the industrial company the choice of metric may ultimately affect the strategy and the culture of the organization.

At first, the culture of the company is a factor in determining the choice of metrics. But, after the metrics is applied and experience with it is gained, the strategy and the culture may change.

Figure 2

A Generalized Model of Metric Selection



In this respect Hauser and Katz have cited the case of Not-Invented-Here (NIH) as a cultural impediment to the selection of appropriate metrics.⁸ They have also considered the example of the use of the R&D-Effectiveness Index (E.I.) as a single or principal metrics of R&D evaluation. This choice may ultimately lead to costly distortions in the resource allocation process for S&T.

A second possible impact is on the behavior of managers regarding S&T. Metrics that encourage certain behaviors that are risk-averse or that promote risk and offer selected rewards for selected modes of behavior may create a chain reaction in the organization. Managers may opt to abandon longer-term projects, hence suffer from the classical

effects of "contamination" of the metrics, so that their behavior is now geared toward satisfying the metric rather than toward an effective discharge of their tasks in S&T.⁹

These possible impacts are mitigated by the way managers perceive the role of S&T in influencing the strategic position of their organization. The more they perceive S&T to be critical to strategic success, the more the metrics they select will tend to influence their behavior and their strategic decisions.¹⁰

POOL OF METRICS

Based on the review of the relevant literature there emerges a comprehensive set or pool of metrics for science and technology. These metrics are generic, as they will be applicable to different stages and dimensions of the innovation process. The key metrics are shown below, classified by eight major categories.

1. INPUT/INVESTMENTS IN S&T*

- Expenditures for each stage of R&D/I.
- Expenditures per time-frame, for one time period or over several time periods.
- Distribution by categories of expenditures: personnel, equipment, etc.
- Source of funding (business unit or central lab in the industrial company).
- Comparison of expenditures, per item category: by competitors, industry averages, and sector averages.
- Expenditures by discipline, technology, and scientist & engineer(S&E).
- Expenditures related to a product-line or other commercial unit of reference (such as customer or market).

B. ECONOMIC/FINANCIAL METRICS

- Cost savings (including ratio of savings in cost of goods sold)
- Return on investment.
- Return on assets.
- Payback of investment.
- Economic measures (such as price differential per unit quality characteristics that can be attributed to science and technology, computed as differential minus cost per unit to which quality characteristics are attributable multiplied by sales volume of this given unit.
- "Dollarization": Profit/Cost of R&D employees.

*Measures are listed not necessarily in order of importance nor any other ranking or ordering criteria.

1. COMMERCIAL AND BUSINESS METRICS

- New sales ratio (sales revenue from products and product or process enhancements that can be attributed to commercialization of S&T outcomes in a given year).
- Projected sales and income (from S&T projects in progress or in the innovation pipeline, by project or by categories of products and processes).
- Profit ratio: profits as percentage of sales that can be attributed to S&T outcomes that had been commercialized and integrated into new and existing products and processes that had generated the profits.
- Market share ratio (relative market share per product category or unit that can be appropriated to the contribution of S&T outcomes to product sales that have generated the market share).
- Customer satisfaction (overall ratings of the company and its products, and specific ratings of the competitive features of products or services that can be attributed to S&T outcomes incorporated in them. This measure may also be in the form of ratios of expenditures for S&T, and specific S&T outcomes, such as patents).
- Interactions with customers (*internal* contacts of scientists and engineers with their corporate customers, such as marketing and production, and *external* contacts by these S&Es with the corporate customers, such as S&Es of the client organization, as well as managers—technical and commercial—of the client organization)
- Regulatory compliance (contribution of the outcomes of S&T to the compliance of products and services with regulations, that can be attributed to the outcomes from S&T that are incorporated into products and services—including issues of safety, health, and ethics).
- Quality and Reliability (contributions of the outcomes of S&T to the level of quality and reliability of the products and services sold by the organization that are considered acceptable by customers and regulators).
- Time-to-market (response time) differential to match or to surpass competitors' new products and processes. This measures the improvement (reduction) in the time needed by the organization to introduce new products or services, as competitive weapons, that can be attributed to S&T outcomes. This measure can also be in the form of ratios of the reduction in time-to-market, for the cost and investments in S&T, or per specific S&T outcomes.
- Proprietary sales and revenues ratio (the portion of sales of products and services, as well as revenues from licenses and similar income categories that are protected by patents and other instruments of trade secrets, that can be attributed to those patents that offer specific protection of product characteristics that provide exclusive features for the organization over its competitors).

4. BIBLIOMERIC METRICS

- Publications (includes scientific papers, technical reports, and articles in scientific journals, book chapters, and proceedings of conferences and symposia. These measures can only be in the form of ratios to the investments in S&T that have generated these measures, or selected expenditures by category of type of industry and academic discipline—all in a given time period).
- Citation Analysis (includes counts of citations of scientific and technical articles, as one measure of impacts on scientific community and quality of the scientific effort that has generated the publications cited. This increase may also be in the form of ratios to investments in S&T, or by academic discipline and type of industry or sector in which the scientific effort has been conducted).
- Co-word Analysis and Database Tomography (DT). (These are measures of analyses performed on large databases of S&T bibliographical outcomes, in a form of data mining. DT, for example, provides a roadmap to topical areas found in relevant literatures. Key word analysis also assist in obtaining a measure of topical coherence, as well as topical lineage, which offers a measure of thematic history, thus helping in identifying the path of development of a field of S&T and a specific thematic area. Such measure is also helpful in a more accurate attribution of the origins and path of evolution of scientific breakthroughs, for both individuals and organizations).
- Special presentations and honors (including keynote presentations of scientific conferences and symposia and other *ad hoc* contributions to the literature).

5. PATENTS

- Count of patents (produced by S&T unit and per S&E in the unit and in the organization. This measure may also be in the form of ratio of number of patents in a given time frame per expenditures for S&T, namely, a measure of the cost per patent by type of industry so as to account for different patenting practices).
- Relevant or current patents (the percent of patents that are current and provide the organization with competitive proprietary advantages. Also in the form of ratio of these current patents per the total number of patents produced over a time-period classified by industry, to account for different patenting practices)
- Comparative patent standards (ratios of the unit and organization pool of relevant and useful patent, per the benchmarks in the industry, for the key technological areas in which the organization is competing.)

- Cost of patents (considering the time lag from the application of investments in S&T, to the point of registering a patent, this is a measure of the cost of a patent. Also to be considered are the issues invoked by identifying the link between investments and the patents that emerge, and differences in quality and competitive impacts of the various patents).

6. PEER REVIEW METRICS

- Internal evaluation (subjective rating of the S&T unit, its activities and its outcomes, by other people and units in the organization, such as marketing and production. This type of measure may be in the form of written evaluations and some ranking scale on an instrument that measures judgment of respondents, *ad hoc* or in a periodical manner.)
- External evaluation (subjective evaluation of the S&T unit, its activities, its outcomes, and its *overall* quality—by a panel of experts. This measure may be in the form of an invited effort requested by the S&T unit or its organization or external S&T experts, consultants, and other knowledgeable people in the community, or in the form of routine evaluation, as part of an on-going assessment of S&T.)
- Targeted reviews (including specific panel evaluations of any outcome from S&T, such as a specific scientific paper, project, or program. Also includes specific judgmental assessment of a product, a patent, and individual scientist and engineer. This measure may be considered a measure of quality, as viewed by expert reviewers).

G. ORGANIZATIONAL, STRATEGIC, AND MANAGERIAL METRICS

- Project management: internal/cycle time (This is a measure of the period between starting a given S&T project to the point of transforming an outcome to the downstream unit within the organization.) This may serve as a measure of internal efficiency in the company.
- Project management: external or commercial cycle time (includes a measure of the time period between the starting of the S&T project, until the ultimate sale of a product or service to an external customer).
- Existence of project champion (includes a measure of the number or portion of current S&T projects which have an identifiable "champion" in the form of a manager from *outside* the S&T/R&D unit).
- Projects with inter-functional teams (includes measures of the number of projects that employ teams composed of people from units across the organization and outside the S&T/R&D unit. Such "hybrid" teams may, for example, have representatives from marketing, production, and finance. Another measure here may be the point in the project life time in which such teams are formed. It is widely assumed that the earlier such a team is established, the greater the probability of the *commercial* success of the project)

- Evaluation of the scientific and technical capabilities of the S&T unit (and by extension of the total organization. This is a measure of external evaluation primarily by the various customers of how the firm and its S&T unit are capable of meeting the scientific and technological challenges of their changing markets. The measure may also be used to evaluate publicly funded laboratories and agencies, such as NASA, CDC, EDA, and even academic institutions)
- Project progress and success (includes measures of the progress of S&T projects in terms of the objectives and milestones that were established, over a given period of time. Also includes measures of the number or percent of projects that exhibited *technical* success and have done so on-time and on-budget).
- Evaluation of projects and programs. (This is a financial measure of the degree to which S&T projects and clusters of projects-programs have had technical *and* commercial success Measures include the cost per *technically* successful projects, and cost of *commercially* successful projects. Differences between these measures usually indicate problems in the downstream flow of S&T outcomes. Another type of measure includes averages of cost per project, by type of project. Similar measures may be applied to programs and to the entire S&T portfolio of projects and programs).
- Ownership, support, and funding of projects and programs. (These measures include percent of projects supported and funded by other units in the organization and that are directly related to a product line or similar commercial entity in the organization. These measures also provide the distribution of projects and programs by source of organization. They may indicate over or under reliance or relation to given units and functions that support S&T).
- Human relations measures of S&T personnel (including such measures as morale of S&T personnel and satisfaction with their work).
- Relation of S&T to strategic objectives. (These are measures of the degree to which S&T objectives are related to the strategic objectives of the organization, and are current with any changes in the organization's strategy. Differences or discrepancies between the two sets of objectives may lead to problems in S&T performance and its relevance to where the organization is heading.)
- Benchmarking project and program performance. (These are measures that relate financial, economic, and project management metrics to benchmarks that are standards or averages in the industry, as well as benchmarks that are established in view of the performance of the "best practices" in the industry and sector. Additional measures in this set are the extent to which these benchmarks influence the strategic direction of both S&T and the total organization.)

H. STAGES OF OUTCOMES

- Immediate outputs (includes measures of the proximal or direct outputs from the S&T/R&D activity, such as bibliometric measures).
- Intermediate outputs (includes outputs of the organizations and entities that have received the immediate outputs, transformed them, and are providing the transformed outputs to other entities in society and the economy).
- Pre-ultimate outputs (includes measures of the products and services that are generated by those social and economic entities that had received and transformed the intermediate outputs).
- Ultimate outputs (measures of the things of value to the economy and society that were impacted by the pre-ultimate outputs).
- Index of leading indicators (manipulation of *core* and *organization-specific* measures, in a weighted procedure. For each stage of outputs—immediate to ultimate—the index of leading indicators is constructed. The index offers a quantitative appraisal of the value of S&T at each of the stages of its flow downstream the innovation process).
- Value indices for leading indicators (includes measures of the value of each index, at each stage in the flow downstream the innovation continuum. Value indices are computed by subtracting the value of each leading index from the index that succeeded it. For example, the value index for an organization in the pre-ultimate stage minus the value in the intermediate stage would measure the added value to the continuum from S&T. Net value is computed by comparison with costs of S&T and transformation at each stage)
- Portion of S&T at each stage. (These are measures of the role that S&T has in each of the stages, for each of the recipient/transforming organization. The outputs from these organizations—products, services, processes, methods, etc.—are possible, in part, because of S&T absorbed and adopted by the organization. These measures offer a look at the size and value of the S&T contribution, for each output, as well as *in toto*).
- Organizations in the private as well as in the public sectors employ various combinations of these metrics, with a variety of measures, quantitative as well as qualitative.

CRITERIA FOR SELECTION

What is a good metric and what are the criteria that guide our selection of the "right" or "appropriate" metric for the evaluation of science and technology? These are important questions, in view of the fact that there is a pool of different metrics and measures available for use.

Ronald Kostoff, for example, had suggested ten principles of R&D evaluation that is based on metrics.¹¹ In addition to such principles as senior management commitment and competency of the technical evaluator, Kostoff also included (1) the relevance of the metric selected to future action by the evaluators and (2) the integration or connectivity of the metrics selected to the operations, processes, and strategy of the organization. This latter criterion is crucial to the consequent use of the metrics-based evaluation by the organization's decision-makers.

Kostoff had also proposed that the cost of using certain metrics should be a criterion in their selection. Finally, he listed: (1) availability and quality of the data; (2) complementarity of other metrics, and (3) accuracy and quality of the metrics. Thus, his criteria include the predominant factors found in the literature.

To a large extent many criteria found in the relevant literature refer to the evaluation and its components, then specifically to selection of metrics. Yet, these same criteria may also be applied to selection and usage of metrics. For example, the need to track the progress of projects over long periods of time may also be a criterion for selecting the metrics for this evaluation effort. We would select those metrics that allow for measurements across different stages of the project's time-line, at the expense of metrics that are restricted to measuring only one phase in the project's life. Another criterion that emerges from this issue is the selection of metrics that, although different, would nevertheless be comparable and allow for integration and convergence across the various stages of the project's life.

A more comprehensive attempt to classify such criteria consists of three categories: methodological, ontological, and managerial and organizational. Underlying these categories are the following objectives in selecting metrics for S&T. First, the metrics selected should be able to do the job. Second, they should be available, accessible, and affordable. Third, they should allow for manipulation, analysis, interpretation, and comparison with other metrics. Hence, we now have the necessary background to create a set of criteria, as shown in Figure 3.

Figure 3. Criteria for Selecting Metrics for Science and Technology*

CATEGORY	CRITERIA**
<p>Methodology (criteria related to measurement and analysis of metrics: What they should do and how)</p>	<ul style="list-style-type: none"> ○ Quality of underlying data. ○ Provide data that allow for meaningful analysis & interpretation. ○ Allow for standardization, generalizability, or normalization across fields and disciplines. ○ Comprehensiveness of set to measure a substantial portion of phenomenon. ○ Relative ease of data collection, manipulation, and interpretation.
<p>Ontology</p>	<ul style="list-style-type: none"> ○ Comprehensiveness,

<p>(criteria related to the construction of metrics: What they should be like)</p>	<p>covering critical aspects of the phenomenon.</p> <ul style="list-style-type: none"> ○ Integration and convergence with other related metrics. ○ Validity: measures what it purports to measure. ○ Includes measures of different aspects of the phenomenon.
<p>Organization and Management</p> <p>(criteria related to the use of metrics by managers and organizations: What they should accomplish)</p>	<ul style="list-style-type: none"> ○ Relevance to organizational and managerial objectives. ○ Relevance to potential action by managers/decision makers. ○ Credibility of metrics. ○ Capability of evaluation to deal with the metrics, to gather and to analyze them. ○ Cost and cost-effectiveness. ○ Relation and relevance to other metrics used in different evaluations.

*This list is illustrative and not an exhaustive list of all criteria.

**Not in any order of importance or relevancy.

Do The Job!

Metrics selected should be able to measure what the evaluators wish to be measured. They should be intimately linked to the objectives and motives of the evaluators. Simply measuring specific attributes of S&T is not an effective way of assessment, unless the measures indeed describe relevant aspects of the phenomenon and are useful to managers. If this criterion is not met, the metrics become a set of irrelevant and useless quantities.

In the final analysis, the indicators and metrics in general are the actual means we utilize to assess the S&T phenomena.¹²

Availability, Accessibility, and Affordability

This second objective in selecting metrics for S&T. It simply calls for the data used by these metrics to be available, accessible, and affordable. If these factors are not satisfied, the quality of the data will suffer, and in certain instances the metric will be useless since it will not contain the necessary and adequate data items.

When desired data are not available, are inaccessible or too costly, the tendency is to explore alternatives, some of which may be in the form of surrogate data items, which in many cases are inadequate or even misleading. The urge to measure and to apply metrics to the evaluation is usually strong enough as to accept compromises that are harmful to a thorough assessment of S&T.

Manipulation and Interpretation

Finally, the S&T metrics should allow for manipulation of the data thus collected, for adequate interpretation, and for comparison with other metrics. Being able to statistically manipulate data in these metrics is not enough, unless the statistical analyses can also provide an interpretation of what these metrics mean for the assessment of S&T. For example, the number of patents generated by an industrial sector and its manipulations (by cost, employees, by assets, or over time) is meaningless unless we also understand the process of patenting in the firm. Thus, by simply arriving at some numbers about patents, we lack the ability to interpret them, unless we also possess additional knowledge into which this metric can be integrated, hence interpreted.

As illustrative of currently used metrics, this paper offers below an analysis of the following metrics: (1) economic and financial; (2) bibliometric; and (3) co-word analysis and data mining. Strengths, weaknesses, and examples of uses are also discussed.

ECONOMIC AND FINANCIAL METRICS

Perhaps the most preferred metric of science and technology, economic and financial measures of S&T evaluation have been amply used for many years. Originally within the sole realm of economic analysis, these metrics have continually advanced to the fore of other forms of usages, primarily to support policy decisions on the nature of S&T outputs and for decisions on resources allocation for S&T.

Economic and financial measures of the S&T activity are a broad spectrum of measures and techniques. They range from simple financial indices of inputs to S&T, to more complex measures of cost-effectiveness and the economic impacts of S&T, such as on national and corporate productivity.

The link between productivity growth (primarily in manufacturing) and S&T has been tortuous and inconclusive. Economists measure productivity in two modes: neoclassical production function and a more recent approach that considers reproducibility of capital and the effects that occur between sectors. Much of the research utilized the neoclassical

approach, so that R&D data were considered a specialized type of resource (capital) and included in the production function. The estimated returns were quite low.

The Productivity Paradox

However, even with the use of more recent approaches, the link between R&D and productivity growth has not been satisfactorily established. This phenomenon of a puny measurable impact of R&D and S&T on growth in productivity (however defined) has been termed by economists "The Productivity Paradox."¹³

The discussion in the literature of this paradox has developed along two main lines: (1) R&D and S&T in general and their impact on industrial productivity, and (2) information technology (hardware, software, and other office automation) and their impact on industrial productivity.

In the case of the general impacts of S&T on industrial productivity, Griliches (1998) had summarized the issue in his book on the body of knowledge in the relation between R&D and industrial productivity.¹⁴ He and other economists have suggested that inadequate measures of productivity are partly to blame for the seemingly slowdown in the 1970s, 1980s, and early 1990s. The other factor seems to be the emergence and the growth of the service sector, while the manufacturing sector, although still very strong, has become smaller.

WEAKNESSES OF ECONOMIC/FINANCIAL METRICS

- 1) Difficulties measuring the outputs and impacts of S&T
- 2) Difficulties in isolating precise impacts of S&T and allocating them to inputs to S&T
- 3) Complexity of the innovation process makes it very difficult to apply ROI or ROA techniques
- 4) existence of a temporal gap between the long term nature of S&T and the short term objectives and the financial tools of corporate management.

STRENGTHS OF ECONOMIC/FINANCIAL METRICS

Figure 4. Strengths of the Economic/Financial Metrics*

- Offers quantitative measures that describe economic factors, variables, and phenomena.
- Allow for the gathering of time-series data of expenditures and other economic/financial variables, hence makes possible the statistical correlations among these data sets.
- Allows for comparisons with financial measures of other (non-R&D) activities in the

firm and in the industry.

- Under certain assumptions, conclusions can be drawn on the economic viability of R&D and the allocation of economic resources to this effort.
- Allows for a link (causal or relational) between investments in R&D or S&T, their economic outputs, and accounting and managerial ratios and indices of financial activity and financial success.
- Allows for the quantification of certain externalities and their relation to the R&D function.
- Inserts an econometric input to the inventive effort.

*This list is not exhaustive, nor in any order of importance.

OUTPUTS FROM SCIENCE AND TECHNOLOGY

Outputs from S&T are events or material/physical objects (such as products, processes, and reports), or organizational/social/economic phenomena (such as savings and productivity). They include non-tangibles such as items of knowledge, satisfaction of people, and other benefits or harmful effects raised (directly or indirectly) by S&T.

Outputs from R&D/S&T are seldom clearly and unequivocally defined in the literature. Various terms are used as descriptors of outputs, sometimes as surrogate terms, other times even as euphemisms. Terms such as "impacts," "effects," "consequences," "benefits," "value," "advantages," "spin-offs," "returns," and "payback" are some of those found in the literature on evaluation and assessment of R&D and S&T.

A major reason for the difficulty in advancing a comprehensive definition of outputs is the fact that due to the complexity of the innovation continuum, there are many different outputs. They are clustered in multiple and diverse sets, all along the continuum from research to engineering to commercialization. Such a variety of outputs leads to issues in measurement, linkage across gaps, and the use of selected outputs in different schemes of evaluation.¹⁵

CATEGORIES OF S&T OUTPUTS

A survey of the literature and the practices of universities, industry, and government reveals a large number of outputs from R&D and from S&T. These outputs are shown in a list of generic outputs ("laundry list") in Figure 5.

Figure 5. A Generic List of Outputs From Science and Technology*

- Science and technical ideas**
- Scientific and technical publications, reports, and citations
- Intellectual challenges
- Technical assistance
- Presentations to learned societies
- Training of scientific and technical people
- New and improved products, materials, and processes
- Patents
- Transfer of technology
- Development of new testing methodologies
- Development of R&D/S&T management practices and techniques
- Start-up of new ventures, new companies, establishment of partnership
- Development of strategic technology alliances
- Development of scientific and technical benchmarks and standards
- Cost-savings in production, product design, and re-design
- Increased productivity and utilization of resources
- Improved product/process/service quality
- Reduced dependence on outside sources
- Facilitator in ability to outsource
- Savings in materials
- Contribution to maintenance/protection of lead or position in the discipline/industry/market
- Facilitation of use by client
- Contribution to adequate response to environmental and other regulatory pressures
- Contributions to potential adaptability of manufacturing to new processes and methods
- Contributions to the competitive features of a product or product-line
- Contributions to creation of new market,

- market segments, and new customers
- Contributions to technology and business planning, and to the strategic management of the organization
- Development, manipulation, and exchange of new knowledge in S&T

- Provision of scientific and technical information to assist managers in areas such as licensing, mergers and acquisitions, and other activities imbued with content of S&T
- Contributions to institutional memory
- Contributions to the identification of opportunities and needs for S&T
- Contributions to improved project selection and resources allocation for S&T and for the innovation process
- Contributions to sales, profits, and other economic criteria of performance
- Contributions to the perception of S&T by the sponsors of this activity and by the public at large
- Increased ability to anticipate and to effectively deal with barriers to application and implementation of results from S&T
- Contribution to expending the state of the art in S&T
- Contribution to the prestige of S&T organizations and their impactees

**This is a comprehensive but not an exhaustive list. Some outputs not listed here may appear embedded in these outputs that are listed here, or perhaps used as independent measures.*

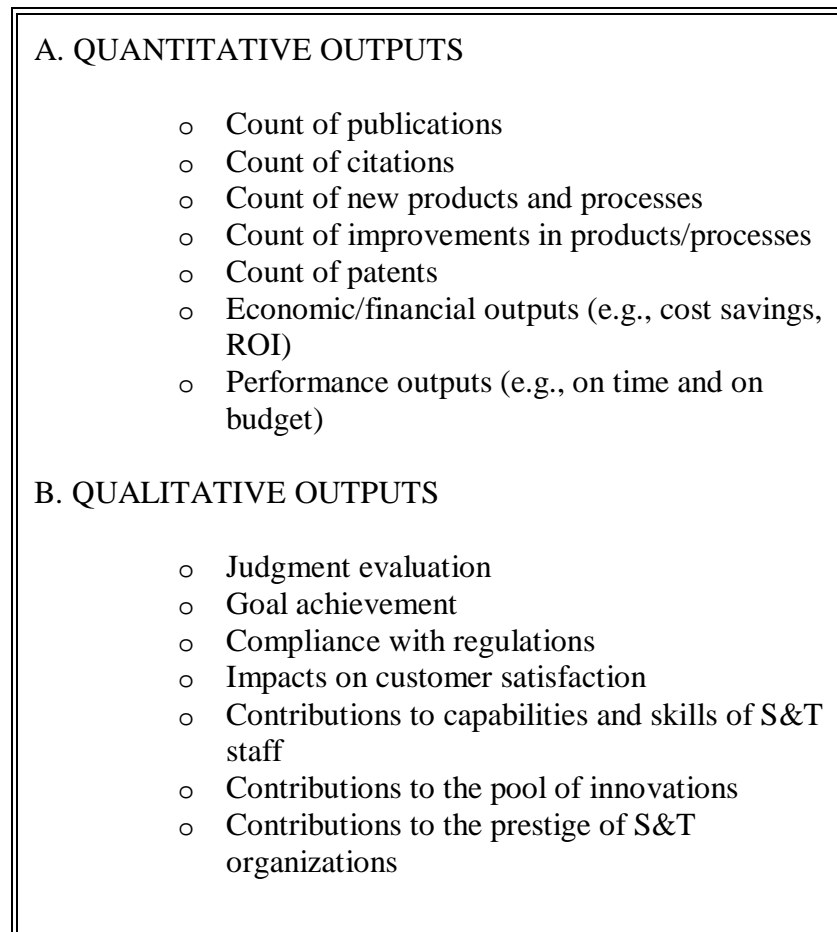
***Not necessarily in any order of priority or other ordinal scale.*

QUANTITATIVE AND QUALITATIVE OUTPUTS

This taxonomy is based on the criterion of how the output is measured: by the use of some quantities or by a qualitative method. Figure 6 shows the distribution of some outputs as they are clustered into these categories.

As is the case with taxonomical analysis, there is not a very definitive distinction between quantitative and qualitative outputs and their measures. For example, although peer review is a process of combining the judgmental evaluation of individuals or a committee, these subjective assessments lend themselves to scoring, hence to some manipulation of quantities.

Figure 6. Illustrative Outputs From S&T By Degree of Quantification *



**This list is not an exhaustive listing of all outputs from S&T. Classification into the categories was a subjective decision by the author*

DIRECT AND INDIRECT IMPACTS OF S&T OUTPUTS

Figure 7. Illustrative Outputs From S&T by Direct or Indirect Impacts*

- | |
|---|
| <p>A. DIRECT IMPACTS**</p> <ul style="list-style-type: none">•Publications•Citations•Patents•S&T performance•Improved products and processes•New products and services <p>B. INDIRECT IMPACTS***</p> <ul style="list-style-type: none">•Start-up of new companies•Improved performance of product line•Cost savings and reductions in such areas as production, design, and maintenance•Compliance with regulations•Goal achievement |
|---|

**This list is not an exhaustive listing of all outputs from S&T.*

***Impactees directly receive/acquire the outputs from those who generate them.*

****Impactees are affected by factors that are related to S&T outputs.*

OUTPUTS FROM SCIENCE AND FROM TECHNOLOGY

Figure 8. Illustrative Outputs From Science and From Technology*

- | |
|----------------------------------|
| <p>A. OUTPUTS FROM SCIENCE**</p> |
|----------------------------------|

- Ideas
 - Knowledge
 - Publications
 - Citations
 - Prestige
 - Training and education
 - New methods and techniques
 - Intellectual challenges
- B. OUTPUTS FROM TECHNOLOGY**
- Patents
 - New products and processes
 - New materials
 - Technical assistance
 - Technical benchmarks and standards
 - Contributions to productivity
 - Improved quality
 - Cost-savings

**This list is not exhaustive. Classification of the outputs into the categories was done by the author based on the literature and personal judgment.*

***Not necessarily in any order of importance or other criteria.*

BIBLIOMETRICS

Bibliometrics is a general term that refers to measures of scientific and technical *published* outputs from science and its publications in scientific and technical journals, and the indexing of citations of these publications by other scientists in these journals. The first measures *quantity* of these outputs, the second *quality*.

But bibliometricians, those who professionally count and analyze these outputs have argued that there is more to these measures than just counts. They see in the data from these measures strong indices of the progress and the dynamics of academic disciplines, and the generation and evolution of new disciplines.

HOW BIBLIOMETRIC ANALYSIS WORKS

Bibliometrics is a procedure that counts the number of S&T publications and other means of knowledge transfer, and also performs an analysis of citations and co-citations in order to assess the quality of these publications and their trends. Bibliometric analyses are routinely conducted by every researcher who investigates the "state-of-the art" on a given topic in the researcher's disciplinary area. But a more comprehensive analysis is performed by commercial organizations dedicated to this issue, by some government agencies, and by selected universities.

STRENGTHS OF BIBLIOMETRICS

Figure 9. Strength of Bibliometrics in the Evaluation of S&T*

<p>A. STRUCTURE</p> <ul style="list-style-type: none">○ Bibliometrics can be applied to various levels of generators of intellectual outputs, such as individuals, groups, institutions, and countries.○ The cost of collecting the data and conducting meaningful analysis is relatively adequate.○ The measures are already built into the metric, thus there is no need to establish them and to test them for validity. <p>B. MEASUREMENT</p> <ul style="list-style-type: none">○ Bibliometrics allow for quantitative assessment of S&T outputs by counts of papers and citations, and for qualitative assessment by analysis of core journals and their relative impacts.○ Bibliometrics and their analysis is a relatively
--

straightforward approach, relying on a few assumptions.

C. REPRESENTATION

- Bibliometrics can be applied to the entire spectrum of S&T where outcomes take the form of reports, papers, and citations.
- Bibliometrics, through citation analysis, help to determine the role that individuals and institutions have in the evolution of a scientific discipline.
- Bibliometrics analysis allows for the identification of trends and developments in science and technology and in scientific disciplines.
- By convention, bibliometrics have been accepted by the S&T community as valid representation of the outputs from intellectual and inventive activities.

*This list is not exhaustive nor in any specific order.

WEAKNESSES OF BIBLIOMETRICS

Figure 10. Weaknesses and Problems with Bibliometrics*

A. COVERAGE

- Published articles are only one measure of outputs from scientific activity, hence the metric does not cover reports, other written communications such as electronic mail, letters, and personal communiques.
- Articles published in peer-reviewed journals and their citations analysis disregard the outputs and intellectual contributions in articles published in technical outlets, as well as work-in-progress.

B. MEASUREMENT

- The "Pied-Piper Effect": bibliometrics, particularly citation analysis, measures influence, not quality. Citations are selective and refer to those papers that "toe the line" and do not "rock the boat."
- Published articles measure output in a given sub-discipline or discipline, hence cross-disciplinary analysis may be difficult to validate, because of the different structure and procedures of the scientific investigation in each discipline; particularly ease and rate of publishing, and the nature of the peer selection processes (the "Apples and Oranges Effect").
- Count of publications and citations analysis tend to disregard the influence of the stage in the life of the discipline or area, such as "mature" area versus an "evolving" new area ("The Gulliver Effect").

C. GENERALIZABILITY

- Count of publications and citations lack a standard for their validation as a measure of quality. When compared with inputs (investments in R&D), the resulting analysis relies on co-variation of two distinct phenomena and disregards the complexities of the R&D process.
- The only standard for validation of bibliometrics is convention of a small and elite group of influential scientists.
- As in the "Hindsight" project, the problem is: "How far in time should citations go?" Should articles in quantum mechanics cite Einstein's work *and* that of the Greek philosophers and mathematicians? Citations are thus highly selective and refer to a relatively short time-frame, preferably within a few years of the focal paper in which they are cited.

D. BIASES

- Because of the almost incestuous nature of the small group of prolific publishers, there is an inordinate amount of self-citations and citations of "friends" and other members of this elite group. Thus, authors who

publish in areas that are near the boundaries of their discipline or in cross-disciplinary topics are much less likely to be published in top journals or to be cited in them.

- Criteria for the selection of articles for publication in juried journals are built into the process and will bias the resultant counts of papers and citations in favor of those authors preferred by the reviewers.
- Selection and analyses of key journals and the interpretation of the counts of papers and citations are based on assumptions of validity of these metrics as measures of quality and internal dynamics of the discipline. Such assumptions are the product of opinions and judgment, hence biased.
- The selection and analysis process of key journals and papers to be included in this process is biased in that analysts who are generally outside the discipline or the area impose their own views and criteria in making determinations and drawing conclusions that transcend the data.

*This list is not exhaustive nor in any specific order.

PROCESS-OUTCOMES METRICS

Process-outcomes metrics are based on a model of stages of the innovation process. The model has four stages or categories of outputs: (1) *immediate* or direct/proximal outputs such as publications and patents; (2) *intermediate* outputs, which are outputs of the transforming organizations such as companies, and inputs to social and economic subsystems, such as new products, new materials and new methods; (3) *pre-ultimate* outputs, which are products and services generated by the economic and social subsystems; and (4) *ultimate* outputs, which are those things of value to the economy and society.

The *immediate* outputs are the outcomes from the R&D activity itself, measured by indicators such as publications, citations, and patents. These outputs are then transferred to transforming organizations such as the business entities of the corporation, engineering, and manufacturing. These entities process and transform the immediate outputs they have absorbed. The outcome of these transformations are the *intermediate* outputs, measured by indicators such as new products, new materials, new methods for analysis, reductions in cost and improved productivity. These are outputs internal to the organization. These two categories of outputs are usually used in assessments of the R&D function. Intermediate outputs are compared with inputs to R&D and with the immediate outputs, particularly in economic models.

However, the innovation flow includes two additional categories of outputs seldom used in R&D assessment. *Pre-ultimate* outputs are generated by the social and economic entities which then export them downstream to their environment. Such outputs are external to the organization. They are measured by such indicators as improved competitiveness, opening new markets, and finding a cure for a disease. Finally, these *pre-ultimate* outputs are absorbed by the economy and the society and produce the ultimate outputs, measured by such indicators as quality of life, economic growth, as well as negative effects such as increased pollution and other environmental concerns.

The stages in the R&D/innovation flow are temporal and conceptual. They represent a serial movement of outcomes along a time dimension, and are also measuring distinct phenomena. For example, immediate outputs assess the outcomes from the R&D activity itself, whereas intermediate outputs assess outcomes from the business (or non-R&D in not-for-profit organizations) side of the organization. In this latter case, the generation of the intermediate outputs is governed by a set of factors very different from those in the R&D phenomenon (see Figure 11).

The movement from one stage of outputs to another is negotiated through diffusion mechanisms that are inherent in the innovation process. It flows by the mechanisms of diffusion and the factors hindering or facilitating the flow as explanatory variables needed for an in-depth analysis of each individual case of R&D/innovation flow and the host organization. Problems of imputation are not eliminated, but they are reduced by the tracking of the flow through the various stages or categories of outputs. This is primarily due to the fact that additional insights are obtained of the way the R&D/innovation flow occurs and assessed at progressive milestones, rather than in one comprehensive "leap of faith" from *inputs* to R&D—to *intermediate* or *pre-ultimate outputs*.¹⁶

The model thus tracks the downstream contributions of R&D via sets of leading or key indicators. These indicators are roughly grouped in two categories: (1) *core* indicators, which are applicable across R&D organizations (generic indicators), and (2) *organization-specific*, which are unique to the organization performing the research.

There may be a cross-over between the core and organization-specific indicators, so that they are not mutually exclusive. It is assumed that the core indicators measure the phenomenon of the performance of the R&D organization. Organization-specific indicators measure attributes of the focal organization.

Figure 11. A Preliminary Conceptual Model of the Linkages Between the R&D Process and Social Systems

Inputs to R&D (in particular R&D sector or the total national R&D system), e.g.

People and Skills
Funding

Figure 12. Illustrative Leading Core Output Indicators for the Immediate and Intermediate Output Categories

1. IMMEDIATE OUTPUTS

Core Indicators

1. Index of written scientific and technical outputs

Measures

- 1.1 Number of publications in refereed journals
- 1.2 Number of technical reports
- 1.3 Number of patents
- 1.4 Number of patent disclosures
- 1.5 Number of citations in refereed journals

2. Index of hardware/software/other outputs

Measures

- 2.1 Number of new products conceived
- 2.2 Number of key improvements suggested
- 2.3 Number of new and improved test methods, models, standards, concepts, and databases transferred downstream
- 2.4 Number of new ideas transferred downstream
- 2.5 Number of problems solved for users/clients downstream
- 2.6 Improved understanding of phenomenon

3. Index of overall reputation of R&D performers

Measures

- 3.1 Number of complaints by clients/users
- 3.2 Judgment by clients/users
- 3.3 Judgment by other R&D performers (peers)
- 3.4 Number of awards received
- 3.5 Milestones/objectives met

II INTERMEDIATE OUTPUTS

1. Index of scientific/technical impacts on direct user of R&D outcomes

Measures

- 1.1 Number of improved and/or new products
- 1.2 Number of improved and/or new materials
- 1.3 Number of improved and/or new tests and methods of analysis

2. Index of economic impacts on direct user of R&D outcomes

Measures

- 2.1 Actual cost reduction/savings in products or processes

- 2.2 Actual improvement in productivity of material/equipment/techniques or people
- 2.3 Actual improvements in performance in sales, profits, sophistication of new and/or improved products and services

3. Index of responsiveness of R&D

Measures

- 3.1 Judgment by the direct client/user downstream
- 3.2 Judgment by other (non-R&D) organizations
- 3.3 Judgment of other R&D performing organizations

Figure 13. Illustrative Organization-Specific Indicators for the Immediate and Intermediate Output Categories

1. IMMEDIATE OUTPUTS

Organization Specific Indicators

1. Level of technical expertise

Measures

- 1.1 Ratio of doctorate holders to scientific workforce
- 1.2 Relative experience of R&Es: total years of technical work

2. Attractiveness of R&D organizations

- 2.1 Number of candidates applying for each position in the technical/scientific area
- 2.2 Age profile of S&Es
- 2.3 Judgment by other R&D performers (peers)

II. INTERMEDIATE OUTPUTS

1. Level of investment in exploitation of R&D outcomes

Measures

- 1.1 Annual funds allocated to technology commercialization
- 1.2 Number of personnel from non-R&D units working with R&D

2. Level of importance of R&D outcomes

Measures

- 2.1 Role of new products/services in the organization's success and survival (judgmental)
- 2.2 Perceived success (trade record) of outcomes transferred from R&D in the organization's performance (judgmental)
- 2.3 Judgment by other organizations (peers)

ILLUSTRATIVE PRACTICES

How do successful companies evaluate this technology? In this section I provide two illustrations of different companies and their unique way of assessing the impacts and benefits of technology. In general, the "best practice" in the examples below are all based on very few chosen methods and measures. Perhaps it is not the method or measure used, but *how* it is counted and assessed that makes all the difference.

General Electric

Walter Robb, who was GE's senior vice-president for corporate research and development, had pointed out that his company has used a mix of subjective and more objective measures. An overall assessment is based on the subjective view of the company and its successes, *without* the existence of the R&D laboratory and the technology it has generated over the years. Robb calls this the "Jimmy Stewart Test," based on the actor's role as Mr. Bailey in Frank Capra's "It's a Wonderful Life." In this movie Mr. Bailey is shown how different and less valuable his town and the life of its inhabitants would have been if he hadn't lived.

In the area of more precise and objective measures, GE has resorted to a count of patents and various attributes of patents, such as cost per patent, comparisons with other companies, and income from licensing. Another measure was the amount of royalty payments the company saves by having a specific patent.

Finally, GE also used a discounted rate-of-return applied to the transfer of a technology from the laboratory to a commercial use. To ensure objectivity, GE has asked a consulting company, Booze, Allen, & Hamilton, to calculate the return on investments. By using a methodology similar to that of historical description of key events used in the Bullpup and Traces studies, the consultants conducted well-directed interviews with key informants in both the technology center and the business units. They then proceeded to assess the level of credit that the technology center can be assigned of the benefits from each of the 190 such technology transfer projects they had analyzed.

Based on the level of credit, the part played by the technology/R&D center in the returns from each project/product was determined. Compared with the expenditures for the R&D center, a rate of return was thus computed.

Robb had emphasized that the methodology thus described had also helped to alert GE to the fact that a very small number of products (technologies) generated by the center and transferred to commercial applications had accounted for much of the payback. This knowledge, according to Robb, was instrumental in redesigning the company's allocation of resources for R&D.

Eastman Kodak

At Eastman Kodak the preferred methodology for evaluating technology was a modified risk analysis described by the company as an "R&D Portfolio Analysis." The methodology is driven by the business aspects of desired products. Two main flow

diagrams are constructed. The first starts with the features of the product and contains the other factors that will influence the net-present-value of the product once it is marketed. The second flow-diagram (described by the company as an "influence diagram") shows the various R&D tasks or activities needed to accomplish specific features.

What follows is the application of probabilities (in the form of a decision tree) of the potential performance of the R&D tasks. How probable it is that specific tasks will generate the R&D outputs (technology) necessary for the achievement of product features (such as sharpness, speed, and grain).

In essence, the methodology compares what a certain product, with proposed features, will gain in the marketplace—with the potential of the R&D function in the firm to generate technology that would help form such a product. Thus, when dividing such a value of R&D by the projected cost of this R&D effort, what emerges is an indication of how well R&D performs.

Inherent in this methodology and in the "influence diagrams" is the assumptions that the risks are fairly well understood, and that all the critical stages and tasks have been included in the analysis. The methodology starts with the product itself and its projected successful commercialization in the marketplace. It then looks back to "what can R&D do for us in this specific case?".

This method had also contributed to reallocation of resources at Eastman Kodak, in favor of those activities that had shown higher commercial promise. Such bias is consonant with the more stringent non-linear models of research and technology evaluation.

WHERE WE SHOULD GO FROM HERE

The following are key points of the agenda for measuring technological progress and transfer:

1. Use mix of metrics, such as bibliometrics, peer-review, and process-outcomes.¹⁷
2. Process-outcomes is essential to gaining true understanding of how S&T progress and how it impacts organizations, society, and the economy.¹⁸
3. Use a mix of quantitative *and* qualitative/judgment metrics.¹⁹
4. Take into account the weaknesses of metrics, so that promises as to their benefits can be more realistic.²⁰
5. Metrics and measurement by themselves are of little value.²¹ They must be integrated with social-economic systems.²²

NOTES AND REFERENCES

1. This paper is based on the author's two decades of research and consulting in the area of evaluation and measurement of science and technology. The main source for material for this paper is: Geisler, E., *The Metrics of Science and Technology: Evaluation and Measurement of Research, Development, and Innovation* (Westport, CT: Quorum Books, forthcoming 2000).
2. Geisler, E., "Industry-University Technology Cooperation: A Theory of Inter-Organizational Relationships," *Technology Analysis & Strategic Management*, 7(2), 1995, 217-229.
3. Werner, B., and W. Souder, "Measuring R&D Performance: State of the Art," *Research-Technology Management*, 40(2), 1997, 34-42.
4. Geisler, E., "Measuring the Unquantifiable: Issues in the Use of Indicators in Unstructural Phenomena," *International Journal of Operations and Quantitative Management*, 1(2), 1995, 145-161.
5. Kostoff, R., "Metrics for Strategic Planning and S&T Evaluation," *R&D Enterprise—Asia Pacific*, 1(2-3), 1998, 30-33.
6. Geisler, E., "Key Output Indicators in Performance Evaluation of Research and Development Organizations," *Technological Forecasting and Social Change*, 47(2), 1994, 189-204.
7. Geisler, E., and A. Rubenstein, "Evaluating the Outputs and Impacts of R&D Innovation," *International Journal of Technology Management*, Special Publication on the Role of Technology in Corporate Policy, 1991, p. 181-204.
8. Hauser, J., and G. Katz, "Metrics: You Are What You Measure," *European Management Journal*, 16(5), 1998, 517-528.
9. Geisler, E., and A. Rubenstein, "The Use of Indicators and Measures of the R&D Process in Evaluating Science and Technology Programs," In: J. D. Roessner (Ed.), *Government Innovation Policy* (New York: St. Martin's Press, 1988), pp. 185-204.
10. Geisler, E., and A. Rubenstein, "Long Range Planning of Corporate Research and Development," *Planning Review*, 13(2), 1985, 39-42.
11. Kostoff, 1998, *op. cit.*
12. Geisler, E., "An Integrated Cost-Performance Model of Research and Development Evaluation," *Omega*, 23(3), 1995, 281-294.
13. Wakelin, K., "R&D and Productivity: The Econometric Evidence," *The World Economy*, 22(1), 1999, 148-149.
14. Griliches, Z., *R&D and Productivity: The Econometric Evidence* (Chicago: University of Chicago Press, 1998).