5-1-1985

Measuring R and D Productivity

Richard A. Pappas

Donald S. Remer
Harvey Mudd College

Recommended Citation

Pappas, Richard A., and Donald S. Remer. "Measuring R&D Productivity." Research Management 28.3 (1985): 15-22.

MEASURING R&D PRODUCTIVITY

Richard A. Pappas and Donald S. Remer

Although current quantitative techniques are not satisfactory, certain semi-quantitative techniques based on qualitative judgments can be effective.

Measuring the productivity of an R&D organization is extremely tricky. Productivity is usually defined as a ratio of an output, like number of cars produced on an assembly line, to an input, like the wages paid the workers. While R&D may have a measurable input, the output is often intangible and difficult to quantify. This is further complicated because the return from an R&D department may not be realized for one or two decades, which means the time lag is much higher than in factory measurements. Furthermore, many researchers believe that this kind of measurement may be counter-productive, since the mere act of measurement could reduce R&D productivity. Nevertheless, companies continue to evaluate R&D with the crude methods available as they desperately look for more effective, quantitative methods.

After reviewing the literature, we divided the R&D evaluation techniques into three general categories: (1) quantitative, (2) semi-quantitative, and (3) qualitative. *Quantitative techniques* usually follow a specific algorithm or predefined ratio to generate numbers that can be compared with other projects and past experiences. In many cases, this involves having key managers rate different aspects of the effectiveness and importance of the project using probabilistic weighting factors. These numbers are then combined using a rigid algorithm, as described later in this article.

Semi-quantitative techniques are basically qualitative judgments that are converted to numbers. These techniques differ from quantitative techniques in that no attempt is made to use a sophisticated formula to compile the data, though techniques like averaging are sometimes used to simplify the output.

Qualitative techniques are intuitive judgments. We will not analyze qualitative techniques in detail because our survey was aimed at quantitative methods, and because little has been written about qualitative methods.

Richard Pappas is a regional sales engineer at Hewlett-Packard. He received his B.S. in engineering from Harvey Mudd College, Claremont, California. Donald Remer is the director of the Energy Institute and the Oliver C. Field professor of engineering at Harvey Mudd College. He received his M.S. and Ph.D. in chemical engineering and business economics from California Institute of Technology and his B.S. from the University of Michigan. Prior to coming to Harvey Mudd, he served with Exxon Chemicals in several positions including forecasting, planning, economics, coordination, and technical service. Prof. Remer is also a founding partner of the Claremont Consulting Group which specializes in engineering and project management and economics, management training, and energy and environmental analysis.

Qualitative techniques are, however, in widespread use today.

In our literature search and interviews with some 20 experts, we found that people using today's measurement methods do not accurately define what stage of research they are attempting to measure. This i a major flaw, not only in current efforts to improve techniques to measure R&D productivity, but also in th application of the methods already in use. The problem is that R&D has so many different stages, and that no single measurement technique is best at each stage. Thus, we propose the system shown in Table 1, where each of the three evaluation techniques described earli are compared with the research stage to which they ar best suited. Understanding this simple figure is imperative before a useful analysis of R&D may be attempted, since it reflects the current areas where quantitative measures are most applicable.

The R&D stages can be defined as:

1. Basic Research—directed to the search of fundamental knowledge.

2. Exploratory Research—to determine if some scientific concept might have useful application.

3. Applied Research—directed to improving the practicality of a specific application.

4. Development—engineering improvement of a particular product or process.

5. Product Improvement—directed to changes for a product or process that can increase its marketability, reduce its cost, or both.

In basic research, a quantitative method is less applicable because the output is often too abstract. Thus, most companies use a qualitative method based upon the intuitive feel of managers to evaluate basic research. But, on the other end of the scale, product improvement usually has a more quantifiable output that is more easily modeled by a rigid algorithm. As a result, quantitative techniques used today are usually aimed at this stage of R&D, though it is generally not explicitly stated.

Between these two extremes, there is a mix of techniques used, but often the semi-quantitative approach proves to be the most useful. Applied researc usually does not produce an output that is readily quantifiable. For this reason, the rigid algorithm of a quantitative technique is usually not applicable given today's state-of-the-art. However, often the output is n

MEASURING R&D PRODUCTIVITY

Richard A. Pappas and Donald S. Remer

Although current quantitative techniques are not satisfactory, certain semi-quantitative techniques based on qualitative judgments can be effective.

Measuring the productivity of an R&D organization is extremely tricky. Productivity is usually defined as a ratio of an output, like number of cars produced on an assembly line, to an input, like the wages paid the workers. While R&D may have a measurable input, the output is often intangible and difficult to quantify. This is further complicated because the return from an R&D department may not be realized for one or two decades, which means the time lag is much higher than in factory measurements. Furthermore, many researchers believe that this kind of measurement may be counter-productive, since the mere act of measurement could reduce R&D productivity. Nevertheless, companies continue to evaluate R&D with the crude methods available as they desperately look for more effective, quantitative methods.

After reviewing the literature, we divided the R&D evaluation techniques into three general categories: (1) quantitative, (2) semi-quantitative, and (3) qualitative. *Quantitative techniques* usually follow a specific algorithm or predefined ratio to generate numbers that can be compared with other projects and past experiences. In many cases, this involves having key managers rate different aspects of the effectiveness and importance of the project using probabilistic weighting factors. These numbers are then combined using a rigid algorithm, as described later in this article.

Semi-quantitative techniques are basically qualitative judgments that are converted to numbers. These techniques differ from quantitative techniques in that no attempt is made to use a sophisticated formula to compile the data, though techniques like averaging are sometimes used to simplify the output.

Qualitative techniques are intuitive judgments. We will not analyze qualitative techniques in detail because our survey was aimed at quantitative methods, and because little has been written about qualitative methods. Qualitative techniques are, however, in widespread use today.

In our literature search and interviews with some 20 experts, we found that people using today's measurement methods do not accurately define what stage of research they are attempting to measure. This is a major flaw, not only in current efforts to improve techniques to measure R&D productivity, but also in the application of the methods already in use. The problem is that R&D has so many different stages, and that no single measurement technique is best at each stage. Thus, we propose the system shown in Table 1, where each of the three evaluation techniques described earlier are compared with the research stage to which they are best suited. Understanding this simple figure is imperative before a useful analysis of R&D may be attempted, since it reflects the current areas where quantitative measures are most applicable.

The R&D stages can be defined as:

1. Basic Research—directed to the search of fundamental knowledge.

2. Exploratory Research—to determine if some scientific concept might have useful application.

3. Applied Research—directed to improving the practicality of a specific application.

4. Development—engineering improvement of a particular product or process.

5. Product Improvement—directed to changes for a product or process that can increase its marketability, reduce its cost, or both.

In basic research, a quantitative method is less applicable because the output is often too abstract. Thus, most companies use a qualitative method based upon the intuitive feel of managers to evaluate basic research. But, on the other end of the scale, product improvement usually has a more quantifiable output that is more easily modeled by a rigid algorithm. As a result, quantitative techniques used today are usually aimed at this stage of R&D, though it is generally not explicitly stated.

Between these two extremes, there is a mix of techniques used, but often the semi-quantitative approach proves to be the most useful. Applied research usually does not produce an output that is readily quantifiable. For this reason, the rigid algorithm of a quantitative technique is usually not applicable given today's state-of-the-art. However, often the output is not

---

Richard Pappas is a regional sales engineer at Hewlett-Packard. He received his B.S. in engineering from Harvey Mudd College, Claremont, California. Donald Remer is the director of the Energy Institute and the Oliver C. Field professor of engineering at Harvey Mudd College. He received his M.S. and Ph.D. in chemical engineering and business economics from California Institute of Technology and his B.S. from the University of Michigan. Prior to coming to Harvey Mudd, he served with Exxon Chemicals in several positions including forecasting, planning, economics, coordination, and technical service. Prof. Remer is also a founding partner of the Claremont Consulting Group which specializes in engineering and project management and economics, management training, and energy and environmental analysis.

Table 1—General Uses of Evaluation Techniques

RESEARCH FUNCTION	EVALUATION TECHNIQUE
BASIC RESEARCH ↓	QUALITATIVE ↓
EXPLORATORY RESEARCH ↓	
APPLIED RESEARCH ↓	SEMI-QUANTITATIVE ↓
DEVELOPMENT ↓	
PRODUCT IMPROVEMENT	QUANTITATIVE

as abstract as basic research, so that it is possible to assign quantitative values to qualitative judgments. Thus, the best measurement technique for this R&D stage is one where the evaluations of persons near the projects are quantified, i.e., a semi-quantitative technique. Therefore, as a concept filters through the different stages of R&D, all three evaluation techniques could be used as shown in Table 1. However, it should be noted that this is a *general* trend, and exceptions are possible.

The Intent of Productivity Measurement

Whether R&D professionals like it or not, the productivity measurement that management seeks is bound to be used to allocate salary raises and bonuses. These rewards may not only be personal, but they may dictate which projects are funded, which is perhaps of more importance to the R&D scientist. Thus, the measurement will provide incentive to produce those outputs that top management deems necessary.

But before R&D productivity is ever measured, top management must first *decide what they expect to get* from their research center, and, second, *what is the intent* of the productivity measurement system. Some research centers are nothing more than glorified technical service centers, because they are being evaluated on a short-term basis rather than long-term payouts.

This philosophy is reflected in the measurement techniques used today. The current algorithms for quantitatively measuring R&D can only model incremental improvements rather than the dramatic breakthroughs. If management runs its business based upon only quantitative measures, then R&D would be rewarded better if they concentrate on product improvements, which are outputs that can be more easily quantified. But this incentive is in reality short-sighted, because it is the breakthrough technologies that can propel a business to more fruitful horizons. Incremental improvements can be very useful, but there is a point in the life of a technology when the time and money spent on another incremental improvement could have diminishing returns.

Recognizing when the returns for a technology are diminishing is the basis for an analysis technique of R&D projects proposed by Richard Foster at McKinsey & Co. (*1*) Foster claims that technological progress proceeds along an S-shaped curve when a plot of effort versus performance is drawn. The slope of this curve is considered to be R&D productivity, and its peak occurs at the midpoint. The idea here is to follow the S-curve of a technology until this peak is reached, and then switch to a new technology and a new S-curve. This has the advantage of never expending high effort for small gain in a mature technology. Thus, to improve R&D productivity, a business should concentrate on technologies that have the most technical and economic potential, and recognize when it is time to move on to a new technology. Of course, the difficulty here is knowing when to switch technologies, but it is clear that Foster believes that the most productive research centers will provide many different opportunities to pursue.

Former director of research at Cyprus Research, Larry Ferreira, told us he agrees with this philosophy, but with a slight twist. He claims that to survive, businesses must "*renew*" themselves. He defines renewal as being equal to profits plus depreciation, and strategies should be built around maximizing renewal. The function of R&D, then, is to establish new technologies and ideas to renew the entire business, not to be a technical service center. In fact, Ferreira's R&D department was so good at producing new products and opportunities that the president at Cyprus once told him to slow down. His reply was simply that he had constructed a "candy store" of products so that top management had a complete array of "candy" to try, provided the time was right. Thus, both Foster and Ferreira agree that the intent of R&D should be to provide the company with many possibilities for renewal, and that measurement of R&D should motivate scientists to that end. Unfortunately, measuring those developments which achieve the goals and objectives of the company can be the most difficult, and herein lies the problem with today's measurement methods.

Quantitative Techniques for Measuring Productivity

1. *Benefits.*—Management wants to have numerical data to help them with their decisions. The reasons are obvious—numbers are easily compared, both between companies and historically within a company. Thus, management hopes to make *better decisions* with a quantitative measurement of productivity, since they would be able to tell whether R&D is becoming more or less productive. These measurements would also help management allocate funds and resources, and provide insight to selection of future ventures.

Another advantage of the quantitative technique is the valuable information that will be discovered during the quantification process. First, the goals and direction of the R&D department must be understood. Then, the development of a quantitative measure requires exploring the communication lines and idiosyncrasies of the R&D department. An analysis like this of an R&D department is rarely done.

Quantitative techniques will probably never develop to the point where a generalized formula fits all stages of research in all fields. On the contrary, it is this *tailoring* of the measurement process that makes the rigid quantitative method so attractive.

2. *Problems.*—Forcing the efforts of an entire R&D department into a rigid formula is not practical or desirable given today's state-of-the-art in quantitative measurement of R&D productivity. There are subtle differences between projects. How can one project that searches for fiber optics applications be compared with another that finds methods for better fiber optics materials? The ramifications of each are far different, and quantitative methods can falsely treat each project alike. Furthermore, quantitative methods are not sensitive to the subtle differences in work effort by individuals on a project.

Another major problem with quantitative techniques is the unavoidable time lag in judging the effectiveness of an R&D project. Roland Mueser, supervisor of innovative studies at Bell Laboratories, believes that a time lag of 7 to 19 years is about average for an R&D endeavor. If that is true, then estimating the future potential of an R&D project is like predicting the stock market 19 years from now! But quantitative techniques are only useful to management if they can measure current productivity, not what it was like ten years ago.

R&D is difficult to measure quantitatively. Typically, productivity measurements can best be made on *repetitive tasks* as opposed to one-shot, creative ones. But the work performed in R&D does not appear repetitive to the casual observer. The authors believe that research should be directed to finding models that relate a common thread of repetition in the creative development of ideas, as outlined in the earlier diagram. Until that is found, quantitative technqiues will continue to monitor those tasks which are most repetitive, i.e., incremental product improvements. Unfortunately, these algorithms do not consider breakthrough ideas, which are one of the greatest benefits from R&D.

3. *The Business Opportunity Concept.*—The business opportunity concept is a quantitative technique for the measurement of R&D productivity that is presently being used at Borg-Warner and has been discussed in articles by Donald Collier of Borg-Warner and Robert Gee of Du Pont (*2,3*). The measurement system used at Borg-Warner actually combines all three techniques mentioned earlier, namely quantitative, semi-quantitative, and qualitative. The semi-quantitative portion is discussed in the next section.

The business opportunity concept is based upon the premise that the objective of research is to generate opportunity. Thus, research productivity is measured in terms of the amount of opportunity generated. Efficiency is then measured in terms of opportunity generated per dollar expended, since a more efficient R&D organization will generate opportunities at a minimum cost. This method is used to evaluate

Forcing the efforts of an entire R&D department into a rigid formula is not practical or desirable given today's state of the art in quantitative measurement of R&D productivity.

opportunity of entire projects that have "transisted"—that is, R&D has completed their efforts as prescribed by another department at Borg-Warner, so it is in a transitional phase. Thus, the business opportunity concept cannot be used to evaluate individual participants in a project, but rather rates their collective efforts.

To determine a new technology's business opportunity, four steps are followed:

- Estimate the market for the newly developed technology.
- The customer's total cost to accomplish the function is then estimated, assuming that the best present alternative to the newly developed technology is used. This will establish how much a customer might be willing to pay.
- Using the customer's cost as a ceiling, a price is set on the newly developed technology by working backward.
- Annual income resulting from sales of the new technology to the entire market is calculated, using the hypothetical price. (See example below.)

The annual income generated is the business opportunity that a specific development might realize. It is unrealistically high since capture of the entire market is unlikely, but it has been useful in making project comparisons. Furthermore, a return-on-research index can be constructed by adding together the annual income for each project and then dividing by the total cost of operating the research unit. This index can be used to make year-to-year comparisons.

Gee uses a rigid formula derivation to illustrate just how this technique can work. A simple example provided by Collier more easily describes the idea. It begins with the assumption that a new development is 10 percent more efficient than the best competitive unit, though it can be manufactured for the same cost. First, the market is estimated at 4665 units sold per year. Then, the savings from the higher efficiency of the new device are estimated at a present value of $14,200, using a discount rate of 12.5 percent over 20 years. Also, it is estimated that the current units yield an average profit of $6,200 per year. Thus, the price that a customer would be willing to pay to break even would be \$14,200 + \$6,200 = \$20,400. Finally, assuming the new development draws the entire market, the business opportunity would be 4665 × \$20,400 = \$95 million.

In a conversation with Dr. Collier, he stated that the business opportunity concept is still hampered by problems. The biggest of these is the fact that marketing has not been able to accurately estimate the markets where a new development might serve. Marketing apparently does not understand enough about the customer's needs. However, Collier maintains it is not the wrong approach, they just need to obtain better marketing data. He did note that this system is only used to judge incremental improvements, and the few breakthroughs they find at Borg-Warner are still measured in strictly qualitative terms. Thus, the system at Borg-Warner appears to follow our earlier outline.

There are several other limitations to the business opportunity concept. For instance, the method contains only economic considerations, while other qualities such as aesthetics or convenience are ignored. Therefore, this particular system is more feasible for the industrial market than for consumer products. In addition, Schainblatt points out that there is no relationship between the value of a business opportunity created by a new development and the difficulty of achieving that new development (*4*). Finally, Borg-Warner's system is not sensitive to the length of time spent on the project, so the results may balloon in some years when several projects "transist," and tailspin in others even though the R&D department has been just as productive.

4. *Program Value Method.*—A good example of a quantitative technique to evaluate R&D programs was reported by Schainblatt (*4*). The value of a program is based upon four different factors:

- Potential Annual Benefit. This is defined as the annual pretax income which will result from successful commercialization of the R&D program output. Financial benefits from R&D programs are estimated by their ability to be marketed as a new sales item, their applicability to be added on to existing items, or their ability to reduce costs. In each case, these financial benefits must be estimated separately.

- Probability of Commercialization. Management attempts to rate how well the R&D project fits with overall strategic plans and long term goals. A low fraction might be assigned if the technology has low interest level in business, while a high fraction would be attributed to a project that would be immediately fruitful.

- Competitive Technical Status. This probabilistic factor attempts to recognize the historical and scientific significance of the project. For example, a project that has had continued historic significance and is ahead of competitive activities would rate a high probability, but a project that is an alternative to more promising solutions would yield a low fraction.

- Comprehensiveness of the R&D Program. This factor discounts those projects that may only aim at part of the potential annual benefit. For example, a program which is targeted to a general area of opportunity and only has vague connection to the potential annual benefits is not very comprehensive. Thus, high-fraction comprehensiveness projects are perceived as having direct benefits for the entire problem.

The calculation of "program value" follows this algorithm:

1. Estimated annual new or projected sales for complete product

2. Estimated annual cost improvements

3. Potential annual benefit = (assumed percentage of line 1 that represents average incremental pretax income plus 100% of line 2)

4. Probability of commercialization

5. Competitive technical status

6. Comprehensiveness

7. Program Value = line 3 × 4 × 5 × 6

8. Total program value is summed over all businesses or products

9. A discount factor can be incorporated depending on the number of years to potential annual benefit.

The authors of the program value method know that the values generated are only rough estimates, and they only treat differences *in order of magnitude* as being significant enough to help with decisions regarding their R&D program mix. Thus, it does not seem relevant to use the program value to measure productivity. Rather, this method was presented to show how one company uses probabilities to help quantify the qualitative judgments of its managers.

5. *Other Quantitative Methods.*—Many companies have used the "bean counting" approach to quantitatively measure the productivity of their R&D personnel. This involves keeping records of patents, technical publications, or honors and awards from peer groups. For the high achieving R&D person, this might seem to make some sense. But so much of industrial research is carried out at a project level that counting patents or publications might be misleading. Also, much of today's research involves software, which, as of now, is very difficult to patent. In addition, some companies do not apply for patents because they feel they are better protected by keeping their research results a secret.

Furthermore, professionals are more prone to publish miniscule contributions if the number of publications is used as a measure of productivity. We interviewed a medical researcher who said that rather than publish three or four results together, he publishes each one separately so as to impress his funding sources. He has done research at three different organizations and he found this to be the standard method for reporting medical research. This researcher happens to do quality work, but he "plays the game" in order to have more publications.

Roland Mueser at Bell Labs suggested using *citations* of publications as an indicator of productivity. The fact that others had cited an R&D person's publication may show contribution to the field. However, this indicator could not be used exclusively since many researchers do not publish or may be working on proprietary matters.

Defining technical innovation as a new technical event like an invention, discovery, or theory that has proved to have practical utility, Mueser believes that technical innovation, unlike patents and publications, can be defined to gauge the results of all kinds of scientific and engineering work. Thus, counting technical innovations has been used at Bell Labs to measure fundamental research and product development (5).

Dundar Kocaoglu, professor of management at the University of Pittsburgh, suggested the development of *models* of the R&D department to help with the measurement of productivity. He is working on a way to combine control theory complete with feedback and delays to model the R&D function. The advantage of this approach is that a degree of understanding will result from studying the characteristics of R&D before it is modelled. Nevertheless, successfully modeling an R&D department appears to be a monumental task.

Another quantitative technique by Michael Packer of MIT suggests the use of a productivity information system that allows management to choose specific criteria that can be plotted for easy evaluation (6). This method appears to be quite promising because it stresses the need to present the productivity measurement in a useful format. The algorithm involves using a factor analysis to convert both objective criteria (like the number of patents) and intuitive indicators (like an undisputed reputation) into the underlying abstract concepts that managers use to evaluate R&D productivity (for example, "quality" vs. "quantity"). The factor analysis allows the output of R&D to be plotted according to these abstract concepts, and a trend analysis can be done to show the optimum positioning according to management's collective attitudes. The strength of this technique is the fact that management can easily adjust the plot with subjective inputs much easier with a plot than they can with hard numbers. However, this method could suffer in that it chooses specific criteria based on current objectives and assumes that they can apply over time. Since manager's attitudes and objectives tend to shift, it seems unfair to base R&D's productivity against criteria that could be unstable. This is a common problem in quantitative methods, though this method might still prove to be a valuable tool.

Five other methods to measure productivity as found by Schainblatt are summarized in Table 2. The location where every method is being used is not revealed because of promised anonymity. This list, when coupled with the examples described earlier, gives a good cross-section of the kinds of quantitative techniques now being used, even though it is not all-encompassing. Note that the examples we have presented range from the complicated algorithms presented before to the simple ratios in Table 2.

Table 2—Some Quantitative Methods Used in Industry Today

Methods	Disadvantages
Oil company counts outputs like flow sheets, cost estimates, and drawings; standards were developed, and complexity factors assigned. (this is engineering and not R&D).	Staff using it questioned the meaningfulness of the output; method obtrusive to professionals.
Comparative analysis and trends; science panel judges quality & impact of discoveries; use indices like number of analytical tests per professional to measure trends of developmental research.	How can different projects be rated on the same scale? Some projects may need more tests than others.
Figure of merit = (pre-tax profit over last 5 yrs)/(R&D expenditure); used in trend analysis for pay increases.	Assumes changes in sales are due to R&D expenditure, and that marketing, etc. has no effect on those sales.
Measure productivity increase due to investment in labor-saving equipment and instruments.	Assumes R&D just conducts tests, does not consider impact of improvements on business.
Quantitatively rate by patent attorney (for technical excellence) and by VP of R&D (for relevance to business); then check with (costs of R&D)/(goods produced) and (patents)/(professional).	Patents can be misleading, goods produced are dependent upon many other departments; VP ratings are semi-quantitative.

Nevertheless, each of these methods has problems that seem unavoidable. They are too rigidly defined to incorporate the broad spectrum of activity in an R&D department, and in some cases, such as the first in Table 2, would serve to irritate rather than motivate R&D professionals. It appears that the underlying problem of all quantitative methods is that they have attempted to quantify R&D using, in our opinion, poor models. Thus, it seems logical that research be directed in the area of improved model building so that quantitative measurements will be more accurate. A start for this modeling approach might be the use of the simple concept we outline in Table 1.

Semi-Quantitative Measuring Techniques

Semi-quantitative techniques appear to be among the best methods for evaluating R&D productivity. The subleties of different projects are not lost through the use of formal algorithms because the evaluation process is performed by assessors who are located near the project. Basically, this technique asks people close to the project to write down what they think, and it is these opinions that are then quantified according to different rating factors. These numbers may be crudely manipulated; for instance, by averaging them so that a condensed output may be provided. Thus, management has a number that they can readily use for comparison, as opposed to a purely qualitative statement.

But semi-quantitative techniques are not without their limitations. Often people can be swayed by bandwagon effects in making their qualitative judgments. Furthermore, using numbered scales can always be misleading since some people believe there can never be a perfect "10" while others think everything is perfect. Thus, relative differences in qualitative scaling can distort the output if these measures are incorrectly combined.

Discussion of several different examples and their problems will help to illustrate why we like this approach.

1. *Borg-Warner.*—The semi-quantitative technique used at Borg-Warner was described by Donald Collier (*2*). This system is used with the business opportunity concept described earlier. While the business opportunity concept analyzed the R&D efforts according to a specific algorithm, this semi-quantitative measure describes how well the research department met the agreed upon objectives according to qualitative judgments.

At the end of each year, the "customer," or other division within the company, compares the actual R&D performance with the initial objectives brought to R&D, and rates the effort on a scale from 0 to 3: "0" means that the objectives were badly missed; "1" means the project over ran time and cost; "2" indicates the objectives were met; and "3" is assigned to a project that exceeded the stated objectives or completed them well below the budgeted expenditure of time and money. This rating is then multiplied by the money spent on each project, and finally normalized by the total amount spent on R&D. Thus, a performance rating can be generated for each project for comparison and from year-to-year for historical trend analysis.

Borg-Warner uses these ratings to help determine annual bonuses. They believe this system has helped to improve the quality and clarity of the objectives for each project by encouraging better communication. The method has provided incentive to R&D staff not to overrun time and cost. The company believes productivity has improved as a result of using this productivity measurement.

However, this and many other measurement techniques are too strongly related to the estimates made prior to the project. Thus, the measurement is not of productivity, but rather of how well R&D can estimate its own abilities. Collier claims that his R&D workers are "perennial optimists", but it has been our experience that if someone is faced with making more money by meeting objectives, he tends to set his objectives lower. This is especially dangerous in an R&D environment where original thought and personally risky ventures should be encouraged rather than thwarted.

Semi-quantitative techniques appear to be among the best methods for evaluating R&D productivity.

Nevertheless, this example is typical of many semi-quantitative methods. Note that no formulas or obscure ratios are used, just the qualitative judgments of those nearest the work have been asked to assign numbers to their opinions.

2. *The Union Carbide Questionnaire.*— At Union Carbide, an annual "R&D Categorization Questionnaire" helps to evaluate the research efforts. This questionnaire, as described by Whelan, is the result of several years of work to establish definitions of different kinds of projects that are understood and accepted by both management and R&D personnel (*7*). The intent was to avoid good-bad connotations so that each active R&D project could be categorized fairly. Ratings are provided annually by line managers in charge of typically $1–$5 million in R&D funding.

Union Carbide has found it useful to convert the responses to each question to dollar average rating. The average response is weighted by the project funding. Since responses generally fall within a continuum from a defensive to an offensive posture, it is helpful to define this common scale as one in which zero represents defensive and one represents offensive. These defensive to offensive (D/O) averages are reported with the detailed data. Further condensation has proven helpful by combining selected D/O averages into composite indices. For example, Union Carbide numerically analyzes historical trends by combining the D/O averages for the sections on the questionnaire that judge the purpose, stage, type, and organizational implication for a project as well as the two ratings of success probability.

This system has several advantages. First, it is possible to ascertain the relative amounts of corporate R&D effort being devoted to *high risk versus low risk* projects. In addition, the profile serves as a *communication tool* between R&D and corporate management. Also, *trend analysis* of R&D posture reflects the R&D response to corporate goals as well as the impact of budget variations on R&D objectives. Finally, this method incorporates analysis of *all types of research*, from basic research to product improvement.

However, this measure does not necessarily measure productivity, but rather converts the impressions of line managers to numbers as they perceive how their workers are performing. Experience has shown that quite often line managers can distort the truth to impress their superiors or can be misled by their subordinates. In addition, line managers in R&D sometimes lack the qualifications necessary to rate each project. Thus, this method is limited because the sample may not be broad enough.

3. *The Peer Rating Approach.*—Quite simply stated, a peer rating system merely asks all project members to evaluate themselves and one another on some quantitative scale. A supervisor then correlates the data and condenses it to some grading factor. The system has worked quite well in evaluating the participation of students in industrial-funded projects at Harvey Mudd College, and it might be applicable to an R&D environment in several ways.

Obviously, R&D professionals could be asked to rate each other. But they could also be polled as to their opinions about the type of tasks they think they should be doing and how well they perceive other projects are doing. Naturally, it is expected that this kind of information would be compared with a supervisor's own appraisal. Thus, this is not just a personal evaluation, but a project rating as well, which could be used to measure the productivity of an entire division.

The advantages of a peer rating system are numerous. First, the subtle differences between people and projects are certainly highlighted by peer ratings. It seems that true breakthroughs might surface faster as well, since peers can recognize the value of an idea before it is reduced to the layman's terms of top management. This would also occur without ignoring the benefits of significant incremental improvements. Thus, it seems that many of the problems of quantitative techniques can be avoided using peer ratings.

Another important benefit of the peer rating approach is the added understanding management can obtain about their own employees. In R&D departments, the seemingly unproductive, purely creative individuals sometimes make the most fruitful discoveries. Sometimes that "unproductive" person is a "gatekeeper"—the kind of person who always answers everyone else's questions, but never seems to have time for his own work. These types of individuals might be weeded out in a strictly quantitative method. However, a peer rating system will alert management to these extremely valuable R&D professionals.

Peer ratings will also help in managing R&D. It is management's responsibility to formulate the right mix of talents on a research project. If management knows more about the strengths and weaknesses of each professional, it stands to reason that a better mix could be achieved. For example, one of the authors recalls that the Jet Propulsion Laboratory could have used this kind of information when a new supervisor was hired for a group of three world-class mathematicians whose only drawback was their inability to communicate. The previous supervisor had extremely good communication skills, so the mix was very good. But the new supervisor was not as effective because his skills in communication were not as good as his predecessor, and the talents of these mathematicians were not fully

Many of the problems of quantitative techniques can be avoided using peer ratings.

used. A peer rating system would have encouraged management to locate a communicator to work with the mathematicians, or improve the communication skills of the new group supervisor.

A peer rating approach is not without its problems. In R&D, there are occasionally those individuals that work on their own project by themselves. There are also individuals who work on many projects at once. Thus, there is an immediate problem to assess just who is to evaluate whom so that a valid evaluation may be generated. In addition, individuals in one project may have limited or no knowledge of certain other projects because of secrecy requirements. This means that the secret projects will have a more limited sample of evaluations.

Often it is claimed that people cannot accurately rate the performance of their peers. But a study done at the Air Force R&D Laboratories shows that professional colleagues are suited to evaluate the innovativeness and productivity of researchers' output (*8*). The experiment asked 2, 3, 4, or 5 people to rate their peers on a scale from 1 to 9. After applying various mathematical tests to their data, they concluded that peer ratings of R&D people are both *reliable and valid*. They also found that those individuals who were classified by their peers as innovative were generally productive as well.

Another problem with peer ratings is the aggrandizement effect. This is the hypothesis that a rater will almost always rate his own ability or output higher than others would have rated them. For example, it is mathematically implausible to have 40 percent of all mathematics departments in the top 5 percent, but if they are asked individually, they will all certainly claim that they are. A study was conducted to test for the aggrandizement effect in 55 sets with six organizations per set. The organizations ranged from Camp Fire girls to insurance associations. It was found that the raters overestimated the prestige of their organization eight times as frequently as they underestimated it, and net overestimation could be discerned in every one of the sets (*9*). Our experience in using peer ratings at Harvey Mudd College to evaluate students substantiates this study. But we have found that aggrandizement can be minimized by eliminating an evaluator's rating of himself.

A peer rating system as applied to an entire R&D department could be disaster if it were not handled correctly. Careful consideration should be given as to just how the peer ratings are administered and how the results are presented. Promoting the feeling that everyone is evaluating you at all times might stifle communication that is vital to the success of an R&D organization.

Laurie Larwood, professor of psychology at Claremont McKenna College, suggests the use of the Delphi Method to develop an effective peer rating system. The Delphi method has been used for years to determine the best strategy for a business to pursue by interviewing prominent people in the field. The results are then combined and reviewed by top management, and a new set of questions for the experts is filtered through. This process is reiterated until some focus has been determined. Presumably, this technique eliminates the "band wagon" effect of a brain storming session. Prof. Larwood believes this technique could be used to help find out what criteria the R&D personnel would like to be rated on, and how. In addition, the questionnaire could be formulated so as to include those aspects that professionals feel are important to the well-being of a project and an entire R&D organization. It is hoped that a valid questionnaire would result from this stage, one in which the majority of professionals would find effective.

Overall, the peer rating approach seems to have met with general approval from several of the experts we interviewed in the course of this study. According to Kocaoglu, an R&D department at Westinghouse Electric is happy with their peer rating system which involves having each of the 16 professionals rate the strengths and weaknesses of each other, the organization, and the mix of R&D projects that were adopted.

Peer ratings appear to be better than most other semi-quantitative techniques. Most other rating systems use only the supervisors of R&D, or some outside specialists in the field. In fact, one of the authors worked for a large company that ranked each employee based upon ratings of their supervisors, not by peers. We believe both systems should be used and then compare the results. Differences between the two should raise important subtle questions about how R&D is run and where it is going.

Semi-quantitative techniques seem to attack the problem of R&D productivity measurement better than quantitative techniques. They deal with the measurement on a more flexible level because the R&D function itself is so flexible. However, given time and research into new techniques of measurement, this current trend could change.

In sum, a great deal of research needs to be done to discover better ways to measure the productivity of R&D. Since so many different intangible factors come into play, perhaps an integrated group of individuals might be able to make significant contributions. A team consisting of scientists, psychologists, economists, engineers, and management scientists would be necessary to formulate a better model to evaluate the R&D process. For instance, the Claremont Colleges, and Harvey Mudd College in particular, have all of these elements represented in a contiguous area so that the work could proceed smoothly. We are interested in pursuing such an endeavor. As this kind of research progresses, measuring R&D productivity may become a valuable tool to guide and motivate R&D organizations. □

References

1. Foster, Richard N., "Boosting the Payoff from R&D," *Research Management*, January, 1982.
2. Collier, Donald W., "Measuring the Performance of R&D Departments," *Research Management*, March, 1977.
3. Gee, Robert E., "The Opportunity Criterion—A New Approach to the Evaluation of R&D," *Research Management*, May, 1982.
4. Schainblatt, Alfred H., "Measuring the Productivity of Scientists and Engineers in R&D; A State of the Practice Review," The Urban Institute, Washington, D.C., May, 1981.
5. 1980 IEEE Engineering Management Conference Record, "What is Technical Innovation?" by Roland Mueser, Catalog #CH1603-0.
6. Packer, Michael B., "Analyzing Productivity in R&D Organizations," *Research Management*, Jan.–Feb., 1983.
7. Whelan, J. M., "Project Profile Reports Measure R&D Effectiveness," *Research Management*, September, 1976.
8. Stahl, Michael J. and Joseph Steger, "Measuring Innovation and Productivity—A Peer Rating Approach," *Research Management*, January, 1977.
9. Caplow, Theodore and Reece J. McGee, *The Academic Marketplace*, Doubleday & Company, Inc., Garden City, New York, 1965, pp. 88–90.

IMPROVING THE RETURN ON RESEARCH AND DEVELOPMENT

A framework that allows R&D practitioners to estimate the potential for improving the performance of industrial R&D organizations is provided in this study by a subcommittee of the IRI's Research-on-Research Committee. More than 40 operational suggestions for improving performance are also presented. This report appeared in condensed form in the January-February and March-April issues of RESEARCH MANAGEMENT. The complete 40-page report is available from the Industrial Research Institute, 100 Park Ave., Suite 3600, New York, NY 10017. Single copies $10; 10 percent discount on 10–99 copies and 25 percent discount on 100 or more.