

Design principles for the development of measurement systems for research and development processes

Inge C. Kerssens-van Drongelen¹ and Andrew Cook²

¹ Faculty of Technology and Management, University of Twente, The Netherlands

² CSC Consulting and Systems Integration, Preston, PR1 1RE, UK

Based on a comprehensive literature review and the activities of numerous case study companies, it is argued in this paper that performance measurement in R&D is a fundamental aspect to quality in R&D and to overall business performance. However, it is apparent from the case companies that many companies still struggle with the issue of R&D performance measurement. Excuses for not measuring are easily found, but there are also empirical examples and literature available with suggestions how it can be done. In this article this literature is reviewed and placed within the context of general performance control and contingency theory. Furthermore, the main measurement system design parameters are discussed and some basic system requirements are described as well as several design principles that can be useful for those who accept the challenge of establishing a meaningful measurement system.

Introduction

Research and Development (R&D)¹ was once considered to be a unique, creative and unstructured process that was difficult, if not

¹ In this paper, we will use the abbreviation R&D as a general term to indicate activities varying from basic research, to product development and introduction. If at any point basic research or any other specific R&D category is intended then it will be clearly distinguished in that context.

impossible, to manage and control. The standard management and control techniques used in other parts of the organization were therefore considered inappropriate for R&D (Roussel et al., 1991). However, recent changes in the business environment — intensified competition, splintered mass markets, shortened product life cycles, and advanced technology and automation, etc. — have focused senior managers' attention on R&D's contribution to competitive advantage (Kumpe and Bolwijn, 1994; Wheelwright and Clark, 1992). Today, R&D is no longer considered to have a mere supportive role to the primary business processes, but to be a vital part of it (de Weerd-Nederhof et al., 1994). These changes challenge companies to improve their R&D processes in terms of efficiency, internal and external customer focus, time to market and innovativeness (Kumpe and Bolwijn, 1994, Wheelwright and Clark, 1992). Although managers still acknowledge that R&D processes have several characteristics that make them different from other business processes, they no longer accept that this should mean they are unmanageable. Several companies have now adopted management techniques traditionally used in other parts of the company, like Total Quality Management and activity control, in R&D (Miller, 1995). One of the key aspects of quality management is the need for performance measurement (Miller, 1995; Schumann et al., 1995). Measurement drives behaviour and, even more importantly, behaviour change. Furthermore, it supports the prioritization of actions and enables comparing and tracking of performance changes and differences

(Schumann et al., 1995). If TQM is to be used effectively in R&D, a measurement system needs to be in place to give accurate and timely data on R&D performance. Furthermore, feeding back this performance information should motivate the researchers to excel and improve. This approach to measuring R&D performance contrasts with the traditional approach, which is often limited to annual budgetary and informal professional control (Roussel et al., 1991). Thus, there is a clear need to develop new measurement systems that meet the requirements of modern R&D management and accommodate the peculiarities of R&D processes. This article focuses on the issues involved in the development and content of such measurement systems.

The article starts with an introduction of the research method, followed by a brief overview of the theory of performance control and performance measurement. We continue with a discussion of the practical problems encountered when trying to apply these general control principles to R&D processes. Taking these problems into account, two critical issues that need to be considered to reduce the risk of failure when designing a measurement system for R&D — the purpose of the measurement and contingency factors — are detailed. Although we argue that the specifications of the performance measurement system should be developed by the users themselves to meet their particular needs, we will give an overview of measurement system requirements and design principles that can be helpful to this development process.

Research method

This article is based on a review of the R&D performance measurement literature. It appeared that many authors have focused on one aspect of R&D performance measurement systems, mostly the metrics or measurement methods that were or should be used to measure R&D performance. Often, the positioning of the metrics in a more general framework of control theory and a classification of metrics and methods based on application areas was lacking. To fill this gap, we have placed R&D performance

measurement in the context of general control theory and will present the findings from literature in such a way that they become useable for a manager who is responsible for developing an R&D measurement system.

The literature review is illustrated with empirical evidence and examples gathered in two separate research projects. One project was carried out in the Netherlands, whereas the other investigated British, Dutch, Swedish and Japanese companies. The Dutch project focused on different aspects of R&D performance measurement procedures (metrics, methods, norms, frequency and timing of measurement), the context in which they were applied (amongst others characterized by the type of R&D and the hierarchical level at which was measured), and the perceived effectiveness of the procedures. It consisted of a survey which was returned by respondents from 48 large and medium-sized companies, and nine in-depth interviews with R&D managers. A detailed discussion of the research method and the findings is presented elsewhere (Kerssens-van Drongelen and Bilderbeek, 1996). The other research project consisted of semi-structured interviews with managers and engineers involved in R&D in a variety of industries. Although the research addressed new product introduction performance in different operating scenarios, performance measurement issues were also investigated.

The role of performance measurement in performance control

In the broadest sense, control can be defined as: 'any form of goal-directed influence' (de Leeuw, 1982). Thus, performance control of organizations and processes can be seen as ensuring that the combined efforts of the people involved, using multiple resources, are in line with company objectives and plans. In the literature, two basic approaches to performance control can be identified (Ansari, 1977; Bruggink, 1989). First, there is the structural or *measurement approach* adopted in the literature about cybernetics, accounting and management information systems. Here, control is seen as decision making on the basis of feed forward and feedback of (quantitative) information about

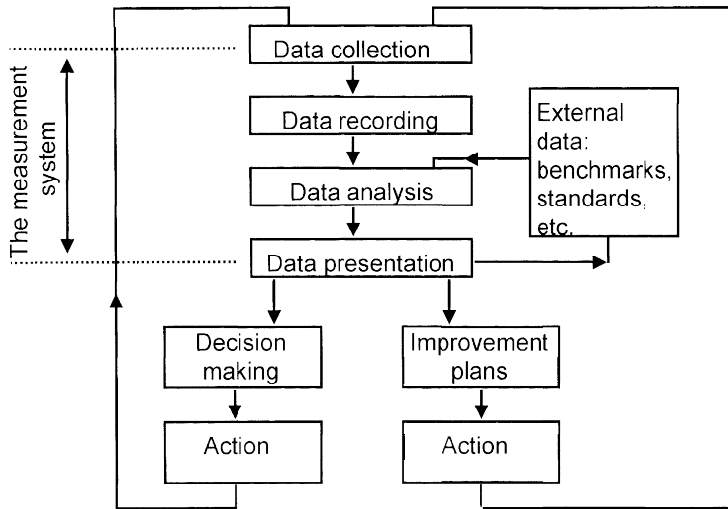


Figure 1. The measurement and control process.

actual performance and influential external factors. Thus, *measurement* is at the core of this approach. The other part of the control literature is dominated by the *behavioural approach*. Behaviourists see control as a set of formal and informal mechanisms for coordination, supervision and motivation of people that are applied to direct their behaviour towards organizational goals. At first glance, performance *measurement* does not seem to have a clear role in this approach. The decision which control mechanism to apply in a specific situation could be based on general views, assumptions or personal preferences. However, it seems more reliable to base this decision on clear information about the current and expected status of organizational goal attainment and about factors that might influence goal attainment. In that case, the behavioural approach needs to be supplemented with measurement. Combining the behavioural and the measurement perspectives in this way, performance control can be defined as a process consisting of the following activities:

the *acquisition and analysis of information* and the *interpretation* of this information to *determine what to do and how to do it* and the *application of the chosen measures* to influence people so that their efforts are aligned to company objectives and plans.

In accordance with this definition, *performance measurement* can be interpreted as a specific part of the control process, namely: the *acquisition and analysis of information* about the actual attainment of company objectives and plans and about factors that may influence plan realization. Finally, a performance measurement *system* can be defined as a set of tools and procedures supporting the measurement process. It is the mechanism by which the performance information is *gathered, recorded and processed* (see Figure 1). In the next paragraph, we will discuss the problems encountered when trying to set up such a system in R&D.

Problems with performance control in R&D

Several problems with designing and implementing R&D measurement systems have been reported in the literature (Kerssens-van Dronghen, 1994). The most prominent of these problems are listed and some approaches to address them are described below.

The problems

Obviously, it is desirable to measure R&D performance in terms of the standard

financial performance measures applied to other business functions — profit, return on investment, return on net assets, etc. However, when this is attempted two main structural problems are encountered. The first problem is the difficulty of isolating from other business activities the contribution of R&D to company performance; it is also problematic, with traditional accounting methods, to identify the contribution of R&D to the profits resulting from individual new products. The second problem is the time lag between R&D efforts and the potential financial rewards, which makes it difficult to use this information for timely decision making. This is especially applicable to basic research, but can also apply to applied research or development projects. For example, companies designing automotive components for vehicle manufacturers sometimes have to wait for up to two years for the production start and the first product sale. At Bell Labs there is a typical time lag of between 7 to 19 years for basic research projects (Pappas and Remer, 1985). In addition, it can often not be foreseen in which products or processes basic and applied research output may be used. For example, several years ago a venture capital company that participated in our empirical research invested heavily in developing core pen computing technology that is still being incorporated in today's new products. Over a 5-year period they have launched more than 20 products based on the core technology and more are planned. Nobody could have reliably predicted this 5 years ago.

Related to the problems of selecting performance metrics is the determination of the correct *norms* for comparisons. Some interviewees in our empirical research indicated that they found this an even more difficult issue than the selection of metrics. By their very nature R&D projects are unique and involve non-repetitive processes. It is consequently difficult to compare and contrast two projects as they are always different, and researchers tend to object to targets based on past projects, or use project uniqueness as an excuse for not meeting those targets. However, it should be noted that project uniqueness is more applicable to (basic) research projects than to product development projects; usually, the activities of

development projects can largely be defined and scheduled in advance. Setting norms for metrics at a more aggregated level (e.g. the performance of the R&D department in general) can be problematic as well. Many companies lack records of past performance from which trends and norms can be derived, and relevant public data on these subjects are also hard to obtain.

The last problem we wish to discuss here is the *acceptance* of performance measurement in R&D. Many engineers and scientists believe the design and implementation of such a system is counterproductive since the very act of measurement is thought to discourage creativity and to reduce motivation among highly educated technical people (Pappas and Remer, 1985). Brown and Svenson (1988) give two possible explanations for this negative attitude towards measuring R&D. The first explanation is that engineers and scientists may fear such systems because they may highlight their inadequacies and lack of productivity. The second explanation, which they regard as more significant, is that they have bad experiences with inappropriate measurement systems leading to improper decision making, and consequently believe that all measurement systems do not work.

How to deal with the problems

However, the problems mentioned above should not be used as excuses for not measuring R&D performance at all. As the vice presidents responsible for R&D at General Electric and at Alcoa have argued: a credible R&D organization needs to prove that it is working on the right things in the right way, with as much quantifiable arguments as possible (Bridenbaugh, 1992; Robb, 1991). Furthermore, performance measurement is at the heart of any (R&D) quality management system (Francis, 1992; Miller, 1995; Schumann et al., 1995). Thus, measuring R&D performance is necessary both for external and internal reasons. The challenge is to overcome the problems discussed and to find solutions that can cope with the peculiarities of R&D processes and the attitude of the people working there.

For example, the behavioural problems of R&D measurement acceptance might be

reduced by having the R&D employees involved in the design of the system (Brownell, 1985). In organizations with empowered (multifunctional) teams, employees might even ask themselves for an appropriate measurement system that supports them in their own decision making (Meyer, 1994).

The problem of norm setting should be addressed first of all by building up a structured project data system and getting started with measuring performance. Many companies we visited in our empirical research still lacked structured project data systems, making it difficult to learn from the past. Once measuring and recording have been started, records of past activities and performance will become available that can be used to forecast new project performance more accurately and to develop realistic norms (Patterson, 1983). To ensure that the norms will not be merely inward looking and not competitive in the marketplace, managers should also actively engage in benchmarking initiatives as currently offered by consultancy firms or should themselves try to find benchmarking partners.

To overcome the success definition problem, many R&D managers and academic researchers have looked at alternatives for success measurement, e.g. process or direct output measurement (see for example Brown and Gobeli, 1992; Griffin and Page, 1993; Kerssens-van Drongelen and Bilderbeek, 1996; Moser, 1985; Packer, 1983; Pappas and Remer, 1985; Patterson, 1983). Unfortunately, this theoretical body of knowledge is rather fragmented. Different measures and measurement methods have been proposed, but a general framework that defines what concepts to use in what context (e.g. purpose of the measurement, organizational level, type of research) is lacking. Developers of company-specific R&D measurement systems should take care adopting one of these approaches without in-depth situation analysis, as they may run a serious risk of failure. Instead, they should first address two crucial issues: what is the specific purpose of the measurement procedure (i.e. incentive or diagnostic purposes) and what are the contingencies? Only after these issues have been addressed properly, system developers should turn to metrics, measurement techniques, frequency, etc. matching with their specific situation

(Kerssens-van Drongelen, 1994). If different purposes or contingencies are at stake at the same time, the last step should be to align the different measurement procedures with one another, resulting in an efficient performance measurement *system*. In the next section of the paper we will discuss the crucial issues of measurement purpose and contingency factors in more detail.

Purposes of measurement

There are effectively two clusters of purposes for performances measurement; each requiring its own approach to measuring. First, performance measurement can serve the purpose of *motivating people*. The assumption is here that by feeding back information about their performance, possibly coupled with incentives, people will be motivated to change their behaviour. The second group of purposes are associated with *diagnosing* activities (e.g. projects) and organizational units. For example, a diagnostic approach can be used to assess if problems can be expected, or if organizational changes have had an effect. Measurement systems for these purposes include, for example, project progress-monitoring systems, post project evaluations, and organizational audits. A major difference between motivational and diagnostic purposes is that the former needs to include only those factors that *can be controlled by the (group of) person(s)* subjected to the measurement, but the latter should preferably assess the *combined effects* of personnel, technology and environment (Pritchard, 1990). Furthermore, if for motivational purposes an incentive or reward programme is coupled with the measurement (e.g. bonuses, promotions), one should ensure that all the important influential aspects of work are measured. This is because there is a human tendency to direct more energy towards tasks that are being measured, giving less attention to tasks which are not. The diagnostic purposes can frequently be satisfied with incomplete performance measures (Pritchard, 1990).

The purpose or purposes of the measurement system determines the 'how, what, when and where' of the data collection, analysis and reporting. Thus, the purpose

must be defined before any attempt is made to design the measurement procedure. Clearly, the data collection, analysis and reporting methods ideal for one purpose will not necessarily suit the other. At Philips Central Research, researchers involved in the development of a self-audit tool for project teams were very suspicious about what would be done with their audit data. They said they would not use the tool if the data would also be used by higher level management without any additional consultation. They felt that, although the data would suit their own purpose very well, it could easily lead to misunderstandings and improper decision making by others (Deetman, 1994). But whenever possible, one should obviously try to couple the data collection step of different measurement procedures into one measurement system to minimize duplication of effort and reduce the administrative burden. The collected data can then be analysed and reported according to the different individuals' needs. Spreadsheet packages or relational database management systems seem useful tools for this.

Contingency factors for measurement procedures

The contingency model of organizations suggests that organization design should be in harmony with the contingency factors in the environment. As argued by the contingency theory of management accounting, these external factors, as well as specific internal factors, also have an impact on the design of measurement systems (Emmanuel et al., 1990). External influential factors include, for example, market predictability, the nature of competition and the number of different product-markets faced. Relevant internal features include the size of the organization, organization structure and the interdependence between units, the degree of decentralization, the organization's strategy, the control policy and the nature of the primary business process (Emmanuel et al., 1990; Snellenberg, 1995). For a specific measurement procedure, which could be part of a larger measurement system, the aggregation level at which is measured can also be regarded as an influential factor.

In the R&D performance measurement literature as well as in our own empirical research, the influence of several of the general contingency factors on R&D measurement system design has been confirmed. For example, Loch et al. (1996) found evidence that the nature of competition, especially the degree in which technical competence is a competitive factor, has an impact on the choice of R&D process performance measures. Furthermore, Griffin and Page (1996) have confirmed the influence of strategy on R&D metrics selection. They found that the most appropriate programme-level and project-level success measures depend on the company's innovation strategy and the project strategy respectively.

A specific instance of the contingency factor 'nature of the primary business process' in an R&D environment could be the type of R&D. R&D activities can be classified in a number of ways. The most used classification is the one by the OECD, which identifies three categories: basic research, applied research and (product and process) development. Product development efforts can be further subdivided into runners, repeaters and strangers (Cook et al., 1996). A runner is a relatively trivial product change that involves minimal effort. A repeater is a product that the organization is familiar with and a stranger is a radically different product for the organization and consequently requires more effort and is more risky. Wheelwright and Clark (1992) use a similar classification, but labeled them breakthrough, platform and derivative projects. Several authors have suggested that the type of R&D influences one or more measurement system design parameters such as the metrics, measurement techniques, norms and frequency of measurement (Brown and Gobeli, 1992; Pappas and Remer, 1985; Quinn, 1960). In basic research projects it is often not known when or for which products or processes the outputs will eventually be used, if at all. In such a situation, criteria concerning market opportunities, economic benefits, etc. are usually not obtainable or relevant. The measurement technique proposed for this type of research is 'qualitative evaluation' (e.g. peer reviews). Conversely, for development projects economic and market outcome measures are relevant and important, which can and should be measured

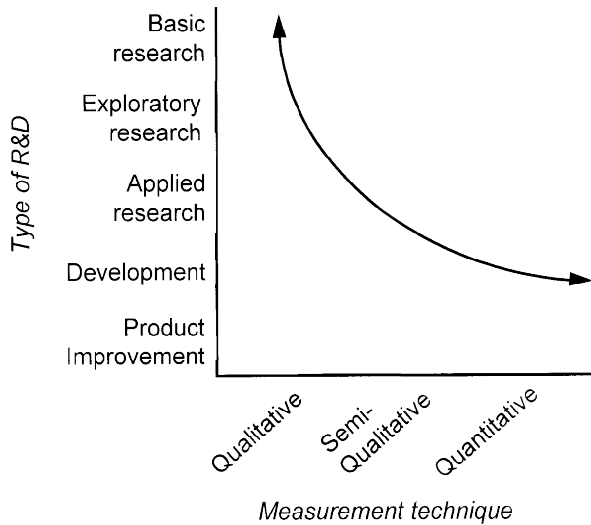


Figure 2. The relevance of different measurement techniques (based on Pappas and Remer, 1985).

quantitatively. Figure 2 shows a classification of measurement techniques that Pappas and Remer (1985) advocate. In spite of the fact that these propositions seem evident, they are not proven with substantial empirical data. While the differences between the two extremes, basic research and product development/improvement, may instinctively be accepted, it becomes more problematic with the R&D types in between. In our Dutch empirical research for instance, we found no difference between the measurement procedures for applied research and for product development/improvement. The interviewees explained that they did not have different departments for these types of R&D. The same researchers worked on both types of R&D and in some organizations these types of R&D were also combined within one project. Thus, it was not feasible to have separate measurement procedures for applied research and development.

In our Dutch empirical research project, the contingency factor 'aggregation level at which is measured' appeared to have a profound influence on the design of measurement procedures. The metrics, 'norm setting methods', 'frequency and timing of measurement' and 'measurement techniques' were all found to be quite different in kind and/or in frequency of occurrence, depending on the level at which

was measured. For example, we found that at the individual level subjective assessment by superiors was the most applied measurement technique, measuring metrics such as working speed, accuracy, ability to work in teams and achievement of milestones. On the other hand, at the organizational level the most used measures were: number of projects completed, customer satisfaction, percentage of sales due to new products and time to market. The measurement techniques most frequently mentioned for this level were quantitative measurement and satisfaction questionnaires/verbal feedback from customers (Kerssens-van Drongelen and Bilderbeek, 1996).

The performance measurement system

In this last section of the paper, we will deal with the actual design of a performance measurement system. First, the basic system requirements will be presented. Then the major system design parameters will be discussed.

Measurement system requirements

Any measurement system must satisfy some basic criteria. It should allow the right information, at the right time, to be collected

reliably and economically. Many measurement systems we have seen do not do this. Some systems have developed in an uncontrolled fashion over a number of years. The result is an unwieldy measurement system that does not fit the business structure or activities. In other cases the processes and the organization have changed significantly, but the measurement system has remained the same. In both situations the system creates an unnecessary overhead burden that collects information that may be irrelevant to business needs.

A measurement system should be designed using a holistic approach. All relevant factors should be taken into consideration — the cost of running the system, the timing of the reporting, both the normal and *ad hoc* information requirements of the decision makers, available benchmarking criteria, etc. Information technology can play a useful role in reducing the data collection and analysis overhead, increasing the frequency of reporting and expanding the set of possible metrics, thus giving the ability to have *ad hoc* reports as and when required.

Other writers support these views. Packer (1983) identified five criteria that the measurement system should achieve. The information must be understandable and easy to interpret, it must be relevant, and it must be reliable (that is representative, unbiased and verifiable by others). Furthermore, the system must be acceptable and cost effective.

The sophistication and motivation of the stakeholders involved (the people who are subject to the evaluators and the users of the outcomes) will also impose requirements; in most cases any one group of people might have more than one role. If, for example, the people subject to evaluation have a positive attitude towards performance measurement, they will be more inclined to co-operate in a quantitative assessment, but if they are resistant to evaluation, qualitative assessment might be the only option. The requirements of the users will differ with the organizational level. The higher the level, the more aggregated and the more quantitative (financial) the metrics tend to be.

Major design parameters

Measurement system design parameters can be defined as the topics that need to be

addressed when developing a measurement system. The main parameters are: metrics (performance measures), measurement system structure, standards against which to measure performance, measurement techniques, reporting formats and the frequency and timing of measurement and reporting.

Metrics. As stated previously, the measures of performance (metrics) must align with the purpose of the measurement and the contingency factors. Furthermore, they must reflect the *objectives for and responsibilities of the person(s) or activities that are being measured.* It is widely recognised that the measures of performance used, have a strong influence on business activities and results (Kaplan and Norton, 1992). The old adage 'you get what you measure' holds also true in R&D. It would be fundamentally wrong to measure laboratories that have objectives that include identifying future technologies and undertaking basic research that may first be incorporated in products in 20 years time on traditional short term oriented criteria such as payback or return on investment. If a basic research laboratory is measured on such criteria it would undoubtedly change over time and turn towards work, such as incremental product development, that has an immediate return.

Many different metrics are used in practice. In fact, there is an unlimited number of criteria that can be used as measure of performance. All the metrics that have been identified in the literature and in the case study companies can be categorized as operationalizations of one or more of the five *top level measures* — *cost, quality, time, innovativeness and contribution to profit.* As argued by Bolwijn and Kumpe (1994), efficiency (cost), quality, flexibility and innovativeness reflect the critical responses of a firm to the increasing demands of its customers. ('Flexibility' in R&D is translated into time issues like throughput time and timeliness). Performance on these four aspects should eventually result in performance on the fifth metric: *contribution to profit.* The five top level metrics also align with the four perspectives mentioned by Kaplan and Norton in their 'balanced scorecard' (1992, 1993, 1996): 'quality' corresponds with the 'customer perspective'; 'efficiency' and 'timeliness' with

the 'internal business process perspective'; 'innovativeness' with 'innovation and learning'; 'contribution to profit' with 'financial perspective'. With measures from the first three perspectives, R&D managers (or empowered teams) can diagnose whether individual researchers, teams, departments, etc. are currently focusing and performing well on those aspects of their work that are assumed to be critical to business success. In addition, measures from the financial perspective help to analyse — be it with a time lag — whether the R&D strategy itself was right.

Although the application of all top five measures have been found in practice, most companies often only use a limited sub-set. In a study of 124 R&D managers Moser (1985) found that from a given list of metrics, the ones mentioned in Table 1 were used most frequently and consistently in practice. In Table 1 we have indicated to which top measures we think the metrics reported by Moser are linked. Innovativeness measures (e.g. number of patents, numbers of professional rewards) were also used, but less frequently. Contribution to profit measures were not included in the survey and no distinction was made between the different types of R&D or organization levels. It can be observed that 'soft' measures are the ones most commonly used, even though they are difficult to calculate accurately and reliably in practice. This finding was confirmed in our empirical research.

Griffin and Page (1993) identified a somewhat different list of metrics for diagnosing product development success and failure.

They made a distinction in firm level measures, programme level measures and project level measures (subdivided in development aspects (cost/time), financial aspects (contribution to profit indicators) and customer acceptance (quality)). The ones most often mentioned are presented in Table 2. Again, the innovativeness measures were lacking. The contribution to profit measures were found to be used quite often. In fact, they found that managers use on average about four criteria, mostly from the quality and contribution to profit categories.

Measurement system structure. As more than one metric is often and should be used, the problem is posed if and how these measures should be systemized in some kind of a framework. This systemization should help to check whether all the important aspects of performance are measured, and to elucidate compensation effects between measures. In our case study companies, we rarely observed such systematic approaches.

In business economics, a pyramid structure is usually preferred for a metrics framework. Such a framework has one aggregated primary measure on top (e.g. return on R&D investment), from which several elucidating and supporting variables are derived (Bruggink, 1989). The lower level metrics could serve performance measurement and decision making at the lower levels of the organization, whereas the aggregated primary measure is used by top management. Although attempts to apply this approach in R&D have been reported (Foster et al., 1985), this mathematical deduction

Table 1. Measures of performance used in practice (adapted from Moser, 1985).

Actual measure (as defined by Moser) — in order of descending use	Top level measure		
	Quality	Cost	Time
Quality of output or performance	✓		
Unit's degree of goal attainment	sometimes	sometimes	sometimes
Amount of work done on time	sometimes		✓
Unit's level of efficiency		✓	
Percentage of project completions		sometimes	sometimes
Percentage of results adopted by the company	✓		

Table 2. Five categories of metrics for product development (adopted from Griffin and Page, 1993).

Firm-level	Programme-level	Project-level		
		Development aspects	Financial aspects	Customer acceptance
% of sales by new products	Programme met new product objectives	Development cost	Break-even time	Customer acceptance
Strategic fit with business	Impact of new product programme on corporate performance	Launched on time	Attain market goals	Customer satisfaction
		Product performance	Attain profitability goals	Met revenue goals
		Met quality targets	IRR/ROI	Revenue growth
		Speed to market		Met market share goals Met unit sales goals

method is not widely used in R&D because of the problems mentioned earlier (acceptance, time lag, etc.).

Another systemization concept which we think could be more suitable for R&D performance measurement, is the balanced scorecard we already mentioned in the previous section (Kaplan and Norton, 1992, 1993, 1996). In this approach, strategic objectives are formulated looking from four perspectives (financial, customer, internal process and innovation and learning perspective). In Figure 3, an example is given of a balanced scorecard for R&D. For each objective, one or more metrics can be formulated. Some of the objectives can easily be translated into a metric, e.g. 'current development time' divided by 'reference development time' reflects performance on 'time to market improvement'. But for other objectives, a construct has to be chosen which is *assumed* to reflect the performance on the objective (e.g. ranking by key accounts as a metric for customer satisfaction). In fact the objectives themselves are also *assumed* to have a causal relationship with better financial performance. This is also recognized by Kaplan and Norton (1996), who present it as an instrument of *strategic learning* about the validity of the strategy and the quality of its execution. For motivational purposes, the top level balanced scorecard might be translated into units' or even individual scorecards. This could be done through translation of the strategic objectives into objectives that can

be influenced by the people at that organizational level and formulating corresponding performance measures.

Standards to measure performance against. Performance measures are meaningless unless they are compared against generally accepted reference points (Kerssens-van Drongelen, 1994). Industry standards, self-established targets (for example budget or schedule adherence) or externally generated benchmarks from competitors or companies outside the industry can all be used. In our empirical research among Dutch companies involved in R&D, we found that the procedures for norm setting depended largely on the organization level at which was measured. Further, it must be noted that if performance is measured against self-defined goals then these goals must be meaningful. For example, the owner-manager of an automotive company that participated in our empirical research in the UK, imposed time schedules that the staff found to be unrealistic. This consequently created a fire fighting style when projects invariably slipped from target.

Measurement techniques. A metric may be measured *qualitatively* or *quantitatively*, or a *hybrid* of the two methods (Pappas and Remer, 1985). Qualitative methods inherently result in subjective and intuitive measures. This method is often used for individual performance appraisal and stage gate project assessment. An expert, normally a line manager, evaluates the performance

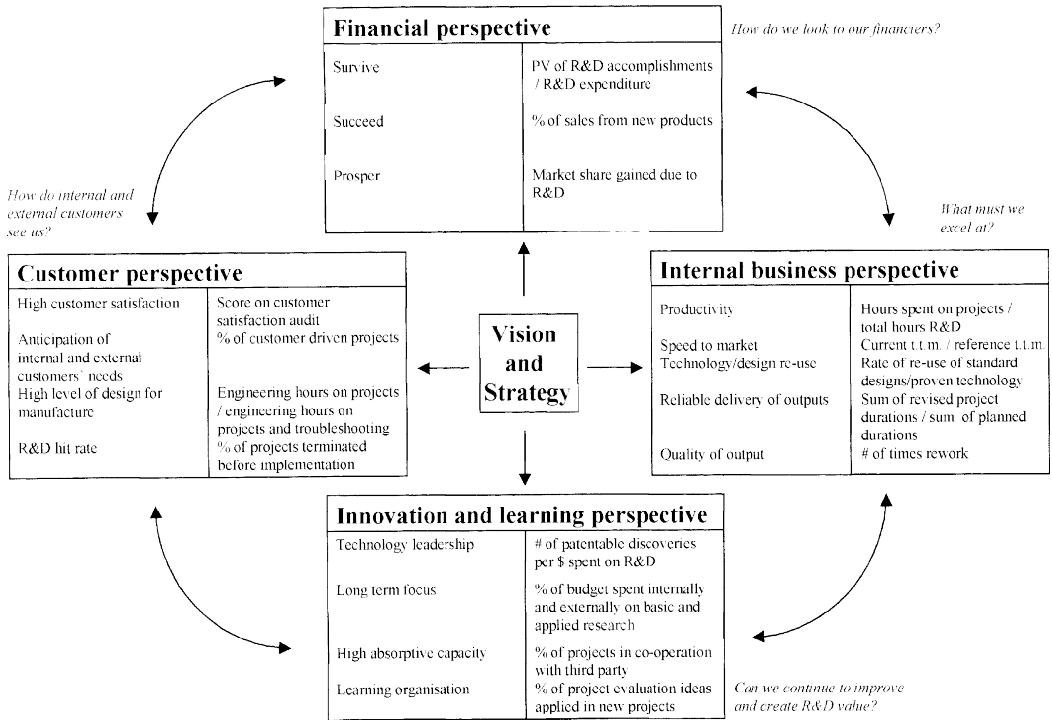


Figure 3. Example of a balanced scorecard for an R&D organization. (Adapted from R.S. Kaplan and D.P. Norton, Using the balanced scorecard as a strategic management system, *HBR* January–February 1996.)

with criteria such as pass or fail, very good, good, poor or very poor. In our field companies we also found this practice a lot (Kerssens-van Drongelen and Bilderbeek, 1996). The better companies use an unbiased staff manager to assess project performance in this situation. It is difficult to compare one qualitative appraisal with another as the answers are often arbitrarily awarded and never scored or ranked. Semi-quantitative methods convert a qualitative answer to a 'number' by applying a rank or a score. The accuracy and the reliability of the method may be improved by averaging the score for a metric from a number of experts. Personal preference and bias is a major problem with non-quantitative methods and should consequently be avoided if possible. To prevent a company bias, it is also recommended to involve the actual customer or a customer test group in the assessment when they can be identified. Some of our case companies used such an approach. In general, they were very satisfied with this measurement technique.

Quantitative methods use pre-defined algorithms to generate numeric measures that can be compared with reference points. A further distinction can be made between 'bean-counting' *bibliometric methods* and *economic methods*. Examples of metrics used in bibliometric analysis are number of patents approved per employee, number of new products per dollar spent on R&D. The main problem with these bibliometric methods is that they explicitly assume that there is a direct causal relationship between the metric and overall business success. Several people have argued against this (Brown and Svenson, 1988; Packer 1983). For example, it is often assumed that having more patents from a given level of input is good for a company. However, a company may have more 'ideas' than it can turn into new revenue earning products. In such a situation R&D resources are wasted. This example also illustrates the 'cause and effect' aspect of metrics and the old adage 'you get what you measure'. In order to look good, management may direct

resources to maximise the number of patent approvals and not to maximise R&D's contribution to business performance. Thus, such metrics should not be used as the single performance measure. In a balanced combination with other metrics, they might be useful.

Economic methods use more sophisticated algorithms than the bibliometric methods. Economic methods assume that both inputs and outputs can be translated into monetary costs and benefits. Economic methods are, due to the problems already given, used by only a small number of companies. For example, Alcoa uses the 'present value of an accomplishment' as a performance measure (Patterson, 1983) and Hitachi monitors R&D performance with the ratio 'R&D's contribution to profit / R&D costs' (Kuwahara and Takeda, 1990).

Frequency of measurement and reporting. It is generally accepted that, unlike some production activities, continuous measurement of R&D performance is not appropriate. Performance should be measured for key decision points, for example project milestones, annual budgets and periodic progress reviews. The milestones should be set so that prompt corrective action can be taken but not too frequent that they become a burden on the project. This applies especially for basic research where project deliverables usually occur sporadically. If the frequency of measurement is too great, staff may be demotivated if 'no progress' is repeatedly reported, or conclusions will be drawn too soon. Interesting in this respect is the finding by Bean (1995) that corporate labs in highly productive firms had substantial longer planning and review horizons (4–6 and 2–3 years respectively) than less productive firms which operated on an annual cycle. Of course this does not mean that project progress, project efficiency etc. should not be measured between the formal reviews.

Discussion

In this paper we have explored the state of the art in R&D performance measurement. Starting from general control theory, we have identified the most dominant problems

with performance measurement in research and development processes. In our field work we have sometimes encountered the attitude that, as it is so difficult to measure, one is better off not measuring at all. In some other companies, people were aware that their measurement system was not accurate, but still they used it because the metrics were easy to measure. However we strongly disagree with both opinions, as there are both external and internal demands for accurate performance measurement. External demands come from top management (or other (internal) customers if a market control mechanism is used for R&D funding), to which the R&D organization has to qualify for resources by proving that it is working on the right things rightly. Within the R&D organization, the need for performance measurement will also increase if quality management programs are to be implemented effectively. Without an effective, efficient, appropriate and meaningful performance measurement system, R&D may be isolated from the true business needs and its role in the business may be questioned.

To support developers of R&D performance measurement systems, we have listed some basic system requirements, the main design parameters and several design principles for the development of meaningful measurement systems. We have illustrated with empirical evidence found in literature and in our own empirical research projects. However, we have also argued that managers should not unthinkingly copy the concepts proposed by others but design their own tailor-made system, suiting the purpose(s) of measurement and the peculiarities of their R&D setting (type of R&D, organizational size, etc.).

From the literature overview it has also become apparent that although quite a lot has already been written about R&D performance measurement, there are still many areas that need further research. For example, research on the influence of contingency factors on R&D measurement system design is only just started and so far has mainly focused on metrics. However, from our empirical Dutch research we learned that contingency factors also influence the other system design parameters (Kerssens-van Drongelen and Bilderbeek, 1996). This seems a new area for

further research. Furthermore, we would like to mention the relationship between measurement purposes and R&D measurement system design as a future research topic. As has been summarized in this article, general performance control theory suggests that different measurement purposes require different measurement procedures. However, in the R&D management literature this topic has received little attention so far. Finally, we remind of the need for a general framework that would give a clear overview of what measurement concept to use for what purpose and in which context. Our empirical research has taught us that such a practical tool would be warmly welcomed by those assigned to develop a measurement system for their R&D process.

References

- Ansari, S.L. (1977) An integrated approach to control system design. *Accounting, Organisation and Society*, 2, 2.
- Bean, A.S. (1995) Why some R&D organizations are more productive than others. *Research • Technology Management*, January–February.
- Bridenbaugh, P.R. (1992) Credibility between CEO and CTO - a CTO's perspective. *Research • Technology Management*, November–December
- Brown, M.G. and Svenson, R. (1988) Measuring R&D productivity. *Research • Technology Management*, July–August
- Brown, W.B. and Gobel, D. (1992) Observations on the measurement of R&D productivity: a case study. *IEEE Transactions on Engineering Management*, 39, 4.
- Brownell, P. (1985) Budgetary systems and the control of functionally differentiated organizational activities. *Journal of Accounting Research*, 23, 2.
- Bruggink, A. (1989) Performance control in banking, theory and application. Thesis University of Twente, Enschede.
- Cook, A. et al. (1996) *The CSC Manufacturing Industry Handbook*. Solihull: CSC.
- Deetman, G. (1994) Project evaluation within Philips Research. *Course Book 'Managing the R&D Process' - Part II*. Enschede / Milan, pp. 277–288.
- Emmanuel, C., Otley, D. and Merchant, K. (1990) *Accounting for Management Control*. London: Chapman&Hall, pp. 57–67.
- Foster, R.N., Linden, L.H., Whiteley, R.L. and Kantrow, A.M. (1985) Improving the return on R&D — I. *Research Management*, January–February.
- Francis, P.H. (1992) Putting Quality into the R&D process. *Research • Technology Management*, July–August.
- Griffin, A. and Page, A.L. (1993) An interim report on measuring product development success and failure. *Journal of Product Innovation Management*, 10, 291–308.
- Griffin, A. and Page, A.L. (1996) PDMA success measurement project: recommended measures for product development success and failure. *Journal of Product Innovation Management*, 13, 478–496.
- Kaplan, R.S. and Norton, D.P. (1992) The balanced scorecard - measures that drive performance. *Harvard Business Review*, January–February.
- Kaplan, R.S. and Norton, D.P. (1993) Putting the balanced scorecard to work. *Harvard Business Review*, September–October.
- Kaplan, R.S. and Norton, D.P. (1996) Using the balanced scorecard as a strategic management system. *Harvard Business Review*, January–February.
- Kerssens-van Drongelen, I.C. (1994) Design of a measurement system for R&D: an overview of the issues. *Course Book 'Managing the R&D Process' — Part II*. Enschede / Milan, pp. 261–276.
- Kerssens-van Drongelen, I.C. and Bilderbeek, J. (1996) R&D performance measurement in large and medium-sized Dutch companies. Paper presented at the 3rd International Product Development Conference, Fontainebleau, 15–16 April 1996.
- Kumpe, T. and Bolwijn, P.T. (1994) Towards the innovative firm-challenge for R&D management. *Research • Technology Management*, January–February, pp. 38–44.
- Kuwahara, Y. and Takeda, Y. (1990) A managerial approach to research and development cost-effectiveness evaluation. *IEEE Transactions on engineering management*, 37, 2.
- Leeuw, A.C.J. de (1982) *Organisaties: management, analyse, ontwerp en verandering: een systeemvisie*. Assen: Gorcum.
- Loch, C., Stein, L. and Terwiesch, C. (1996) Measuring development performance in the electronics industry. *Journal of Product Innovation Management*, 13, 3–20.
- Meyer, C. (1994) How the right measures help teams to excel. *Harvard Business Review*, May–June.
- Miller, R. (1995) Applying quality practices to R&D. *Research • Technology Management*, March–April pp. 47–53.
- Moser, M.R. (1985) Measuring performance in R&D settings. *Research • Technology Management*, 28, 5.
- Packer, M.B. (1983) Analyzing productivity in R&D organizations. *Research Management*, January–February.
- Pappas, R.A. and Remer D.S. (1985) Measuring R&D productivity. *Research • Technology Management*, May–June.
- Patterson, W.C. (1983) Evaluating R&D performance at the Alcoa Laboratories. *Research Management*, April–May.
- Pritchard, R.D. (1990) *Measuring and Improving Organizational Productivity: a Practical Guide* New York: Praeger.
- Quinn, J.B. (1960) How to evaluate research output. *Harvard Business Review*, March–April.
- Robb, W.L. (1991) How good is our research?. *Research • Technology Management*, March–April.
- Roussel, P.A., Saad, K.N. and Erickson T.J. (1991) *Third Generation R&D, Managing the Link to Corporate Strategy* Boston: Harvard Business School Press.
- Schumann, P.A. Jr., Ransley, D.L. and Prestwood, D.C.L. (1995) Measuring R&D performance *Research • Technology Management*, May–June.
- Snellenberg, J.A.N.M. van (1995) Rapportage: ontwerpvariabelen voor de managementrapportage. In: *Handboek Management Accounting, D1510*, Samson Bedrijfsinformatie, The Netherlands (in Dutch).
- Weerd-Nederhof, P.C. de, Kerssens-van Drongelen, I.C. and Verganti, R. (1994) Part I: R&D Management. *Coursebook Managing the R&D Process*, Twente Quality Centre, Enschede/Milan.
- Wheelwright, S.C. and Clark, K.B. (1992) *Revolutionizing Product Development* New York: Free.