IJMR

June 2000

Performance measurement in industrial R&D

Inge Kerssens-van Drongelen, Bill Nixon and Alan Pearson

Currently, the need for R&D performance measurements that are both practically useful and theoretically sound seems to be generally acknowledged; indeed, the rising cost of R&D, greater emphasis on value management and a trend towards decentralization are escalating the need for ways of evaluating the contribution of R&D to corporate performance. However, although recent research and writing on the subject shows that the challenge of developing such sound measurements has been taken up by many academics and organizations, it is also clear that there is no generally applicable approach.

In this review, we consider various approaches for measuring the performance in industrial R&D and identify their key characteristics. We also include a brief summary of the 'history' of performance measurement in R&D, which shows that although there are some new ways of looking at the issue there are many examples from the past that can contribute to our current thinking.

The approaches found in the literature and practice are very varied in their application, some being more suitable for the project level, others for the R&D department, and some for the development process or for the organization as a whole. Furthermore, the uses of the approaches tend to be different. For example, some approaches are intended to justify the continuation of investment in R&D to upper management, whilst others are more suited to support learning and self-correction by empowered R&D teams. In this paper these uses, or 'functions', of performance measurement and a taxonomy of typical subjects of measurement in R&D environments are explored.

Finally, we conclude the review with a discussion of some limitations of the growing literature on R&D performance measurement.

Inge Kerssens-van Drongelen is from the University of Twente, the Netherlands, Bill Nixon is from the University of Dundee, UK, and Alan Pearson is from the Manchester Business School, UK.

Trends in R&D Arousing the Interest in Performance Measurement

Several studies have pointed to changes in the business environment that have taken place over recent decades, which have had a substantial impact on the way R&D processes are practised and managed, though it has also been noted that these trends do not apply equally throughout all industries (see, for example, Clark and Fujimoto 1991; Gupta and Wilemon 1996; Kumpe and Bolwijn 1994; Larson 1998; Szakonyi 1998; Wheelwright and Clark 1992).

In general, the following changes in the R&D environment are widely acknowledged:

- Shift from a sellers' market (in the 1950s and 1960s) to a buyers' market with increased domestic and international competition and a global marketplace.
- Explosion of market segments and niches in response to an increased demand by sophisticated, discerning customers for targeted, customized products.
- Faster-changing customer requirements, which truncate product life cycles and demand faster new product introductions.
- The growth of scientific and engineering capabilities world-wide, which nurtures the continuous development of new technologies and results in a decrease in technology life cycles.
- Dramatic increase in the breadth of technologies in many products.
- Growth in the number of technologies from which companies can choose.
- An increase in the rate of technological dissemination, development of global networks and a move towards virtual laboratories.
- Increased government regulations and societal pressures regarding environmental, safety and health issues.
- Increased pressure on R&D departments to be accountable to business needs.

Consequently, while a few decades ago efficiency was regarded as the single most important performance indicator for companies, nowadays companies need to excel simultaneously in efficiency, quality, flexibility and innovativeness (Bolwijn and Kumpe 1990; McNair and Liebfried 1992; Wheelwright and Clark 1992). In addition, a more customer- and profit-oriented approach to product development has increased the requirement for R&D to support a broad array of competitive bases such as differentiation, time-to-market, value, service and economic product proliferation (Cooper 1995; Cooper 1998; D'Aveni 1995; McGrath 1996; Morone 1993; Smith and Reinertsen 1998). These increased demands have brought about a more integrated and strategic role for the R&D function (Athaide and Stump 1999; Comstock and Sjolseth 1999; Edelheit 1998; Khurana and Rosenthal 1997; Stillman 1997; de Weerd-Nederhof *et al.* 1994).

Roussel *et al.* (1991), and Kumpe and Bolwijn (1994) have extensively described the evolution that has taken place in the *management* practices of many R&D organizations in response to these increased pressures on R&D. They have outlined a transition from a 'strategy of hope' characterized by the expectation that – given the right mix of brains, money, equipment and time to pursue ideas – scientists and engineers, left alone, will concoct new profitable products and processes, to a strategically and organizationally embedded form of R&D management. In this 'third generation' R&D management concept, R&D strategies and business strategies are closely linked and, at both the strategic and operational levels, the R&D department has to co-operate with other departments to manage the cross-functional R&D processes. This practice is often strengthened by making a part of the R&D budget directly dependent upon project- or program-based funding to be determined by Business Unit management instead of putting fixed annual budgets at the disposal of R&D managers (Chester 1995; Robb 1991).

The term 'fifth generation' R&D management (Amidon Rogers 1996) endeavours to capture the changes in R&D management since 1991, most notably the move from cross-functional new product development teams to collaborative, networked groups in virtual organizations, from R&D portfolios to integrated technology platforms, from accelerated product development to seamless innovation through concept to customer and from technology creation to technology selection (Chiesa and Manzini 1998; Iansiti and West 1997; Klein 1998; Lewis 1998; Robb 1991).

Besides linking R&D closely to business needs, large multinationals, and increasingly also smaller technology-intensive companies, have responded to the market changes by internationalizing their R&D processes (Bradley *et al.* 1993; Gassmann and von Zedtwitz 1998; Kumar *et al.* 1998; OECD 1998). This way, R&D becomes more sensitive to requirement differences of local markets and to technologies emerging in foreign centres of expertise. At the beginning of the internationalization process, the central R&D facility usually remains dominant. Currently, a trend can be observed towards integrated R&D networks consisting of a small number of leading research centres, which optimally exploit and continuously refine local strengths and expertise (Chiesa 1996; Gassmann and von Zedtwitz 1998). Expertise areas of these research centres may overlap, often resulting in joint multi-site projects as well as a sense of competition between the R&D centres. This competition may be strengthened further by the new trend of establishing research centres in low-wage areas such as Eastern Europe and India. Managing such international R&D networks requires a change from simple control structures to a set of complex co-ordinating structures.

In addition to the changes in R&D *management*, academic researchers have observed profound changes in the way R&D is practised on the 'shop floor'. For example, in a large survey of US service and manufactured goods firms, Griffin (1997) found a steady growth in the number of organizations that have implemented a formal process, e.g. Cooper's (1992) 'stage-gate' approach, for controlling New Product Development projects. She also noted that the use of cross-functional NPD project teams has by now become endemic, especially for more innovative projects. Alongside the increased involvement of other disciplines within the company, academic researchers have also noted an increasing degree of co-operation with *outside* partners, which, although not yet really widespread (Gupta and Wilemon 1996), seems to be growing steadily (Cauley de la Sierra 1995; Steensma 1996; Wood 1998). Finally, Hustad (1996) and McDonough and Griffin (1997) have noted the steady growth in the number of tools, techniques and organization forms that have been proposed and tried as aids to improve R&D performance. These include for example: co-location, heavyweight project leaders, quality function deployment (QFD), focus groups, voice of the customer, critical path method (CPM), rapid prototyping, incentive schemes, skunk works and time allocated for creative activities. Though all these mechanisms have their success stories, it has become clear that their return is contingent (McDonough and Griffin 1997), leaving R&D managers with the problem of determining which mechanisms are appropriate in their specific situation.

Overall, we conclude that the management and the practice of R&D have become more complex, as they involve many parties and have a wide range of, often interrelated, technological, market and organizational options to choose from under constrained conditions. At the same time, it has been observed that top management's attention to R&D is rapidly increasing (Wood 1998), holding R&D accountable for directly contributing to business results not just in the long term but also in the short term (Gupta and Wilemon 1996). Together, the complexity, the growing importance of effective and efficient R&D for company success, and the increased pressure on R&D to become accountable for its actual contribution to company success, have aroused a need to implement performance measurement and control tools in R&D. This need becomes even clearer when one observes that in several best-practice studies (e.g. Cooper and Kleinschmidt 1995; Griffin 1997; Pittiglio *et al.* 1995) 'measuring performance' was found to be one of the discriminating factors between 'the best' and

IJMR
June 2000

the rest, and that on average the capabilities of R&D groups in this area are low (Gupta and Wilemon 1996). Shareholders are also demanding more information about the precise contribution of R&D to corporate performance; institutional investors, in particular, have made it quite clear that they cannot attribute a proper value to a company's reported R&D expenditure unless information is also provided that indicates its effectiveness (Nixon 1998). Fortunately, over recent decades general control and measurement theories have matured and have gradually been elaborated for R&D environments, resulting in a growing stream of publications on this subject. These are briefly reviewed in the next section.

History of R&D Performance Measurement

Background

Performance measurement has been and may always be, a technically difficult and possibly an emotive issue in most areas of management, not least because it is almost inevitably going to influence decisions about the allocation of resources and also may be linked, implicitly if not explicitly, to rewards. This is not different in R&D; however, the difficulty of evaluating R&D performance (see the next section) is compounded by the complexity and dynamic of the R&D environment, by the fact that it is a sub-system of the new product development process, that there is a considerable time lag between inputs and outcomes and by the fear that control may constrain creativity (European Industrial Research Management Association (EIRMA) 1995; Ellis 1997). Nonetheless, the research undertaken for this paper suggests that many companies are now attempting ''to lay aside all the old excuses for not measuring and baselining R&D effectiveness and to do it anyway'' (Francis 1995, 47).

Early writings on R&D performance assessment, such as Cook (1966) and Edwards and McCarrey (1973), might lead one to believe that the problem of allocating scarce resources across competing projects was simply to support those 'experts' who were most likely to succeed in those areas that were most likely to be 'ripe' for development. Problems inevitably arise in deciding who are the so-called experts. An often-used approach in this respects was *peer judgement*. However, as pointed out by Keller and Holland (1982), peer judgement often suffers from the 'Matthew-effect': those who already have would tend to get more, not because their current ideas are better than other researchers', but simply because of their reputation. In respect of which areas to work in, again reliance was often on peer judgement (Reynolds 1965) with the possibility of more of the same rather than backing potential 'paradigm'-breaking initiatives. Roussel *et al.* (1991, 26) typify this era – often referred to as 'first generation R&D management' – as follows:

> R&D is an overhead cost, a line item in the general manager's budget. General management participates little in defining programs or projects; funds are allocated to cost centres; cost control is at aggregate levels. There is minimum evaluation of the R&D results other than by those involved in R&D. The R&D activity is relatively isolated and there is little communication from R&D other than to say, ''Everything is going fine.'' There is only a modest sense of urgency: ''Things are ready when they are ready''

Many of the early attempts at the use of *quantitative approaches* for project selection and portfolio management were based on analysing the link between inputs and outputs, with the most obvious example being the effort put into trying to apply the principles of mathematical programming to R&D [see, for example, Bell and Read (1970) for a specific example, and Gear *et al.* (1971) or Steele (1988) for a review].

Early attempts at *improving*, as distinct from *measuring* performance, include the effort put in by companies such as Allied Corporation and Unilever in trying to relate R&D more closely to company policy, bearing in mind the need not only to seek a closer alignment, but also to allow for some misalignment in order to encourage research in areas that were potentially relevant to the business but for which no relatively immediate need could be identified (Allio and Sheehan 1984; Hubert 1970). This clearly recognized the importance of seeking an appropriate balance between the short and long term – the former being more likely to be emphasized by the 'business' and the latter by R&D. The early attempts at R&D performance management also indicated the value of developing 'competencies', although this particular word was not common currency at the time. The value of allowing 'free time', usually said to be about 10%, for scientists to pursue their own ideas, was often emphasized (see, for example, Reynolds 1965), although it was rare to find this was effectively used, as most scientists were overloaded with current demands. The possibility of overcoming this short termism by separating research from development generally seemed to lead to the 'Ivory Tower' syndrome, and there are few organizations left who follow this path (Kumpe and Bolwijn 1994).

In the R&D management literature, Quinn (1960) was probably one of the first authors to apply the causal chain concept to performance measurement in R&D. He has structured his discussion of R&D performance evaluation around the idea that "the three phases of research" – meaning Fundamental Research, Applied Research and Development – "… produce technology … which has value as opportunities to exploit … and can be exploited by the company … to support company goals" (p. 70). In line with this he proposed to measure:

- current R&D work;
- R&D outputs in terms of "expected economic value compared to costs" and "productivity";
- the eventual profits obtained from technologies actually adopted by the businesses.

Though in a somewhat less formalized form, most elements of the framework presented almost three decades later by Brown and Svenson (1988) were already there. The Brown and Svenson framework (see Figure 1) has been widely used, though sometimes slightly adapted, in R&D management research (see for example Cordero 1990; Lee *et al.* 1996; Loch *et al.* 1996; Schumann *et al*. 1995).

Developments in R&D Performance Measurement Practice

Already by the early 1970s, there were companies that measured certain aspects of their R&D and gained major benefits from it. For example, Hardingham (1970) discussed the format, use and benefits of a simple but effective multi-project monitoring tool used at the Nuclear Power Group. This was used to measure progress and the occurrence of specific causes of delay across all projects. Hardingham reported, amongst others, the following returns provided by the tool:

- For the first time in a complex situation, with projects being started and closed almost every day, it was possible to keep an up-to-date list of basic reference data.
- The discipline of having to review progress on each project each month helped section leaders in reviewing the work under their control and ensured that projects were not forgotten.
- Analysis of patterns and trends in data gathered over the preceding 2 years elucidated several common reasons for delay, which subsequently became subjects for improvement efforts. The measurement procedure provided the feedback as to whether these improvement efforts had the expected impact.
- It helped the section leader to schedule manpower under his direct control.

At Stauffer Chemical (Galloway 1971) and Alcoa Laboratories (Patterson 1983), a more output- and outcome-focused performance

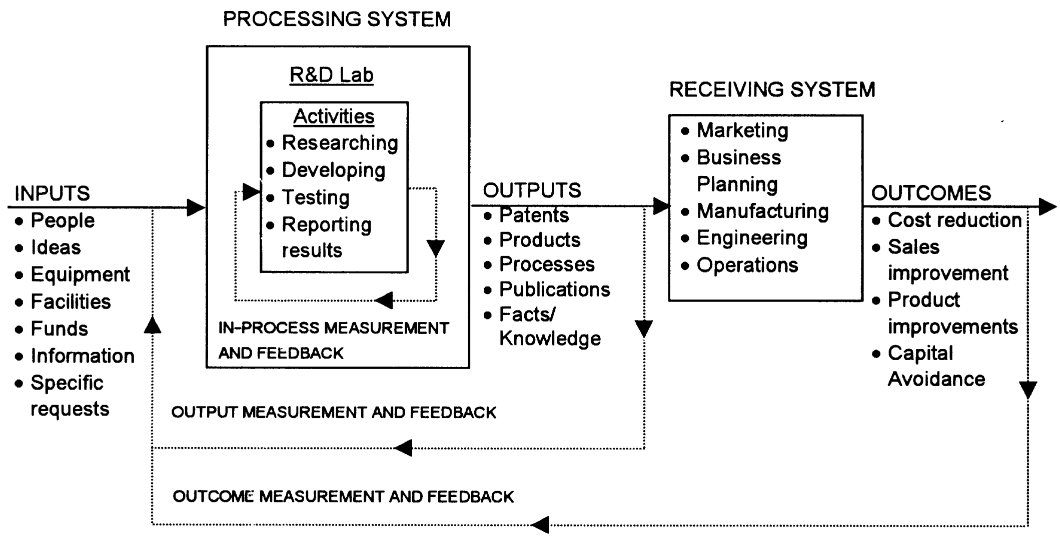

Figure 1. An R&D measurement framework.

Source: Brown and Svenson (1988).

measurement procedure was used. At Alcoa, plant industrial engineers and headquarters' marketing personnel (the internal customers of R&D), together with R&D's technical planners, estimated quarterly the 'economic benefit' (present value) stemming from major R&D accomplishments implemented in that quarter. As not all accomplishments were immediately adopted, they also measured the number of major accomplishments submitted, the number of accomplishments immediately implemented and the resulting stock of accomplishments pending. The trends in these measurements were plotted and analysed and were used in several ways, including, for example, to legitimize R&D for corporate management in financial terms rather than parlance and as input for decision making by those responsible for R&D planning and budgeting. Though acknowledging that the economic benefit estimates were based on uncertain sales and profit projections, Patterson nevertheless reported the benefits of the *ex ante* measurements, as perceived by the Alcoa managers, to include:

- "As major accomplishment achievement has begun to imbue the thinking of researchers, they are more vigorously seeking work with financial benefits to the company" (p. 25).
- "Many are awakening to the need for effective implementation and the propriety of R&D involvement up to the point of economic fruition" (p. 26).
- "Analysis of the statistics reveals that about 80% of economic benefits come from only 20% of the accomplishments. Concentration on a few vital projects is now pursued via centralised project prioritisation and intensified project management of key programs" (p. 26).
- "By demonstrating mastery in implementation of near-term technology, the R&D organisation creates a more favourable climate for sustaining a reasonable balance between short and long-range work" (p. 26).

In the past three decades, various other companies have reported the use of similar output- and outcome-focused methods, e.g. at Borg-Warner (Collier 1977), General Electric (Robb 1991) and Hitachi (Kuwahara and Takeda 1990), or other types of measurement procedures, e.g. a *technical audit* used by 3M (Krogh *et al.* 1988), *project performance measurement procedures* used at GM Hughes Electronics (Chester 1995), Hewlett Packard (House and Price 1991) and Unilever (Rainbow 1971), and an *individual performance measurement procedure* used at Olin Corporation (Endres 1997). Table 1 highlights important characteristics of these measurement procedures. Other papers describing measurement procedures used in practice include Cook (1966), Foster (1996), Reynolds (1965), Schainblatt (1982) and Tenner (1991).

Despite these early examples from practice of working R&D performance measurement systems, recent findings of Gupta and Wilemon (1996), based on 120 R&D directors from technology-based companies in the US, indicate that, although there was a growing need to measure R&D performance, most R&D directors acknowledged that they still had limited knowledge in the area (see also Cooper *et al*. 1998; Szakonyi 1998). The fact that R&D performance measurement is still perceived by many managers to be difficult, and that some even prefer to evade the issue altogether, might partly be attributed to some intrinsic problems, as will be put forward in the next section.

The R&D Performance Measurement Problem

The first problem of R&D performance measurement is the difficulty of accurately *isolating the contribution of R&D* to company performance from the other business activities because it is always the intertwined efforts that eventually result in outcomes in the marketplace (Hodge 1963). Even if this attribution problem could be solved, it would still be difficult to isolate the contribution of the R&D *department* from those of the other disciplines represented on the project team, let alone the contribution of an individual researcher. It is generally acknowledged that the contribution of groups of people cannot be determined objectively, and that one has to rely on subjective assessments. However, it should be noted that subjectivity does not necessarily imply inappropriate measures. As reported by Robb (1991), the percentage of sales and profits from various products attributed to R&D by the operating departments were very similar to those mentioned by the researchers themselves.

Of course, even if it is technically possible to assess the contribution of particular specialist departments like R&D, Design and Marketing to project outcomes, consideration must be given to the possible impact of such evaluations on behaviour and, in particular, on the cross-functional communication and collaboration that are so critical to the success of innovation (Katzenback and Smith 1993; Quinn *et al*. 1997). Performance measures can encourage competitiveness but they can also encourage ‘functional mindsets’ (Majchrzak and Wang 1996) and divisiveness if the focus of the measures is on participant groups rather than on the team. Some companies like Asea Brown Boveri (ABB) allow the pressures for competitiveness to coexist among the group’s 5000 highly decentralized profit centres. However, ABB takes great care to foster collaboration through formal and informal organizational arrangements (Barham and Heimer 1998). At a project level, CCM Ltd, a small UK engineering company, undertakes feasibility studies that evaluate, *ex ante*, the ability of internal functions to make the required contribution; this prospective assessment encourages internal competitiveness because functions are benchmarked against possible external suppliers to determine outsourcing options. However, once projects commence, the focus of the performance measurement system (PMS) is on the team’s performance and no attempt is made to evaluate, *ex post*, the contribution of participant groups such as R&D. Some of the reasons advanced by the joint managing directors of CCM Ltd for this approach are:

- It encourages the collaboration and openness which they regard as a prerequisite to successful new product development.

IJMR
June 2000

Table 1. Examples of R&D measurement procedures reported by practitioners

Company	Procedure characteristics
3M (Krogh *et al.* 1988)	*Technical Audit System* • management/peer review method • numerical rating as well as essay comments on running programs in terms of technical and business factors and on general organization aspects of the unit • every unit audited once in 2--3 years • implementation of audit recommendations not mandatory
Borg-Warner (Collier 1977)	*Project Goals Achievement Ratings* • annual rating per project of actual achievements vs. goals on a 13 scale, aggregation per research group via weighted average (weights based on project budget) *Business Opportunity Evaluation* • potential economic value (assuming 100% market share) of projects completed in the last year • provides a 'return on research' indication when related to R&D investments
General Electric (Robb 1991)	*Jimmy Steward Test* • counting (sales & profits of) BUs which exist thanks to an R&D invention *Direct Output Counting* • e.g. No. of patents related to inputs, licensing income, royalties payment avoided due to patents *Valuation of Technology Transitions* • Estimation of contribution of R&D to (expected) cash-flows of products/processes introduced in the market in the past decade • R&D's credit % is determined by the receiving BU

GM Hughes Electronics (Chester 1995)	*Laboratory Performance Measure*
	• mix of three metrics:
	    • Technical Excellence: accomplishments at or beyond state of the art, as annually evaluated by external experts
	    • Business Unit Objectives: performance against these objectives evaluated twice a year by BUs
	    • General Management Objectives: meeting R&D and indirect expense objectives and meeting diversity objectives (scored by Finance and the Workforce Diversity Staff)

Hitachi (Kuwahara and Takeda 1990)	*R&D Cost Effectiveness Measure* • project sponsors in the business groups annually estimate (using guidelines) the % of profits accredited to R&D from projects finished in the past 5 years ('R&D Contribution to Profit' (RCP)) • RCP divided by the costs of these projects provides the 'R&D cost effectiveness measure'

Hewlett Packard (House and Price 1991)	*The Return Map*
	• chart with money (development cost and returns) on the *Y* axis and time on the *X* axis
	• chart is updated and discussed by the project team at various points in project life, including at the project start, at Manufacturing Release and after Commercial Release (estimates are gradually replaced by actual realizations)
	• metrics derived from the chart:
	    • break-even-time (point where cumulative profits=investment)
	    • time-to-market (start to manufacturing release (MR))
	    • break-even-after-release (time from MR until investment is recovered by profits)
	    • return factor (profits divided by investment)

Olin Corporation (Endres 1997)	*Performance Management Cycle* • an employee's key customers' expectations plus individual growth objectives are set as personal objectives for the coming year • performance progress is discussed with supervisor twice a year • the 'key customers' are asked annually to evaluate performance and to formulate new expectations • aggregated info is discussed with 2nd level manager by supervisor

Unilever (Rainbow 1971)	*Progress Assessment Procedure*
	• monthly report by project leader of:
	    • latest estimates of completion date and cost to completion
	    • technical feasibility index (TFI) (project leader's assessment)
	    • rate determining factors (RTF) (non-scientific factors of delay)
	• the project scores are accumulated in an overall progress exception report per R&D division

- Project problems and aims are everyone's responsibility and cannot be assigned exclusively to a single function or individual without risking sub-optimal results.
- The dominant criterion of a successful product is, in the view of the directors, success in the marketplace, and the market, they contend, rewards an outstanding integrated product rather than superb componenets (Nixon 1998).

A second problem with measuring the contribution of R&D to the company is that a part of the benefits it generates is hardly quantifiable: for example, giving the company a high-tech image may attract new high potential employees. A related dimension of this problem is one that economists have always acknowledged, namely the positive force for economic growth of R&D spillover effects; that is the difficulty that firms have in capturing the benefits of new knowledge for themselves alone and the tendency for knowledge to flow across firm and industry boundaries (Mansfield 1981; Rosegger 1986; Rosenbloom and Spencer 1996).

A third issue is the problem of *matching* specific R&D inputs (in terms of money or man-hours) and intermediate outputs (research findings, new technologies, new materials, etc.) with final outcomes (including new or improved products and processes) (Hodge 1963). At the start, and sometimes even after the completion, of basic and applied research projects, it can often not be foreseen in which products or processes the outputs will eventually be used (Mechlin and Berg 1980). Additionally, it is difficult to indicate which previous R&D projects' findings (and hence how much resources) have actually been used in a development project, because the knowledge currently applied is built on various previous projects, including 'failed' projects that have generated knowledge on approaches that were *not* (yet) feasible (Galloway 1971). Though this implies that a *precise* benefits/inputs measure cannot be obtained, *rough* indications of potential project benefits and of the main knowledge sources used as input can often be made. For many applications, such measures appear to be sufficiently informative (see, for example, Collier 1977).

However, comparisons of the R&D expenditure and outcomes of one company with those of another are especially difficult and can be very misleading unless care is exercised. A House of Lords Select Committee on Science and Technology (HOL/SCST) of inquiry on R&D definitions noted that

> [T]he facts (including R&D expenditure) are clouded by uncertainty over definitions and false comparisons between like and unlike. (HOL/SCST 1990, 5)

The trend towards the more integrated, 'simultaneous' development of new products and processes has accentuated the long-standing problems of defining and measuring R&D expenditure (Nixon 1995).

A fourth major measurement problem is the *time lag* between R&D efforts and their pay-offs in the marketplace. This time lag is especially apparent with basic research, but also applies to applied research or development projects. For example, companies designing automotive components for vehicle manufacturers sometimes have to wait up to two years for production to start and the first product sale (Kerssens-van Drongelen and Cook 1997). At Bell Labs, it was reported that there was a typical time lag of between 7 and 19 years for basic research projects (Pappas and Remer 1985). The time-lag problem makes the (financial) outcome metrics inappropriate inputs for decision making regarding the concerned R&D project, as by the time they become available it is obviously too late for correction. Thus, for this measurement function, other, more timely, measurable metrics have to be sought. However, for other measurement system functions such as learning and motivating people by giving feedback, outcome metrics could still be quite useful.

IJMR
June 2000

Besides problems with the selection of performance *metrics*, there is also the problem of determining the right *norms* to compare with (Schainblatt 1982). By its very nature every R&D project is a unique, non-repetitive process (de Weerd-Nederhof *et al*. 1994). It is consequently considered to be difficult to compare and contrast two projects, as they will always be different. It should be noted, however, that real project uniqueness is more applicable to basic research projects than to product development projects. Usually, the *kinds of activities* required for a development project could largely be defined and scheduled in advance. Moreover, based on past experiences and group discussions, rough norms could be developed for each type of activity, given a certain degree of complexity (Quinn 1960). At a more aggregated level (e.g. the performance of the R&D department in general), the norm setting problem also occurs. Again, by building records of past performance from which trends and norms can be derived, and via benchmarking, norms can be developed. Naturally, such norms should not be static, but should be subject to improvement and adaptation to external developments when required. However, it is generally accepted that, to be able to control a process, one first has to have some sense of what has been normal and possible in the past and from there work on further improvement.

The final problem 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 hold to be more significant, is that they have had bad experiences with inappropriate measurement systems leading to improper decision making and, consequently, believe that all measurement systems do not work. Yet, not all scientists seem to reject performance measurement, as was pointed out by Werner and Souder (1997b), who found that the acceptance differed between countries, Americans being much more amenable to performance measurement than Germans. Overcoming negative feelings towards performance measurement takes persistence from management and time. As Hardingham (1970, 47) noted: "the most important lesson learned in the entire exercise has been the need to move forward slowly, asking for the minimum of data and giving the maximum amount of useful operating information in return". Furthermore, it may be helpful to determine and communicate in advance what will be the purpose of the measurement activities. Subsequently, it has to be ensured that approach adopted suits this particular purpose and the characteristics of the organization.

Range and Nature of Measurement Approaches

Having described several examples of R&D performance measurement approaches and discussed the measurement problems, we now have a notion of what R&D performance measurement is and what the difficulties are, but, in order to compare the different approaches or to evaluate the suitability of an approach for a specific application, it would be useful if we could define the concept of R&D performance measurement more precisely and characterize different approaches according to their characteristics. This will be done in this section.

The 'R&D Performance Measurement' Concept

R&D performance measurement can be seen as a subset of the broader concept of 'R&D

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. Controlling is often done as a repetitive 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 courses of action to influence people and processes so that the efforts and outputs are aligned with company objectives and plans.

In accordance with this definition, *performance measurement* can be interpreted as:

> that part of the control process that has to do with 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 PMS can be defined as:

> the mechanism supporting the measurement process, by which the required performance information is *gathered, recorded* and *processed* (see Figure 2).

A PMS usually consists of several measurement *procedures*. Each procedure is designed to fulfil a certain *purpose* and is characterized by a specific combination of the following aspects (Kerssens-van Drongelen and Cook 1997):

- metric(s) (performance aspects/indicators), sometimes intentionally organized in a specific way;
- measurement method(s);
- norms;
- frequency and timing of measurements;
- reporting format.

The *purpose* of a measurement procedure can be considered as the combination of the *subject* to whom the procedure has to apply (i.e. one or more people), and the *function* the procedure has in the performance control processes of that subject. For example, a purpose can be to motivate a product development team (= the subject) via performance

IJMR
June 2000

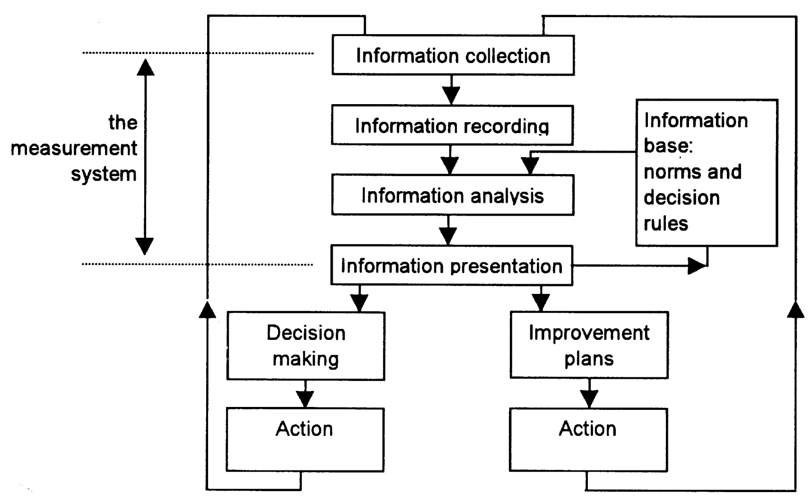

Figure 2. The measurement system's impact on the control process.

Adapted from Kerssens-van Drongelen and Cook (1997).

feedback coupled with rewards (= the function), or to give the R&D manager insight into whether, and if so which, corrective actions are necessary (= the function) in the activities of a technical discipline group (= the subject). These functions and subjects of performance measurement will be discussed later in this paper.

A performance measurement *approach* can be considered as a specific combination of metrics and measurement methods. In many performance measurement publications, metrics and measurement methods seem to be considered as intertwined concepts. For example, Ahaus (1994, 143) described a performance indicator as:

> a description of a subject, measurement scale and a measurement procedure. A performance indicator is the operationalisation of a non-measurable goal.

Kerklaan *et al.* (1996, 208) gave the following definition of a performance indicator:

> an instrument to measure a predefined part of the performance of a process in order to monitor the development of this performance. A complete indicator consists of a measure, a norm, a measurement instrument and a registration technique.

In our terminology, this last definition is closer to the concept of a measurement *procedure* than to the term 'metric', which we consider to be only a part of such a procedure. In this paper, we prefer to separate metrics and measurement methods, since many metrics can be measured in different ways. For example, the metric 'customer satisfaction' of a new product can be measured by an index of items in a customer questionnaire, by asking sales people to assess on a simple 'Likert-type' scale how they perceive the satisfaction of customers, by counting the recorded number of customer complaints or by extracting the monthly sales volume of this particular product from the company's financial records. Thus, it seems useful to consider the various measurement methods and their advantages and disadvantages separately from the various metrics that can be measured by using these methods. This will be done in the next two sections.

A Taxonomy of Measurement Methods

In the R&D literature there are several publications addressing possible classifications of measurement methods for R&D environments (see, for example, Azzone and Masella 1992; Brown and Gobeli 1992; Brown and Svenson 1988; Packer 1983; Pappas and Remer 1985; Quinn 1960; Werner and Souder 1997a). Two basic distinctions seem to underlay these classifications.

The first distinction usually made is whether a method leads to a *quantification* of the metric value or only to a *qualification*. For example *computational methods* clearly lead to a *quantitative* value (e.g. the time-to-market has been 6 months), whereas *assessment methods* usually result in a *qualitative* indication of the metric value (e.g. the time-to-market has been 'good' or 'unsatisfactory'). Quantifying methods are often further subdivided into *financial* methods using monetary data (e.g. profits or NPV) and *non-financial* methods counting and computing other quantitative attributes of a subject (e.g. time-to-market, a quantitative technical product specification such as memory usage or weight, or the number of new products introduced per year). A method that combines financial and non-financial values is usually also classified as financial (e.g. the number of patents per $ invested in R&D). Sometimes, qualitative values are *converted* into a numerical value through anchoring the qualitative indicators to definitions or behaviours that can be expressed as numerical equivalents (e.g. on a five-point scale). For conversion of qualitative into numerical values, four techniques may be used: profiles, scaling models, checklists and scoring models (Werner and Souder 1997a). Measurement methods that use this conversion of judgements into numerical values are often referred to by the term *semi-quantitative*. Within the cluster of semi-quantitative methods, a further subdivision is sometimes made between:

(1) methods resulting in a list of ratings of several items, from which the value of the desired aggregate metric (e.g. effectiveness or quality) has to be deduced *judgementally*, or is simply assumed to be the *average score* of all items (see, for example, the method proposed by Szakonyi 1994b); and
(2) methods that convert the items and numerical values into one or a few aggregate metrics using *weighting*, *sophisticated formulae* or *statistical procedures* (e.g. factor analysis, Packer 1983).

In this last type, we also include methods that use a combination of quantitative values and converted qualitative values, resulting in an *integrated metric* (Werner and Souder 1997a), since these are at least partially still qualitative.

A second basic distinction that can be made is whether a method uses *objective information* or relies on *subjective judgements*. This distinction is usually labelled 'objective versus subjective' or 'non-judgemental versus judgemental'. Though this distinction is sometimes blurred with the terms quantitative versus qualitative, we think it is useful to discern between the two. For example, the *estimation* of the net present value or option value expected for a project is a method to assign a value to the metric 'project profitability'. Though the output of the method is quantitative and the mathematical computation may be rather sophisticated, the basis of the method remains a subjective estimation of future cash flows. Furthermore, it seems useful to make a distinction between *degrees* of subjectivity. Whereas judgements made by one person directly involved in the subject of measurement can be qualified as highly subjective, the judgement of a *group* of *external* experts is usually considered as much more objective and credible (Brown and Svenson 1988). Thus the *degree of involvement* of the evaluator(s) in the subject of measurement as well as the *number* of evaluators seems to be important. Though still based on judgements, we will label methods involving (several) people who are not directly involved with the subject of measurement as *semi-objective non-involved persons' measurements*. These people can be business unit managers, an audit team, or others, but not the subject itself, the direct supervisors of the subject, or peers closely co-operating with the subject. A group of people falling in-between involved and non-involved are *customers* who, especially in business-to-business markets, are often closely involved in the R&D process. Being the customer, they will probably be more critical about the performance than an involved person, but less objective than completely non-involved ones. Therefore, we have made a special cluster of measurement methods in which the customers are involved, labelled *semi-objective customers' measurement.*

Note that the determination as to whether someone is involved or not depends on the subject of measurement chosen in a particular situation. For example, in an R&D department with several sub-departments, if the subject of measurement is an individual researcher or a team, then the R&D department manager can be considered as not directly involved, and subsequently his assessment might be considered semi-objective. However, if his own department is the subject of measurement, than he is clearly involved and his opinion should be regarded as subjective.

Further, it should be noted that our distinction of four clusters – subjective measurement, semi-subjective customers' measurement, semi-subjective non-involved persons' measurement, and objective measurement – is still rather rough. For example, subjective estimations, such as net

IJMR
June 2000

present value predictions by a project team, will normally be considered more objective when they are based on up-to-date market information and data from comparable past projects than when they are made without such information. Also, the already mentioned factor of the number of people involved in the measurement is not taken into account in our distinction. If 20 involved persons assess a certain project quite similarly, their judgement will probably be considered more credible than the judgement of any one of them alone. However, for reasons of simplicity we have not made further subdivisions in these respects. At least, the four identified clusters along this dimension are more distinctive than the dichotomies found so far in literature (see, for example, Brown and Gobeli 1992; or Pappas and Remer 1985). If we combine the two discerned dimensions 'degree of quantification' and 'degree of subjectivity', we obtain a taxonomy with ten main clusters, as presented in Table 2, some of them being split further in two sub-clusters.

To discern between qualitative and semi-quantitative methods, we use the terms *assessments* and *ratings*, respectively, while the distinction between judgemental-quantitative and objective-quantitative methods is made by using the terms *estimation* and *computation*. In most of the clusters, we have listed examples of measurement methods belonging to that cluster, which were reported to be used in the R&D measurement procedures already listed in Table 1. Note that the clusters 'qualitative-objective' and 'semi-qualitative-objective' will not exist, since these are contradictions (Werner and Souder 1997a), but for the other empty clusters examples can easily be found. For example, procedures fitting the 'subjective rating with sophisticated conversion' category are the NewProd method for predicting project success or failure (Cooper 1992) and the RACE II method for rating the maturity of organizations with respect to Concurrent Engineering (de Graaf 1996).

A Taxonomy of R&D Metrics

Very different ideas exist on what the concept of 'metrics' actually includes. In this paper, we adopt a slightly adapted version of the definition of a performance indicator used by Fortuin (1994, 143):

> A performance indicator is a variable which indicates the effectiveness and/or efficiency of a process, system or part of a system when compared with a reference value.

In contrast with the definitions of Ahaus (1994), and Kerklaan *et al.* (1996) cited before, this definition does not include the measurement method. Neither does it include the norm, although it indicates that *having* a kind of 'norm' (reference value) is crucial for meaningful interpretation of the measured value. Furthermore, unlike Ahaus and some other authors (e.g. Smith 1976, 746), this definition does not limit the term metric to operationalizations of *non-measurable* goals, but includes also measurable ones.

We have found considerable academic publications in which various metrics for R&D environments have been proposed (e.g. Chiesa *et al*. 1996; Dressler *et al*. 1999; Foster *et al.* 1985; Griffin and Page 1993, 1996; Hauser and Zettelmeyer 1997; Kerssens-van Drongelen and Bilderbeek 1999; McGrath and Romeri 1994; Moser 1985; Tipping *et al*. 1995; Werner and Souder 1997a). However, it is useful to present existing metrics in a way that makes it easier to compare different measurement approaches and to select a (set of) metrics in accordance with the purpose of measurement and the context. This in turn requires a taxonomy that helps to discriminate between types of metrics. Since we found various sorting principles that could be used to build such a taxonomy, we first discuss some of these principles before selecting those which seem to be the most useful.

Sorting principles for metrics. In the management accounting literature, metrics have

Table 2. The taxonomy of measurement methods in which the examples from Table 1 have been positioned

	Qualitative	Semi-quantitative	Quantitative
Subjective	**Assessment** • project co-ordinator's assessment as to which non-technical factors may cause delay in project (e.g. non-availability of equipment, lack of staff) *(Unilever)* • superior's assessment of degree of achievement of previously agreed departmental, personal growth and customer defined targets (*Olin Corporation*)	**Rating with non-sophisticated conversion** • project co-ordinator's rating of confidence in technical success (= meeting the predefined targets) (*Unilever*)	**Financial estimation** • team's estimation of the number of months after Manufacturing Release till project breaks even (*Hewlett Packard*) • project co-ordinator's estimation of cost till completion or next review *(Unilever)*
		Rating with sophisticated conversion	**Numerical estimation** • section leader's, or team's, or project leader's estimation of the number of months till completion of projects stages remaining (*Hewlett Packard*) (*Unilever*)
Semi-objective customers' measurement	**Assessment** • customers' assessment how good previously expressed expectations have been met (*Olin Corporation*)	**Rating with non-sophisticated conversion** • customers' (?) rating per project, whether actual project achievements are in line with predefined targets; aggregation using projects' dollar value as weightings (*Borg-Warner*)	**Financial estimation** • customers' (?) estimation of NPV of research projects completed, assuming 100% market share (*Borg-Warner*) • customers' estimation of R&D's share' of cash flows realized by products introduced in the market over the past *x* years (Hitachi) (*General Electric*)
		Rating with sophisticated conversion	**Numerical estimation** • internal customers' estimation of % completed of initially agreed research objectives (*GM Hughes Electronics*)
Semi-objective non-involved persons' measurement	**Assessment** • non-involved peers/ management assessment of several technical, business and organization factors of running programs (3M) • external expert panel assessment of research accomplishments (*GM Hughes Electronics*)	**Rating with non-sophisticated conversion** • (non-involved peers/ management's rating of technical, business and organization factors of running programs (*3M*) • non-involved staff rating of degree of meeting budgetary and other objectives (*GM Hughes Electronics*)	**Financial estimation**
		Rating with sophisticated conversion	**Numerical estimation**
Objective			**Financial computation** • computation of sales and profit generated by BUs which exist thanks to an R&D invention; computation of licensing income per year (*General Electric*) • computation of cumulative profits /investments (*Hewlett Packard*) **Numerical computation** • computation of number of months per completed project stage (*Hewlett Packard*) • computation of number of patents/million $ invested in R&D (*General Electric*)

traditionally been divided according to their *format* into a *financial* and a *non-financial* cluster (see, for example, Anthony and Gonvindarajan 1998, 461–462; Kaplan and Atkinson 1998, 442; Vaivio 1999). The problem with this dichotomy is that the cluster of non-financial indicators is still very large, as it contains the output from all measurement methods except the quantitative financial ones. Especially in R&D where, owing to the fact that most financial indicators are lagging and difficult to obtain (see the section on the R&D Performance Measurement Problem), people often have to rely on non-financial methods, this dichotomy is rather pointless. Therefore, distinguishing between the measurement method and the metric, and elaborating different categories of (non-financial) measurement methods – as was done in the previous section – seems to offer better opportunities for formulating guidelines for measurement system design than solely using the dichotomy of financial and non-financial metrics. Therefore, we will not use this dichotomy as a criterion in categorizing metrics.

Another sorting principle that deals with the *format* of a metric is whether the metric is a *composite* or a 'single item' metric. According to Smith (1976, 747), there are essentially three bases for combining 'single items' into a composite summary measure:

- statistical, based on intercorrelations of measures;
- economic, based on reducing organizational goals and behaviour to monetary scales;
- judgemental, either by direct assessment by the decision-makers of the relative importance of various basic items or through derivation of weighting from decisions made in the past.

We can conclude that we have already captured this sorting criterion in our taxonomy of measurement methods and thus we will not use this criterion again to sort R&D metrics.

A third, often-used classification of metric *formats* that should be commented upon classifies metrics according to the *type of scale* used to assign values. Usually the following scales are discerned, in order of increasing restrictiveness of admissible transformations that can be made to the scale without changing its structure (Park *et al.* 1996; Stevens 1959):

- nominal
- ordinal
- interval
- ratio
- absolute.

The type of scale has important implications for the statistical operations that are permissible. Going from the nominal to the absolute scale, the number of permissible statistics increases. Although this is an important aspect of a metric, we will not use it here since, for our classification, we are more interested in the *content* of a metric rather than its *format*.

In this respect, an interesting classification of indicators has been proposed in the organizational psychology literature by Landy and Farr (1983, 4–5), based on Smith (1976, 748–751). Their classification is built on three dimensions in a cubic model:

- the time span covered by the metric;
- specificity versus generality; and
- closeness to organizational goals.

The first dimension of Smith's model, the 'time span covered', considers whether a metric refers to things that have just happened or whether it covers a longer period of time. An example of the first type might be 'the number of hours spent on redesign in the last month', and of the latter 'the *average* number of hours rework per month *over the last 6 months*'. The second dimension along which metrics may vary is 'specificity versus generality'. A metric may refer to a specific performance aspect or it may give an overall impression of performance. This criterion has similarities with the composite-single item criterion discussed earlier, but whereas the latter deals with the *format* of the metric, the specificity–generality dimension concerns its *content*. For example, if a manager gives his overall assessment of the performance of a development team on a five-point scale, the content of the metric is general but, technically, it is not a composite metric because the underlying items have not been addressed separately and subsequently converted into a single value. The third dimension, 'closeness to organizational goals', is considered to be the most important and has therefore been further expanded. Whereas for the other dimensions only the extremes have been mentioned, this third dimension is explicitly split into three sections:

- behaviours
- results
- organizational effectiveness.

The underlying idea seems to be a *causal chain* of reasoning that (aspects of) individual behaviour lead to certain results, which in turn have an impact on the organization's effectiveness in terms of its stated goals. In the next subsection, we will discuss the three dimensions with respect to R&D environments and relate them to ideas proposed in the management accounting and R&D management literature.

Adaptation of the Metrics 'Content' Taxonomy to R&D Environments

For the design of measurement systems in R&D environments, the time span covered dimension seems to be important since (intermediate) outputs may not come in a continuous flow owing to the project size and degree of uncertainty involved in some types of R&D work. This could result in large fluctuations in a metric report if only 'short period' metrics were used. In this case, a metric averaging the score over several months might be more appropriate. We may thus conclude that, in principle, this dimension has a use in our R&D metrics taxonomy. However, the discerned two categories do not seem to be sufficient as they only concern measurement of performance in retrospect. In R&D environments, *prospective* metrics seem to be at least as important for various decisions as *retrospective* ones (Lander *et al.* 1995; Tipping *et al.* 1995). For example, for a project leader of a product development project team, the team members' estimations of the 'time to completion' of their specific assignments will be a very important metric in decision making as to whether corrective actions are necessary. Another example is the expected 'net present value' of each current project, which is an important metric for the Program Board in deciding whether changes should be made in the project portfolio. As can be seen from these examples, prospective measures can again be divided according to whether they cover a short time period or a long one. Thus, we have distinguished four different clusters along the 'time span covered' dimension of our metrics taxonomy (see Figure 3):

- metrics addressing performance in the distant past;
- metrics addressing performance in the recent past;
- metrics addressing expected performance in the near future;
- metrics addressing expected performance in the far future.

Note that the categorizing of metrics according to this criterion is dependent upon personal opinion as to what should be considered short and long periods.

The closeness of a measurement topic to organizational goals seems to be a useful distinction in R&D environments as well, though again we will propose some adaptations. The underlying causal chain approach, as well as the three distinguished categories, seems to have similarities with work presented in management accounting and R&D management literature. One of the first authors to discuss the concept of causal chains of metrics in the management accounting literature was Haynes (1969, 158–160). In his overview of business forecasting techniques, Haynes discussed a relatively new approach, which made a distinction between leading, coincident and lagging indicators. In this approach, leading indicators were supposed to reflect 'commitments for the future'. This idea of causal chains of performance indicators has clearly been picked up by Kaplan and Norton (1996, 28) with their 'balanced scorecard' concept as well as by several other authors (e.g. Lynch and Cross 1995; Hardjono *et al.* 1996). Whereas the financial perspective of a balanced scorecard contains the traditional lagging financial indicators, the customer and business perspective clusters provide more immediate, coincident indicators. Finally, the learning and growth perspective addresses the infrastructure that the company must build to create long-term growth and improvement, which has similarities with Haynes' leading indicators concept. According to Kaplan and Norton, this infrastructure contains three sources for improvement and growth: people, systems and organizational procedures. If we compare Haynes' and Kaplan and Norton's concepts with those of Landy and Farr discussed in the previous subsection, we can see the similarities between the financial and business effectiveness clusters. Landy and Farr's behaviour and results clusters are somewhat more difficult to equate as, for example, Kaplan and Norton's internal business perspective may contain both behaviour and results metrics. In Landy and Farr's clustering, the leading infrastructure indicators seem to be missing.

Brown and Svenson's (1988) R&D process framework referred to in the 'Background' section shows similarities with each of the classifications discussed above. In line with Landy and Farr, a clear distinction is made between organizational effectiveness (outcomes), results (outputs) and behaviours (in-

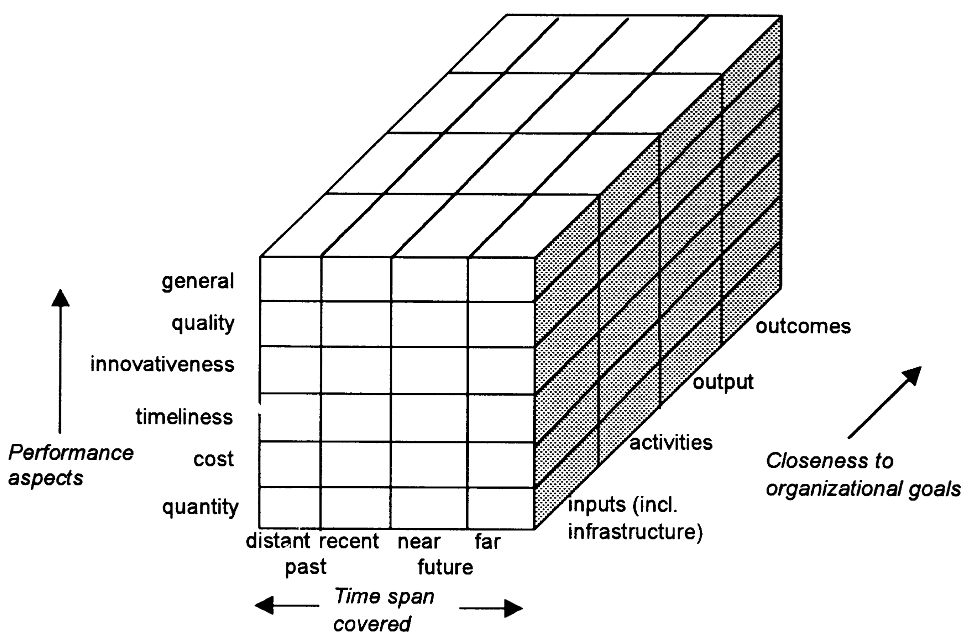

Figure 3. A taxonomy of R&D metrics.

Source: Kerssens-van Drongelen (1999).

process measures). Brown and Svenson's 'inputs' partly consist of the infrastructure elements distinguished by Kaplan and Norton.

Combining the three strands of literature, we can conclude that although the basic causal chain concept is similar in all the approaches, the identification of clusters differs slightly. The classification proposed by Brown and Svenson seems to be the most complete. The following categories are therefore distinguished (in order of increasing closeness to organizational goals, see Figure 3):

- inputs (including infrastructure)
- activities
- outputs
- outcomes.

Finally the usefulness of the specificity–generality distinction has to be discussed. Landy and Farr, as well as Smith, do not elaborate much on this dimension, except briefly discussing what is meant by a specific and a general metric. It can be argued that it is useful to go one step beyond the *degree* of specificity and to search for taxonomies that actually categorize *which* performance aspects are generally discerned in R&D environments. In the literature, we found a few proposals for such clusterings. For example Clark and Fujimoto (1991) and Wheelwright and Clark (1992, 47) have stated that time-to-market, quality and productivity together define R&D performance. Gerritsma and Omta (1998) have expanded on this taxonomy by arguing that productivity should be split into two concepts: productivity in terms of quantity and productivity in terms of cost. These four categories are supported by Brown and Gobeli (1992), who commented that four general attributes were cited by several managers as important parameters of R&D performance: quality, quantity, timeliness and cost. Ahaus (1994, 146) also mentioned that for the various flows in a company (money, information, material), these four items can be considered as the general set of performance aspects. Based on a historical analysis of the drivers of competition over recent decades, Bolwijn and Kumpe (1990) came up with largely similar aspects as important performance criteria for R&D. However, they also built a case that besides efficiency, quality and flexibility (for R&D especially meaning speed), also innovativeness should be identified as an important R&D performance aspect because product uniqueness seems to be the new emerging market demand. Their emphasis on innovativeness as an important performance aspect for the future seems to be supported by several others (see, for example, Moss Kanter *et al.* 1997; Zien and Buckler 1997). For performance appraisal of *individual researchers*, innovativeness – and related concepts such as creativity and serendipity – has already been acknowledged as an important performance aspect for a long time (see, for example, Keller and Holland 1982; Quinn 1960; Stahl and Steger 1977). Overall, we thus arrive at the following five clusters of performance aspects:

- quality
- innovativeness
- timeliness
- cost
- quantity.

In addition to these specific performance aspects, we should of course also maintain a cluster 'general performance', which refers to the intertwined performance on two or more of the performance aspects listed.

Positioning metrics used in measurement procedures reported in literature in the metrics taxonomy. In Figure 3, we illustrate the metrics taxonomy that emerges when we combine the three dimensions and their subdivisions discussed in the previous subsection. Given that, unlike other publications in R&D management and performance measurement literature, we combine in this taxonomy several metric content characteristics through elaboration of the Landy and Farr framework, we believe this

Table 3. Metrics applied in the procedures described in Table 1

Company	Procedure	Metric description	Time span covered	Closeness to goals	Aspect
3M (Krogh *et al.* 1988)	• Technical Audit System *(program part)*	• Overall technology strength	recent past	inputs	quantity
		• Number of personnel	near future	output	innovativeness
		• Personnel skills	recent past	inputs	quality
		• Knowledge of competition	recent past	inputs	quality
		• 3M product performance	distant past	output	quality
		• Remaining R&D investment	near future	activities	cost
		• Manufacturing implementation	near future	output	quality
		• Probability of technical success	near future	activities	quality
		• Programme's financial potential	far future	outcomes	general
		• Competitive position	near future	inputs	general
		• Probability of marketing success	near future	output	quality
	• Technical Audit System *(laboratory part)*	• Strategy focus, clarity of goals	recent past	inputs	quality
		• Number of staff	recent past	inputs	quantity
		• Staff skills	recent past	inputs	quality
		• Balance between product maintenance, related and unrelated new product efforts	recent past	output	innovativeness
		• Amount of co-ordination/interaction with other internal departments	recent past	activities	quantity
Borg-Warner (Collier 1977)	• Project Goals Achievement Rating	• Rating per project whether actual project achievements are in line with predefined targets	recent past	activities	general
	• Business Opportunity Evaluation	• NPV of research projects completed, assuming 100% market share	far future	outcomes	general
General Electric (Robb 1991)	R&D's impact on Business Score, a combination of:				
	• the 'Jimmy Steward' Test	• Sales and profit generated by BUs which exist thanks to an R&D invention	distant past	outcomes	general
	• Direct Output Counting	• Licensing income per year	distant past	outcomes	general
		• Number of patents/million $ invested in R&D	distant past	outputs	quantity
	• Valuation of Technology Transitions	• R&D's share of realized cash flows	distant past	outcomes	general
		• NPV of future cash flows from products/processes introduced in the past decade	far future	outcomes	general
		• Overall rating of R&D's impact on business	distant past	outcomes	general

GM Hughes Electronics (Chester 1995)	• Laboratory Performance Measurement	• Assessment of technical state of the art of research accomplishments	distant past	output	innovativeness
		• % completed of initially agreed research objectives	recent past	activities	quantity
		• Rating of degree of meeting budgetary objectives	recent past	activities	cost
		• Overall performance score	recent past	activities	general
Hitachi (Kuwahara and Takeda 1990)	• R&D cost effectiveness measurement	• R&D's 'share' of realized profits from products/processes introduced in the market in the past 5 years, divided by related R&D cost	distant past	outcomes	general
Hewlett Packard (House and Price 1991)	• Return map	• Number of months per completed project stage	recent past	activities	timeliness
		• Number of months till Manufacturing Release	near future	activities	timeliness
		• Number of months after Manufacturing Release till project breaks even	near future	outcomes	general
		• Cumulative profits/investments	distant past	outcomes	general
Olin Corporation (Endres 1997)	• Performance Management Cycle	• Degree of achievement of previously agreed upon departmental, personal growth and customer defined targets	recent past	outputs	general
		• Assessment as to how well previously expressed customer expectations have been met	recent past	outputs	quality
Unilever (Rainbow 1971)	• Progress Assessment Procedure	• Time till completion of predefined targets	near future	activities	timeliness
		• Cost till completion or till next review	near future	activities	cost
		• Confidence in technical success (=meeting the predefined targets)	near future	activities	quality
		• Non-technical factors which may cause delay in project (non-availability of equipment, lack of staff)	near future	inputs	timeliness

taxonomy to be more complete than those presented up to now. In the framework, we can position the numerous R&D metrics that have been proposed in literature over the past decades. However, it should be noted that only a few authors have empirically validated the appropriateness of the metrics proposed in their publications for specific application purposes and in particular contexts. In addition, for some other metrics, such as for instance the ones used in the measurement procedures presented in Table 1, anecdotal evidence exists that these metrics are useful in a particular context and application. In most literature, however, the authors only *argue* the usefulness of certain metrics or mention the metrics as an example without further reference to evidence of applicability.

In Table 3, we have assigned the metrics reported to be used in the measurement procedures presented in Table 1 to the cluster we think they fit best, acknowledging that some assignments are arguable. From this table, we can deduce that along the three axes, each of the identified clusters has been addressed more than once. Thus, we may conclude that the three dimensions are not overly detailed. However, turning to the 96 possible *combinations* of the three dimensions, only about 25% have been covered. After supplementing the metrics listed in Table 3 with many more R&D metrics found in the literature and in her empirical research, Kerssens-van Drongelen (1999) concluded that still several clusters remained empty, particularly the 'far future' ones. It remains an area for further research to determine which of these empty clusters are actually irrelevant and which clusters need further expansion.

Purposes of R&D Performance Measurement

In the previous section, we described two taxonomies, one of measurement methods and one of R&D metrics, which together enable a *technical* characterization of R&D performance measurement approaches. However, performance measurement procedures can also be characterized according to their *purposes*: whose performance is measured by the procedure (the measurement *subject*) and what is done with it (the measurement procedure's *function*). In this section we will explore these two topics.

Subjects of Performance Measurement

In the R&D management literature, we did not find many overview articles addressing the range of R&D measurement subjects, let alone presenting a systematic list of these subjects. Two exceptions are Ranftl (1978) and Griffin and Page (1993). Ranftl, reporting on a study into R&D productivity measurement, noted that 'the organisational level at which the evaluation is made must be determined for each situation' (p. 40). He subsequently gave five different 'organizational levels' at which R&D work can be evaluated:

- an individual
- a group
- a major design
- an entire program
- a larger organizational segment.

Ranftl has also listed various examples of performance indicators that were deemed by his respondents to be useful in the measurement of individual and organizational segment productivity.

Griffin and Page have identified three subjects for new product development success measurement:

- a single product
- a program
- the total firm.

Based on group consensus and factor analysis techniques, they conclude that these are distinct subjects of measurement with corresponding different metrics.

Most authors focus on a single subject of measurement, for example:

- NPD programs or portfolios (Cooper and Kleinschmidt 1995; Hardingham 1970; Loch *et al.* 1996; Tipping *et al*. 1995);
- a single product development project (Hultink and Robben 1995; Rainbow 1971; Sivathanu Pillai and Srinivasa Rao 1996);
- a product development team (House and Price 1991; Meyer 1994);
- R&D (sub)departments (Chester 1995; Collier 1977; Kuwahara and Takeda 1990; Patterson 1983; Robb 1991; Szakonyi 1994a,b);
- Individual researchers (Domsch *et al*. 1983; Edwards and McCarrey 1973; Keller and Holland 1982).

Noticeable in these lists is the heterogeneity of subjects: there are subjects that consist (partly) of *human beings* (an individual, a group/team, an organizational segment/ department or a firm) and another group of subjects that are the *results* of the work of these people (a major design, a project or an entire program). Normally, one or more persons are made responsible for achieving results within a specific context, thus characteristics of the resulting object, and of the process followed to achieve the results, indicate the performance of these people. Therefore, within the framework of performance measurement for performance control, we believe it is useful and possible to identify subjects of measurement as *people*, or *systems* of people and resources, having to fulfil specific *organizational objectives*. Thus, instead of naming a 'major design' or a 'program' as the subject of measurement, we label the team or researcher responsible for achieving this design or program as the subject of measurement. Furthermore, we distinguish in the lists of subjects presented above two views on R&D: a cross-functional *process view* and a *hierarchical departments view*. Both views can be further analysed to identify the typical subjects of measurement within that view.

From the R&D process view, we can identify the following subjects on which measurement can be focused:

- *the Program Board, Business Unit R&D Portfolio Committee, Platform management*, etc., all having as a common characteristic that they are responsible for *a set* of projects which have to be mutually balanced and co-ordinated;
- *a project team* responsible for accomplishing a single project; possibly divided again in
- *a sub-team* responsible for accomplishing specific parts of a project.

The R&D department or disciplinary performance can also be measured at several levels:

- the *entire R&D department* (the R&D organization at the highest aggregation level);
- the different *R&D sites*, which together make up the entire R&D department;
- a *(sub)department within a site*, e.g. a technical discipline or product group.

Individual researchers could be positioned under both the R&D department and the R&D process view, as they are basic elements in both views (see Figure 4). At the *firm level*, we should look at the *combined* efforts of the R&D department in effectively and efficiently processing technological knowledge, and of all those involved in the cross-functional R&D process in effectively and efficiently creating new products and processes by using this technological and other knowledge (e.g. marketing inputs). Thus, this subject is paramount to both views and includes all people within (and outside) the organization somehow involved in either of these processes (see Figure 4).

With this taxonomy, we can characterize the measurement procedures listed in Table 1 with respect to the subjects they apply to. This is done in Table 4.

Functions of Performance Measurement

Examples of measurement system functions discussed in the literature. As indicated

IJMR
June 2000

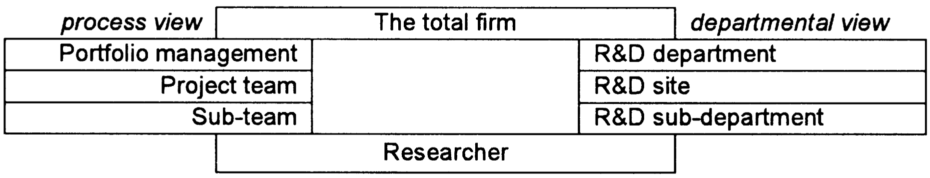

Figure 4. Taxonomy of R&D measurement subjects.

Adapted from Kerssens-van Drongelen (1999).

earlier, performance measurement systems can be used in different ways by decision-makers within the organization to improve or sustain organizational performance in order that the company's objectives will be achieved. In line with the terminology used in the design literature, we call these different uses the *functions* a PMS fulfils within the performance control process. In the literature, several authors have given examples of such measurement system functions, also referred to as 'reasons for measuring' or 'measurement purposes'. For example, Jorissen (1994, 9) indicated that a PMS may *provide insights* to managers whether, and if so which, corrective measures should be taken. In addition, it may help *clarify* general management's organizational objectives to the lower organizational levels. Pritchard (1990) has discussed how performance measurement systems can be built and used as a *motivational tool*. Park *et al.* (1996, 3) have mentioned the following reasons for measuring performance:

- to *characterize*, to gain understanding of processes, products, resources and environments and to establish baselines for comparisons with future assessments;
- to *evaluate*, to determine status with respect to plan;
- to *predict* and thus enable planning;
- to support *improvement* by gathering information that helps to identify problems and by planning and tracking improvement efforts.

Landy and Farr (1983, 3–4), who focus on the measurement of individual performance, come up with three purposes of measurement:

- *administrative purposes*, including determining promotions and demotions, merit payments, lateral transfers, training program assignments, etc.;
- *guidance and counselling purposes*, including supervisory feedback and career planning aimed at improving job satisfaction and work motivation through

Table 4. Subjects focused on in the measurement procedures listed in Table 1

Company	Subject focus
3M (Krogh *et al.* 1988)	• Portfolio management • R&D site
Borg-Warner (Collier 1977)	• R&D department • R&D sub-departments
General Electric (Robb 1991)	• R&D department
GM Hughes Electronics (Chester 1995)	• R&D site
Hitachi (Kuwahara and Takeda 1990)	• R&D department • R&D site
Hewlett Packard (House and Price 1991)	• Project team
Olin Corporation (Endres 1997)	• Researcher
Unilever (Rainbow 1971)	• Project team

providing information on current performance and probable future assignments in the organization;
- *research purposes*, such as validation of selection procedures, evaluation of training programs, compensation plans or job enrichment programs.

Finally, Bonsdorff and Andersin (1995) provided the following summary list of measurement system uses, based on a review of several authors:

- to *motivate* employees;
- to *show employees how they contribute* to the organization's performance;
- to *communicate* performance expectations;
- to provide *management information;*
- to identify *performance gaps;*
- to support decision making.

A taxonomy of measurement system functions. Although there are some recurring themes in the examples referred to in the previous subsection, none of the cited authors seems to give a complete list of measurement system functions. Furthermore, the examples lack a general theoretical framework from which they are derived. This shortcoming was recognized in Kerssens-van Drongelen (1999, 39–50) and subsequently addressed by a research project in which a taxonomy of PMS functions has been developed systematically. Theoretically, this taxonomy is based on three complementary, generally accepted strands of the axiomatic systems theory:

- the *cybernetic systems control theory* which, since Anthony's (1965) publication, is at the basis of most responsibility accounting research;
- the *agency theory*, based on the work of, amongst others, Alchian and Demsetz (1972) and Jensen and Meckling (1976) on which a lot of management accounting and remuneration research is based;
- *empowerment* theories, such as the modern sociotechnical systems design theory (de Sitter 1981; 1989; van der Zwaan 1994) and ProMES (Pritchard 1990), which are often used in human behaviour and organization design research.

By analysing the different roles performance measurement plays, according to these three theories, in performance control, Kerssens-van Drongelen (1999) deduced the following list of measurement system functions. A PMS could:

(1) *provide insight* into deviations of actual performance from objectives, in order to support the *management* in diagnosing whether, and if so which, steering measures should be applied;
(2) fuel *learning* as to how the system that has to be controlled works (in other words, improving the conceptual model of this system's functioning), which enables better planning and control in the future;
(3) facilitate *alignment and communication* of objectives;
(4) support decision making about performance based rewards;
(5) *provide insight* into deviation of actual performance from objectives, in order to support the *staff themselves* in diagnosing whether, and if so which, steering measures should be applied;
(6) provide inputs for *justification* of the existence, decisions and performance;
(7) support *motivating of people through feedback.*

In Figure 5, the systems control model is presented with the *numbered arrows* indicating information flows that are most important for the measurement system function indicated by that number.

The empirical relevance of this taxonomy was established in an experiment involving 14 quality managers in a large Dutch multinational (see Kerssens-van Drongelen 1999, 102–103, 109–112, 150–154), and it was shown that all the examples of measurement system functions discussed in the previous

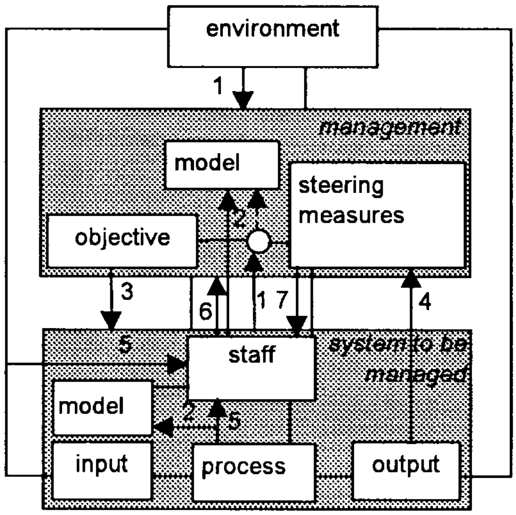

Figure 5. The dominant information flows for each of the seven measurement system functions. Source: Kerssens-van Drongelen (1999).

subsection fit in the developed taxonomy (Kerssens-van Drongelen 1999, 4749). Thus, the developed taxonomy seems to be complete.

Subsequently, Kerssens-van Drongelen established, again based on a literature review and empirical research, that each of the seven PMS functions required different technical PMS characteristics (i.e. different types of metrics and measurement methods). This finding is also illustrated when we identify, based on statements made in the articles about their uses and benefits, the functions of R&D measurement procedures presented in Table 1 (see Table 5). If we look back at Tables 2 and 3, we note that, although at a detailed level each measurement procedure is unique, the procedures aimed at the same PMS function largely use the same *types* of metrics and measurement methods. Such classifications of measurement approaches may facilitate the evaluation by PMS designers whether a particular approach is also suitable for their situation.

Discussion and Conclusions

In this paper, we have reviewed 40 years of literature on R&D performance measurement in a historical, a technical (metrics, measurement methods) and an application (subjects, functions) perspective. More precisely, we have covered the literature on performance measurement in *industrial* R&D; for a review of the literature on evaluation of technology programs in the *public sector*, we refer to Georghiou and Roessner (2000). The major value of such a literature review is that it provides readers with references that explore in greater depth the different issues and dimensions of the subject. It is arguable, however, whether a review should aim at impartiality or take a critical stance in relation to the literature (Butler 1999). At the risk of excessive criticality and of losing impartiality, we set out below our perceptions of some of the limitations of the literature on R&D performance measurement.

One limitation, which is perhaps hardly surprising in view of the fact that this literature is dominated by scientists and engineers, is that it mainly concentrates on the *technical dimensions* of measurement that can be analysed and quantified; there is a relative neglect of the more behavioural, qualitative dimension of PMS *application* in R&D. Although Kerssens-van Drongelen (1999) has found evidence that each of the PMS functions discussed in the previous section requires other PMS characteristics, only a few authors actually discuss for which functions the measurement approaches they present have been found useful, or are expected to be useful. Furthermore, related concepts such as the 'style of evaluation' (Hopwood 1972), 'culture' (Dent 1991; Schein 1991), 'strategic management styles' (Goold and Campbell 1987), the 'cognitive styles' and 'strategic schemata' of decision-makers (Schwenk 1989), 'belief systems' and 'views on risks to be avoided' (Simons 1995), 'organizational climate' (Keen and Scott Morton 1978), 'functional background of top management' (Scherer 1992) and 'institutional' and 'managerial short-termism' (Demirag 1995; Marsh 1990) can all influence the selection and use of performance

Table 5. Functions of the measurement procedures listed in Table 1

Company	Uses mentioned in the publications	Corresponding measurement system functions
3M (Krogh *et al.* 1988)	• Evaluation of current status of laboratory programs • Learning via trend analysis of database about key factors that help improve program selection, prioritization and execution	• Providing insight to support diagnosis by the staff themselves • Fuelling learning
Borg-Warner (Collier 1977)	• Performance comparison between research groups • Identification of improvement opportunities • One of the factors in determining annual bonuses of department heads	• Providing insight to support diagnosis by management • Supporting decision making about rewards
General Electric (Robb 1991)	• Justification of the existence of the corporate R&D lab • Identification of improvement opportunities	• Justification • Fuelling learning
GM Hughes Electronics (Chester 1995)	• Determine laboratory results-based compensation (pay-out % is the same for all laboratory personnel)	• Supporting decision making about rewards
Hitachi (Kuwahara and Takeda 1990)	• Insight into whether corrective actions by the sponsors are necessary • Support management in strategic planning • Learning	• Providing insight to support diagnosis by management • Fuelling learning
Hewlett Packard (House and Price 1991)	• Monitoring a NPD project • Prompt for learning and improvement • Assess impact of decisions on entire project • Encourage collaboration among different functions	• Providing insight to support diagnosis by the staff themselves • Fuelling learning
Olin Corporation (Endres 1997)	• Improve linkage between individual and department objectives • Professional development • Basis for determination of compensation	• Alignment and communication of objectives • Motivating through feedback • Supporting decision making about rewards
Unilever (Rainbow 1971)	• To generate insight whether corrective actions by research management are necessary • Learning	• Providing insight to support diagnosis by management • Fuelling learning

measures, but they are barely discussed in the R&D performance measurement literature.

This latter omission also points to another flaw in R&D performance measurement literature: there is hardly any *systematic, empirical research* into the consistency among R&D measurement approaches and *contingencies*. For example, although several authors (Griffin and Page 1996; Hauser and Zettelmeyer 1997; Kerssens-van Drongelen 1999; Pappas and Remer 1985; Werner and Souder 1997a) suggest that a PMS should match the type of R&D performed, only a few actually try to support this statement with empirical evidence; but owing to the fact that they use different terminology, these studies are difficult to compare. Other contingencies that have received some serious attention are the innovation strategy (Griffin and Page 1996; Hultink and Robben 1995) and the country where an R&D laboratory is established (Hart *et al.* 1998; Werner and Souder 1997b) but the results of these studies are inconclusive. Awareness of the need for

IJMR
June 2000

contingent PMS design is growing, but it seems that we are only at the beginning of an understanding of this topic in R&D environments.

A third limitation of the literature on R&D evaluation is that it fails to reflect adequately the current trends in R&D management and practice that we set out at the beginning of this review. A big step forward has been taken by the various approaches that link R&D to corporate strategy, so that R&D is now much more a strategic issue in its own right (Amidon Rogers 1996; Roussel *et al*. 1991). However, strategy implementation requires the integration and support of many other organizational components, for example, structure, systems (including the reward system), management style, skills and shared values (Waterman *et al*. 1980).

Yet the R&D performance measurement systems that are reported in the literature and at conferences are more consistent with a linear, relay-team ('over-the-partition') approach to new product and process development than with multidisciplinary teams operating in parallel. For example, it is difficult to reconcile the narrow, functional focus of the 14 classes of R&D measurement methods identified by EIRMA (1995) with the emphasis placed in current NPD literature on both internal collaboration among functions and external collaboration with customers, suppliers, research agencies and even competitors (Cauley de la Sierra 1995; Wheelwright and Clark 1992).

Failure to take adequate account of the R&D management process can easily result in a measurement system that contributes to 'segmentalism' (Moss Kanter 1983, 28) rather than the integration of cross-functional teams that is required for rapid, successful product development. The close collaboration that is required for teams to function effectively can easily be eroded by an emphasis on individual contributions; yet such a focus is not uncommon. Although measures that relate to specific functions (e.g. the number of patents or publications obtained) may make responsibility clearer and increase the value of the measures, the narrow, output goals set are frequently of little worth to the project team or the organization. There is probably no easy solution to this difficulty. The very clinical approach to R&D measurement of both academic researchers and consultants leaves a great deal of scope to develop R&D measures that more directly address the wider organizational and behavioural context of R&D evaluation.

References

Ahaus, C.T.B. (1994). Bevoegdheidsverdeling en Organisatie. Thesis, University of Groningen (in Dutch).

Alchian, A.A. and Demsetz, H. (1972). Production, information cost and economic organization. *American Economic Review*, **62**, 777–795.

Allio, R.J. and Sheehan, D. (1984). Allocating R&D resources effectively. *Research Management*, **27**(4), 14–20.

Amidon Rogers, D.M. (1996). The challenge of fifth generation R&D. *Research-Technology Management*, **39**(4), 33–41.

Anthony, R.N. (1965). *Planning and Control Systems: A Framework For Analysis*. Boston, MA: Division of Research, Graduate School of Business Administration, Harvard University.

Anthony, R.N. and Govindarajan, V. (1998). *Management Control Systems*, 9th edition. Homewood, IL: Irwin/McGraw-Hill.

Athaide, G.A. and Stump, R.L. (1999). A taxonomy of relationship approaches during product development in technology-based, industrial markets. *Journal of Product Innovation Management*, **16**(5), 469–482.

Azzone, G. and Masella, C. (1992). Measuring performances of product development teams: a long term view. In *Proceedings of the International Product Development Management Conference on New Approaches to Development and Engineering*, 18–19 May, Brussels, pp. 15–25.

Barham, K. and Heimer, C. (1998). *ABB The Dancing Giant: Creating the Globally Connected Corporation*. London: Financial Times/Pitman Publishing.

Bell, D.C. and Read, A.W. (1970). The application of a research portfolio selection method. *R&D Management*, **1**(1), 35–42.

Bolwijn, P.T. and Kumpe, T. (1990). Manufacturing in the 1990s productivity, flexibility and innovation. *Long Range Planning*, **23**, 44–57.

Bonsdorff, C. von and Andersin, H.E. (1995). Supporting the business process management paradigm by means of performance measurements. In *Proceedings of CE95 Conference, 'Concurrent Engineering: a Global Perspective'*, August, McLean, VA, pp. 529–539.

Bradley, S.P., Hausman, J.A. and Nolan, R.L. (eds) (1993). *Globalization, Technology, and Competition: The Fusion of Computers and Telecommunications in the 1990s*. Boston, MA: Harvard Business School Press.

Brown, M.G. and Svenson, R.A. (1988). Measuring R&D productivity. *Research-Technology Management*, **31**(4), 11–15.

Brown, W.B. and Gobeli, D. (1992). Observations on the measurement of R&D productivity: a case study. *IEEE Transactions on Engineering Management*, **39**(4), 325–331.

Butler, R. (1999). Review: The International Journal of Management Reviews. *BAM News, British Academy of Management*, Autumn, 18.

Cauley de la Sierra, M. (1995). *Managing Global Alliances: Key Steps for Successful Collaboration*. Wokingham: Addison-Wesley Publishers Ltd, Addison-Wesley Publishing Co. Inc. and the Economist Intelligence Unit (EIU).

Chester, A.N. (1995). Measurements and incentives for central research. *Research-Technology Management*, **38**(4), 14–22.

Chiesa, V. (1996). Strategies for global R&D. *Research-Technology Management*, **39**(5), 19–25.

Chiesa, V. and Manzini, R. (1998). Organizing for technological collaborations: a managerial perspective. *R&D Management*, **28**(3), 199–212.

Chiesa, V., Coughlan, P. and Voss, C.A. (1996). Development of a technical innovation audit. *Journal of Product Innovation Management*, **13**(2), 105–136.

Clark, K.B. and Fujimoto, T. (1991). *Product Development Performance*. Boston, MA: Harvard Business School Press.

Collier, D.W. (1977). Measuring the performance of R&D departments. *Research Management*, **20**(2), 30–34.

Cook, L.G. (1966). How to make R&D more productive. *Harvard Business Review*, July–August, 145–153.

Cooper, R. (1995). *When Lean Enterprises Collide: Competing Through Confrontation*. Boston, MA: Harvard Business School Press.

Cooper, R.G. (1992). The NewProd System: the industry experience. *Journal of Product Innovation Management*, **9**(2), 113–127.

Cooper, R.G. (1998). *Product Leadership: Creating and Launching Superior New Products*. Reading, MA: Perseus Books.

Cooper, R.G. and Kleinschmidt, E.J. (1995). Benchmarking the firm's critical success factors in new product development. *Journal of Product Innovation Management*, **12**(5), 374–391.

Cooper, R.G., Edgett, S.J. and Kleinschmidt, E.J. (1998). Best practices for managing R&D portfolios. *Research-Technology Management*, **41**(4), 20–33.

Cordero, R. (1990). The measurement of innovation performance in the firm: an overview. *Research Policy*, **19**, 185–192.

Comstock, G.L. and Sjolseth, D.E. (1999). Aligning and prioritizing corporate R&D. *Research-Technology Management*, **42**(3), 19–25.

D'Aveni, R. (1995). *Hypercompetitive Rivalries: Competing in Highly Dynamic Environments*. New York: Free Press.

Demirag, I.S. (1995). Short-term performance pressures: is there a consensus view? *European Journal of Finance*, **VI**(1), 41–56.

Dent, J.F. (1991). Accounting and organisational cultures: a field study of the emergence of a new organisational reality. *Accounting, Organizations and Society*, **16**(8), 705–732.

Domsch, M., Gerpott, T. and Jochum, E. (1983). Peer assessment in industrial R&D departments. *R&D Management*, **13**(3), 143–154.

Dressler, R., Wood, R.S. and Alvarez, V. (1999). Evaluating R&D performance using the cost savings metric. *Research-Technology Management*, **42**(2), 13–14.

Edelheit, L.S. (1998). GE's R&D strategy: be vital. *Research-Technology Management*, **41**(2), 21–27.

Edwards, S.A. and McCarrey, M.W. (1973). Measuring the performance of researchers. *Research Management*, **16**(1), 34–41.

Ellis, L. (1997). *Evaluation of R&D Processes: Effectiveness Through Measurements*. Boston, MA: Artech House.

Endres, A. (1997). *Improving R&D Performance the Juran Way*. New York: John Wiley.

European Industrial Research Management Association (1995). *Evaluation of R&D Projects*. Paris: EIRMA.

Fortuin, L. (1994). Hoofdstuk III: operationele prestatiemeting: onmisbaar op de weg naar voortdurende verbetering. In Jorissen, A. (ed.), *Prestatiemeting, naar een betere beheersing van bedrijfsactiviteiten*. Antwerpen: MAKLU Uitgevers, pp. 139–177 (in Dutch).

Foster, T.M. (1996). Making R&D more effective at Westinghouse. *Research-Technology Management*, **39**(1), 31–37.

Foster, R.N., Linden, L.H., Whiteley, R.L. and Kantrow, A.M. (1985). Improving the return in R&D–I. *Research Management*, **28**(1), 12–17.

Francis, P., quoted in Schumann, P.A. Jr, Ransley, D.L. and Prestwood, D.C.L. (1995). Measuring R&D performance. *Research-Technology Management*. **38**(3), 45–54.

Galloway, E.C. (1971). Evaluating R&D performance keep it simple. *Research Management*, **14**(2), 50–58.

Gassmann, O. and Zedtwitz, M. von (1998). Towards the integrated R&D network new aspects of organizing international R&D. In Lefebvre, L.A., Mason, R.M. and Khalil, T. (eds), *Management of Technology, Sustainable Development and Eco-Efficiency*. Oxford: Pergamon, pp. 85–99.

Gear, A.E., Lockett, A.G. and Pearson, A.W. (1971). Analysis of some portfolio selection models for R&D. *IEEE Transactions on Engineering Management*, **18**(2), 66–76.

Georghiou, L. and Roessner, D. (2000). Evaluating technology programs: tools and methods. *Research Policy* (forthcoming).

Gerritsma, F. and Omta, S.W.F. (1998). The content methodology facilitating performance measurement by assessing the complexity of R&D projects. In Lefebvre, L.A., Mason. R.M. and Khalil, T. (eds), *Management of Technology, Sustainable Development and Eco-Efficiency*. Oxford: Pergamon, pp. 101–109.

Goold, M., and Campbell, A. (1987). *Strategies and Styles*. Oxford: Basil Blackwell.

Graaf, R. de (1996). Assessing product development: visualizing process and technology performance with race. Thesis, Technical University of Eindhoven, Eindhoven.

Griffin, A. (1997). PDMA research on new product development practices: updating trends and benchmarking best practices. *Journal of Product Innovation Management*, **14**(6), 429–459.

Griffin, A. and Page, A.L. (1993). An interim report on measuring product development success and failure. *Journal of Product Innovation Management*, **10**(4), 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**(5), 478–496.

Gupta, A.K. and Wilemon, D. (1996). Changing patterns in industrial R&D management. *Journal of Product Innovation Management*, **13**(6), 497–511.

Hardingham, R.P. (1970). A simple model approach to multi-project monitoring. *R&D Management*, **1**, (October), 43–47.

Hardjono, T.W., Have, S. ten and Have, W.D. ten (1996). *The European Way to Excellence*. DG III Industry, European Commission.

Hart, S., Hultink, E.J., Tzokas, N. and Commandeur, H. (1998). How industrial companies steer their new product development processes: empirical evidence from Dutch and UK firms. In *Proceedings of the 5th International Product Development Management Conference*, Como, Italy, pp. 503–521.

Hauser, J.R. and Zettelmeyer, F. (1997). Metrics to evaluate R,D&E. *Research-Technology Management*, **40**(4), 32–38.

Haynes, W.W. (1969). *Managerial Economics: Analysis and Cases*, revised edition. Austin, TX: Business Publications.

Hodge, M.H. Jr (1963). Rate your company's research productivity. *Harvard Business Review*, November–December, 109–122.

Hopwood, A.G. (1972). An empirical study of the role of accounting data in performance evaluation. *Empirical Research in Accounting, Supplement to Journal of Accounting Research*, **10**, 156–186.

House, C.H. and Price, R.L. (1991). The return map: tracking product teams. *Harvard Business Review*, **69**(1), 92–100.

House of Lords Select Committee on Science and Technology (HOL/SCST) (1990). *3rd Report, Definitions of R&D*. London: HMSO.

Hubert, J. (1970). R&D and the company's requirements. *R&D Management*, **1**(1), 30–34.

Hultink, E.J. and Robben, H.S.J. (1995). Measuring new product success: the difference that time perspective makes. *Journal of Product Innovation Management*, **12**(5), 392–405.

Hustad, T.P. (1996). Reviewing current practices in innovation management and a summary of selected best practices. In Rosenau, M.D. Jr (ed.), *The PDMA Handbook of New Product Development.* New York: John Wiley, pp. 489–510.

Iansiti, M. and West, J. (1997). Technology

integration: turning great research into great products. *Harvard Business Review*, **75**(3), 69–79.
Jensen, M.C. and Meckling, W.H. (1976). Theory of the firm: behavior, agency cost and ownership structure. *Journal of Financial Economics*, **3**, 305–360.
Jorissen, A. (ed.) (1994). *Prestatiemeting, naar een betere beheersing van bedrijfsactiviteiten*. Antwerpen: MAKLU Uitgevers (in Dutch).
Kaplan, R.S. and Atkinson, A.A. (1998). *Advanced Management Accounting*, 3rd international edition. Upper Saddle River, NJ: Prentice Hall.
Kaplan, R.S. and Norton, D.P. (1996). *The Balanced Scorecard: Translating Strategy Into Action*. Boston, MA: Harvard Business School Press.
Katzenback, J.R. and Smith, D.K. (1993). *The Wisdom of Teams: Creating the High-Performance Organization*. Boston, MA: Harvard Business School Press.
Keen, P.G.W. and Scott Morton, M.S. (1978). *Decision Support Systems*. Reading, MA: Addison-Wesley.
Keller, R.T. and Holland, W.E. (1982). The measurement of performance among research and development professional employees: a longitudinal analysis. *IEEE Transactions on Engineering Management*, **29**(2), 54–58.
Kerklaan, L.A.F.M., Kingma, J. and Kleef, F.P.J. van (1996). *De cockpit van de organisatie*. Deventer: Kluwer Bedrijfswetenschappen (in Dutch).
Kerssens-van Drongelen, I.C. (1999). Systematic design of R&D performance measurement systems. Thesis, University of Twente, Enschede, the Netherlands.
Kerssens-van Drongelen, I.C. and Bilderbeek, J. (1999). R&D performance measurement: more than choosing a set of metrics. *R&D Management*, **29**(1), 35–46.
Kerssens-van Drongelen, I.C. and Cook, A. (1997). Design principles for the development of measurement systems for research and development processes. *R&D Management*, **27**(4), 345–359.
Khurana, A. and Rosenthal, S.R. (1997). Integrating the fuzzy front end of new product development. *Sloan Management Review*, **38**(2), 103–120.
Klein, D.A. (ed.) (1998). *The Strategic Management of Intellectual Capital*. Boston, MA: Butterworth-Heinemann.
Krogh, L.C., Prager, J., Sorensen, D.P. and Tomlinson, J.D. (1988). How 3M evaluates its R&D Programs. *Research-Technology Management*, **31**(6), 10–14.
Kumar, V., Ganesh, J. and Echambadi, R. (1998). Cross national diffusion research: what we know and how certain are we. *Journal of Product Innovation Management*, **15**(3), 255–268.
Kumpe, T. and Bolwijn, P.T. (1994). Towards the innovative firm-challenge for R&D management. *Research-Technology Management*, **37**(1), 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), 134–138.
Lander, L., Matheson, D., Menke, M.M. and Ransley, D.L. (1995). Improving the R&D decision process. *Research-Technology Management*, **38**(1), 40–43.
Landy, F.J. and Farr, J.L. (1983). *The Measurement of Work Performance*. London: Academic Press.
Larson, C.F. (1998). Industrial R&D in 2008. *Research-Technology Management*, **41**(6), 19–24.
Lee, M., Son, B. and Lee, H. (1996). Measuring R&D effectiveness in Korean companies. *Research-Technology Management*, **39**(6), 28–31.
Leeuw, A.C.J. de (1982). *Organisaties: management, analyse, ontwerp en verandering: een systeemvisie*. Assen: Van Gorcum (in Dutch).
Lewis, R. (1998). Membership and management of a 'virtual' team: the perspectives of a research manager. *R&D Management*, **28**(1), 5–12.
Loch, C., Stein, L. and Terwiesch, C. (1996). Measuring development performance in the electronics industry. *Journal of Product Innovation Management*, **13**(1), 3–20.
Lynch, R.L. and Cross, K.F. (1995). *Measure Up! How to Measure Corporate Performance*, 2nd edition. Cambridge, MA: Blackwell.
Majchrzak, A. and Wang, Q. (1996). Breaking the functional mind-set in process organizations. *Harvard Business Review*, **74**(5), 92–99.
Mansfield, E. (1981). How economists see R&D. *Harvard Business Review*, November–December, 98–106.
Marsh, P. (1990). *Short-Termism on Trial*. London: The Institutional Fund Managers Association.
McDonough, E.F. III and Griffin, A. (1997). Matching the right organisational practices to a firm's innovation strategy. *Proceedings of the PDMA Conference, 'Maximising the Return on Product Development'*, October, Monterey, CA, pp. 55–86.
McGrath, M.E. (ed.) (1996). *Setting the Pace in Product Development: A Guide to Product and Cycle-Time Excellence*, revised edition. Boston, MA: Butterworth-Heinemann.

McGrath, M.E. and Romeri, M.N. (1994). The R&D effectiveness index: a metric for product development performance. *Journal of Product Innovation Management*, **11**, 213–220.
McNair, C.J. and Leibfried, K.H.J. (1992). *Benchmarking: A Tool for Continuous Improvement*. New York: Harper Business.
Mechlin, G.F. and Berg, D. (1980). Evaluating research ROI is not enough. *Harvard Business Review*, September–October, 93–99.
Meyer, C. (1994). How the right measures help teams to excel. *Harvard Business Review*, **72**(3), 95–103.
Morone, J.G. (1993). *Winning in High-Tech Markets: The Role of General Management: How Motorola, Corning, and General Electric Have Built Global Leadership Through Technology*. Boston, MA: Harvard Business School Press.
Moser, M.R. (1985). Measuring performance in R&D settings. *Research-Technology Management*, **28**(5), 31–34.
Moss Kanter, R. (1983). *The Change Masters: Corporate Entrepreneurs at Work*. London: Unwin Hyman.
Moss Kanter, R., Kao, J. and Wiersema, F. (1997). *Innovation: Breakthrough Thinking at 3M, DuPont, GE, Pfizer, and Rubbermaid*. New York: HarperCollins.
Nixon, B. (1995). The scope and measurement of research and development expenditure: views of UK company accountants. *Accounting Forum*, **19**(1), 29–46.
Nixon, B. (1998). Research and development performance measurement: a case study. *Management Accounting Research*, **9**(3), 329–355.
Organisation for Economic Co-operation and Development (1998). *Internationalisation of Industrial R&D: Patterns and Trends*. Paris: OECD.
Packer, M.B. (1983). Analyzing productivity in R&D organizations. *Research Management*, **26**(1), 13–20.
Pappas, R.A. and Remer, D.S. (1985). Measuring R&D productivity. *Research-Technology Management*, **28**(3), 15–22.
Park, R.E., Goethert, W.B. and Florac, W.A. (1996). *Goal Driven Software Measurement A Guidebook*. Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University.
Patterson, W.C. (1983). Evaluating R&D performance at the Alcoa laboratories. *Research Management*, **26**(2), 23–27.
Pittiglio, Rabin, Todd and McGrath (1995). *Product Development Leadership for Technology-Based Companies: Measurement and Management A Prelude to Action*. Weston, MA: Pittiglio, Rabin, Todd & McGrath.
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, 69–80.
Quinn, J.B., Baruch, J.J. and Zien, K.A. (1997). *Using Intellect and Software to Revolutionize Growth Strategies*. New York: Free Press.
Ranftl (1978). *R&D Productivity*, 2nd edition. Culver City, CA: Hughes Aircraft Company.
Rainbow, S.F. (1971). The use and development of a research project planning and progressing system. *R&D Management*, **1**(2), 101–106.
Reynolds, Wm.B. (1965). Research Evaluation. *Research Management*, **8**(2), 117–125.
Robb, W.L. (1991). How good is your research? *Research-Technology Management*, **34**(2), 16–21.
Rosegger, G. (1986). *The Economics of Production and Innovation: An Industrial Perspective*, 2nd edition. New York: Pergamon Press.
Rosenbloom, R.S. and Spencer, W.J. (1996). *Engines of Innovation: U.S. Industrial Research at the End of an Era*. Boston, MA: Harvard Business School Press.
Roussel, P.A., Saad, K.N. and Erickson, T.J. (1991). *Third Generation R&D: Managing the Link to Corporate Strategy*. Boston, MA: Harvard Business School Press.
Schainblatt, A.H. (1982). How companies measure the productivity of engineers and scientists. *Research-Technology Management*, **25**(3), 10–18.
Schein, E.H. (1991). *Organizational Culture and Leadership: A Dynamic View*. San Francisco, CA: Jossey-Bass.
Schwenk, C.R. (1989). Linking cognitive, organisational and political factors in explaining strategic change. *Journal of Management Studies*, **26**(2), 177–187.
Scherer, F.M. (1992). *International High-Technology Competition*. Cambridge, MA: Harvard University Press.
Schumann, P.A. Jr, Ransley, D.L. and Prestwood, D.C.L. (1995). Measuring R&D performance. *Research-Technology Management*, **38**(3), 45–54.
Simons, R. (1995). Control in an age of empowerment. *Harvard Business Review*, **73**(2), 80–88.
Sitter, L.U. de (1981). *Op weg naar nieuwe fabrieken*

en kantoren; Productieorganisatie en arbeidsorganisatie op de tweesprong. Deventer: Kluwer (in Dutch).
Sitter, L.U. de (1989). Moderne Sociotechniek. *Gedrag en Organisatie*, **4/5**, 222–252 (in Dutch).
Sivathanu Pillai, A. and Srinivasa Rao, K. (1996). Performance monitoring in R&D projects. *R&D Management*, **26**(1), 57–65.
Smith, P.C. (1976). Behaviour, results and organizational effectiveness: the problem of criteria. In Dunnette, M.D. (ed.), *Handbook of Industrial and Organizational Psychology*. Chicago, IL: Rand McNally College, pp. 745–775.
Smith, P.G. and Reinertsen, D.G. (1998). *Developing Products in Half the Time: New Rules, New Tools.* New York: VanNostrand Reinhold.
Stahl, M.J. and Steger, J.A. (1977). Measuring innovation and productivity a peer rating approach. *Research Management*, **20**(1), 35–38.
Steele, L.W. (1988). Selecting R&D programs and objectives. *Research-Technology Management*, **31**(2), 17–36.
Steensma, H.K. (1996). Acquiring technological competencies through inter-organizational collaboration: an organizational learning perspective. *Journal of Engineering and Technology Management*, **12**(4), 267–286.
Stevens, S.S. (1959). Measurement, psychophysics, and utility. In Churchman, C.W. and Ratoosh, P. (eds), *Measurement: Definitions and Theories*. New York: John Wiley.
Stillman, H.M. (1997). How ABB decides on the right technology investments. *Research-Technology Management*, **40**(6), 14–22.
Szakonyi, R. (1994a). Measuring R&D effectiveness-I. *Research-Technology Management*, **37**(2), 27–32.
Szakonyi, R. (1994b). Measuring R&D effectiveness-II. *Research-Technology Management*, **37**(3), 44–55.
Szakonyi, R. (1998). Leading R&D: how much progress in 10 years? *Research-Technology Management*, **41**(6), 25–29.
Tenner, A.R. (1991). Quality management beyond manufacturing. *Research-Technology Management*, **34**(5), 27–32.
Tipping, J.W., Zeffren, E. and Fusfeld, A.R. (1995). Assessing the value of your technology. *Research-Technology Management*, **38**(5), 22–39.
Vaivio, J. (1999). Exploring a 'non-financial' management accounting change. *Management Accounting Research*, **10**(4), 409–437.
Waterman, R.H., Peters, T.J. and Phillips, J.R. (1980). The 7-S framework. In Quinn, J.B., Mintzberg, H. and James, R.M. (eds) (1988). *The Strategy Process: Concepts, Contexts, and Cases*. Englewood Cliffs, NJ: Prentice-Hall International, pp. 271–276.
Weerd-Nederhof, P.C. de, Kerssens-van Drongelen, I.C. and Verganti, R. (1994). *Coursebook Managing the R&D Process*. Enchede/Milan: University of Twente/Politecnico di Milano.
Werner, B.M. and Souder, W.E. (1997a). Measuring R&D performance state of the art. *Research-Technology Management*, **40**(2), 34–42.
Werner, B.M. and Souder, W.E. (1997b). Measuring R&D Performance U.S. and German Practices. *Research-Technology Management*, **40**(3), 28–32.
Wheelwright, S.C. and Clark, K.B. (1992). *Revolutionizing Product Development.* New York: Free Press.
Wood, R. (1998). Industrial Research Institute's R&D trends forecast for 1998. *Research-Technology Management*, **41**(1), 16–20.
Zien, K.A. and Buckler, S.A. (1997). Dreams to market: crafting a culture of innovation. *Journal of Product Innovation Management*, **41**(4), 239–318.
Zwaan, A.H. van der (1994). *Engineering the Work Organisation.* Assen: Van Gorcum.

IJMR
June 2000