Strategic Management Journal, Vol. 15, 85–100 (1994)

KNOWLEDGE, INTEGRATION, AND THE LOCUS OF LEARNING: AN EMPIRICAL ANALYSIS OF PROCESS DEVELOPMENT

GARY P. PISANO
Graduate School of Business Administration, Harvard University, Boston, Massachusetts, U.S.A.

This paper uses data on 23 process development projects in pharmaceuticals to explore the broader issue of how organizations create, implement, and replicate new routines. A framework is presented which links approaches to experimentation and the structure of underlying knowledge. Although the concept of learning-by-doing is well accepted in the literature, the framework here suggests that where underlying scientific knowledge is sufficiently strong, effective learning may take place outside the final use environment in laboratories (i.e., 'learning-before-doing'). This proposition is tested by comparing how an emphasis on laboratory experimentation impacts process development lead times in two different technological environments: traditional chemical-based pharmaceuticals and new biotechnology-based pharmaceuticals. The data indicate that in chemical-based pharmaceuticals—an environment characterized by deep theoretical and practical knowledge of the process technology—more emphasis on laboratory experimentation (learning-before-doing) is associated with more rapid development. In contrast, in biotechnology-based pharmaceuticals—an environment in which process technology is often characterized as being more of an 'art' than a science—a greater emphasis on laboratory experimentation does not seem to shorten process development lead times. These results suggest that there is no one best way to learn, but that different approaches may be required in different knowledge environments.

INTRODUCTION

The past decade has witnessed a renewed interest by scholars in the role that organizational capabilities, resources, and other firm-specific assets play in competitive performance (e.g., Teece, 1982; Nelson and Winter, 1982; Wernerfelt, 1984; Hayes, Wheelwright, and Clark, 1988; Prahalad and Hamel, 1990; Chandler, 1990). Terms such as 'core competence' and 'organizational capabilities' have joined entry barriers, strategic groups, and kindred terms in the lexicon of strategic management. While the concept of learning has long fascinated organizational theorists, the proposition that competitive advantage stems from firm-specific skills and capabilities has made learning a focal point of concern in fields such as competitive strategy, organizational behavior,[1] and technology and operations management (Hayes *et al.*, 1988; Hayes and Pisano, 1994).

A growing body of empirical evidence indicates that firms in the same industry often possess significantly different levels of capabilities along such performance dimensions as quality (Garvin, 1988), product development speed (Clark and Fujimoto, 1991; Iansiti, 1994), research pro-

Key words: Organizational learning, process R & D, pharmaceuticals, experimentation

[1] See the February 1991 special issue of *Organization Science*.

CCC 0143–2095/94/100085–16

ductivity (Henderson and Cockburn, 1994), and manufacturing productivity (Hayes and Clark, 1986; Bailey, Bartelsman, and Haltiwanger, 1994). If proficiency at a particular activity (such as manufacturing) is critical to competitive advantage, and such proficiency can be improved over time, then learning must play a central role in the competitive advantage of firms. Without learning, it is difficult to imagine from where a firm's unique skills and competencies would come.

While learning can take many forms and occur in many different organizational settings, empirical research on the topic has focused almost exclusively on the learning curve (Wright, 1936; Hirsch, 1952; Rapping, 1965; Hirschmann, 1964; Alchian, 1959; Arrow, 1962; Stobaugh and Townsend, 1975; Lieberman, 1984). This research has documented the tendency for manufacturing performance to improve with cumulative production experience and has provided an empirical foundation for the concept of 'learning-by-doing' (Arrow, 1962). The learning curve, however, reflects only a narrow slice of the broader phenomenon of organizational learning. Firms routinely create and implement new organizational and technical processes through purposeful planning and R&D prior to the start of production. Such planning and R&D are by no means limited to technological innovation. Firms can also 're-engineer' a wide range of business processes such as customer services, order fulfilment, and distribution. One of the chief challenges of innovation lies not only in designing the process, but also implementing and replicating it within the firm's operating environment. If organizational capabilities are embedded in routines, as many scholars are now suggesting (see, for example, Nelson and Winter, 1982), then how firms go about designing, implementing, and replicating such routines must be a central facet of organizational learning.

This paper attempts to shed light on organizational learning by reporting empirical evidence from a study of the development of new production processes in the pharmaceutical industry. The strategy behind this paper is to use the development of production processes as a window into the broader phenomenon of the creation of new organizational capabilities. Production processes are but one of a broader class of organizational routines that can be found throughout an enterprise (Garvin, 1994). Like other types of routines, production processes have an organizational dimension as well as a technical dimension. However, because production processes are the output of formal and reasonably well-documented R&D projects, their development is more amenable to empirical research than other types of organizational processes.

Based on detailed observation of 23 process development projects, this paper identifies two strategies for learning: learning-by-doing and learning-*before*-doing. The basic thesis explored is that each of these approaches is appropriate in different knowledge environments. The paper is organized as follows: the following section provides a conceptual framework for analyzing development as a learning process. This framework posits that the chief challenge of process development is to learn about and predict how different technical choices will influence performance in the actual future operating environment. The state of prior knowledge about the process technology determines the appropriate strategies for acquiring the requisite feedback. In environments where prior knowledge is weak, high-fidelity feedback requires experiments in the actual production environment ('learning-by-doing'). In contrast, when reliable theoretical models and heuristics exist, laboratory experiments, simulation, and other forms of 'learning-*before*-doing' can be productively harnessed. The process development cycle is described in pharmaceuticals and highlights critical differences between biotechnology and chemical pharmaceutical process development. Data from 23 pharmaceutical process development projects are used to examine the impact of different learning strategies on development lead times. The paper concludes with a discussion of implications for further research on capability-based approaches to strategy.

DEVELOPMENT AS A LEARNING PROCESS

Although there are many ways an organization can acquire knowledge, there is broad consensus in the literature that organizational learning is a problem-solving process triggered by gaps between actual and potential performance (Von Hippel and Tyre, 1993; Dosi and Marengo,

1993; Iansiti and Clark, 1994). As a result of a stream of studies demonstrating an empirical link between cumulative production experience and manufacturing performance (Wright, 1936; Hirsch, 1952; Rapping, 1965; Hirschmann, 1964; Alchian, 1959; Arrow, 1962; Stobaugh and Townsend, 1975; Lieberman, 1984), the concept of learning-by-doing has figured prominently in discussions of organizational learning. The basic premise behind the concept of learning-by-doing is that only through actual production experience can an organization discover process problems that cause a gap between actual and potential performance. Thus, as Von Hippel and Tyre (1993: 25) argue: 'The need for learning-by-doing indicates that the innovation process will often be iterative and that developers typically *can't* "get it right the first time"'.

Indeed, anyone who has tried to learn a new skill (such as driving a car) would appreciate that practice (driving around in the high school parking lot with an instructor) is no substitute for actually performing the skill repetitively in the actual use enviornment (a real road with real Boston drivers). However, although it may be impossible to get everything right the first time, organizations also routinely attempt to anticipate and correct as many problems as possible *before* starting production. That is, not all problem-solving associated with learning is a reaction to on-line feedback. For example, when developing a new product, organizations attempt to anticipate the needs and preferences of future customers (Clark and Fujimoto, 1991). When an organization undertakes process R&D to proactively identify potential problems and to design solutions to those problems before production starts, they are engaging in what might be referred to as 'learning-*before*-doing'. The choice between learning-by-doing and learning-before-doing is clearly a matter of degree. The framework below is used to explore the conditions that might lead an organization to emphasize one of these learning strategies over the other.

Process development: Learning and integration

The fundamental challenge of process development is quite similar across industries despite differences in specific activities. The starting point for process development is a description of the product, or a product design. In chemicals, this might be a written description of the molecule, a formula for the required set of reactions, and other data characterizing the molecule. Product designs normally also include a set of functional specifications as targets. At the time process development starts, of course, the description may be incomplete or in a state of flux. While a well-specified product design might allow a sufficiently skilled person to build a replica of the product, it does not include a specific set of instructions for economically making large quantities. This is the role of the process development. The output of process development is an organizational routine for production. In pharmaceuticals, the organizational routines for manufacturing processes include technical specifications (such as equipment designs, reaction conditions, raw materials) and a complete set of standard operating procedures and instructions used by operators and computers to monitor and control the process. Process development creates the organizational routines needed to replicate knowledge embedded in a product design.

Process developers start with a set of targets for process performance. These might be framed in terms of unit cost, capacity, yields, quality levels, critical tolerances, or other operating characteristics. To simplify the exposition, let C represent the set of performance characteristics of the process when operated under expected commercial conditions. The performance of the process (C) is determined by choices over a set of process parameters (p) which define the technical and organizational characteristics of the routine under development (e.g., raw materials, sequence of reactions, reaction temperatures and pressures, control procedures, etc.). The goal of the process developer is to find a set of process parameters, p, which either optimizes $C^*(p^*)$ or at least achieves minimum target levels of it.

There is ample evidence from research on product development that integration across functional boundaries (Clark and Fujimoto, 1991), system–component interfaces (Iansiti, 1994), different scientific knowledge bases (Henderson and Cockburn, 1994), and sequences of projects (Iansiti and Clark, 1994) plays a critical role in development performance. Integrated problem-solving is also critical for successful process development. For example, in developing a new chemical process, choices about

which chemical solvents to use must be integrated with equipment design decisions since some solvents are too corrosive for certain types of vessels. Cross-functional integration between R&D and manufacturing is a particularly salient issue in process development. Since process performance is affected by interactions between technical choices (e.g., the duration of a drying cycle) and the actual operating conditions and capabilities of the future manufacturing site (e.g., how the plant's drying equipment is operated and maintained), technical choices must be tightly integrated with operating choices and conditions. Indeed, given these interaction effects, it makes little sense to describe a process technology in isolation from the actual operating environment. A process technology *is* ultimately a set of technical choices embodied in a set of operating routines.

In searching for $C^*(p^*)$ the process developer needs feedback about how the process will ultimately perform in the future factory environment. Gaining insights about critical interactions between process and operating variables is an important element of this search. Integrative capabilities rest on having mechanisms in place to generate and facilitate the requisite feedback loops. There are a variety of approaches for generating and facilitating feedback, but one of the most important is experimentation (Ulrich and Eppinger, 1995). Experiments lie at the heart of iterative search processes.

Experiments can be conducted in different ways and under different conditions. *Laboratory tests* of the process during development are one way to generate feedback. However, when a process is tested in the laboratory, the researcher does not actually get to observer C(p); instead they observe laboratory performance (L(p)). Whether such tests are a good simulation of actual production conditions, and thus provide a good basis for predicting performance, will depend on the differences between laboratory test conditions and actual operating conditions. If tests are performed under conditions which differ significantly from actual operating conditions, then test results may not be an accurate predictor of future performance. There are many reasons for the lack of fidelity in these tests. In chemical processes, differences in scale can impact process performance; factory workers may perform certain operating tasks differently from Ph.D. chemists in the laboratory; equipment may be different; there may be subtle differences in the raw materials available for research purposes vs. those available in commercial quantities. In many situations intervening variables are not known. A major challenge of process development is to make predictions about C(p) based on observations of L(p).[2]

There are two approaches to minimizing the error between test results and actual operating results. One approach is to make test conditions as close to actual operating conditions as possible. For example, one might run test batches in a pilot production facility or even a commercial manufacturing plant, rather than in the laboratory. While this provides much higher-fidelity feedback about future process performance, it can also have the added benefit of allowing developers to gain a deeper understanding of the factory environment. In-factory tests create a direct feedback loop between the developers and the production environment.[3] The idea that some things can only be learned by running the process in the factory is consistent with the idea of learning-by-doing. This suggests that production plants, rather than laboratories, should become venues for experimentation as early as possible in the development cycle.

Despite the learning value of development in the plant, one must keep in mind that factory experiments are relatively costly. They use capacity which might be deployed to make saleable products. In addition, the experiments themselves might be more costly because of minimum efficient batch sizes or require investments in specialized equipment. For example, one company in the current study calculated that test batches produced in the commercial factory cost 45 times more than batches produced in smaller-scale development facilities. Also, due to the availability of sophisticated instrumentation, laboratory experiments can generally be conducted with a much greater degree of control and precision than factory tests.

---

[2] For simplicity, the added complication that one can have observational errors even in the laboratory, due to instrument calibration, or other factors is ignored. Thus, the true L(p) might differ from the observed L(p).

[3] Tyre and Von Hippel (1993) find that given the 'situated nature' of learning, the physical location of where problems get solved can have an important impact on how they get solved.

An alternative to doing experimentation in the factory is to have knowledge that allows one accurately to predict performance under actual operating conditions (C(p)) from test results observed in the laboratory (L(p)). Such knowledge might be embedded in formal or informal models containing the relevant underlying variables, their interactions, and their impact on outcomes. The model might be based on theory (e.g., the laws of thermodynamics) or an accumulated body of experience. It is not uncommon for developers to use 'rules of thumb' or heuristics. In chemical processes, for example, developers will sometimes refer to a process as having a 'linear scale-up'. This means that results at small scale can be extrapolated to larger scale with a high degree of predictability. Thus, efficient search not only requires researchers to have good knowledge about L(p) but also how L(p) maps into C(p). On the role of scientific knowledge in supporting efficient R&D, see Nelson (1982).

This sample framework suggests that the appropriate learning strategy depends on the state of knowledge characterizing the technology.[4] Where underlying basic theoretical knowledge is strong, one may know enough about the critical variables and their behavior to design laboratory experiments that provide a reasonably accurate prediction of expected commercial performance. If the researcher knows and therefore can control for enough of the critical variables in the laboratory, there should be fewer surprises when the process is transferred into the commercial setting. Under these conditions, it may be efficient to carry out process development under laboratory conditions, and transfer the process to a pilot or commercial plant only after the process design is largely completed. In these cases, prior scientific and practical knowledge provides predictive models (e.g., 'if we observe L′(p′) in the laboratory, we can expect C′(p′) in the plant'). Such models, even if they are quite informal and even tacit, help problem-solvers to predict how different pieces of the puzzle might fit under different conditions. These models and the basic knowledge underlying them represent mechanisms of integration.

In contrast, where theoretical knowledge is weak and experience limited, too many of the critical variables may be unknown, making it virtually impossible to predict how the process tested in the laboratory will perform when run in the factory. Much of what is learned in the laboratory could be irrelevant or misleading. Unless developers have been lucky, a rework of the process may be required to get the process to operate as planned. For such situations, the feedback necessary for integration can only be generated by experiments conducted under conditions which are as close to actual operating conditions as possible. Without good predictive models, integration requires learning by doing.

PROCESS DEVELOPMENT IN PHARMACEUTICALS: BACKGROUND

This study focuses on the development of production processes for therapeutically active chemical or biochemical compounds used in drugs.[5] Since process development activities are part of larger product development projects, a brief overview of product development in pharmaceuticals would be helpful. *Product development* in pharmaceuticals begins with the discovery and synthesis of a molecule which scientists believe to have desirable therapeutic effects. The development of a compound into a drug product involves a sequence of tests to determine its safety, efficacy, and proper dosage strength and form. The compound is first tested on laboratory animals to determine if it has any toxic side effects. If it appears safe, the drug is then tested on human patients to further determine safety (Phase I clinical trials), efficacy at different dosage strengths (Phase II trials), and overall efficacy (compared with existing treatments or a placebo) in a large patient sample (Phase III trials). Data collected from these clinical trials are then submitted to regulatory

[4] Most process technologies lie somewhere along the continuum between having very strong theoretical or experiential knowledge bases and very weak ones. Bohn and Jaikumar (1992) have developed a useful framework known for characterizing different 'stages of knowledge'.

[5] Thus, the sample excludes the process of formulating the final drug form (e.g., capsule, tablet, cream, liquid) taken by patients. For clarification, the term 'chemical' is used to describe small molecules synthesized through traditional organic chemical methods. 'Biochemicals' is used to describe large protein molecules produced from genetically engineered cells (biotechnology).

authorities (e.g., the Food and Drug Administration—FDA—in the U.S.A.) for review. The drug can only be sold commercially after the FDA (or its equivalent outside the U.S.A.) formally approves it. The entire drug development cycle can take anywhere from 3 to 12 years from the time a compound is discovered until it is approved for sale.

Process development occurs somewhat in parallel with product development. While the specifics of process development for chemical and biotechnology-based drugs are quite different, the basic challenge is the same. Initially, when product research scientists (e.g., chemists or molecular biologists) first discover or synthesize a new molecule, they have a technique for producing it in very small quantities. The techniques used by discovery scientists are not commercially viable production processes. They are generally capable of producing extremely small quantities of the compound, at very high cost, and at very low purity levels. A commercial process must be capable of producing the compound in relatively large quantities (metric tons vs. grams), in extremely pure form, at economically feasible cost levels, and within relevant regulatory constraints.

Processes go through three (often iterative) development stages: process research, pilot development, and commercial plant scale-up. *Process research* involves defining the basic structure of the process. For chemical processes, this usually involves searching for and selecting among alternative 'synthetic routes' or the sequences of reactions used to synthesize the molecule. Process research for biotechnological processes typically involves deciding which type of cell (e.g., bacterial or mammalian) will be used to produce the protein, what genetic manipulations will be required, and what type of purification processes will be required. The goal of process research is to define the basic architecture of the process, rather than all the details. In this sense, it is akin to the 'concept development' phase in most product development projects. As an organizational process, this stage of the development cycle has some distinctive characteristics. It is generally performed by Ph.D. chemists or molecular biologists in laboratory settings using very small-scale equipment (such as shake flasks and test tubes). Although small-scale experiments generate important data and validate knowledge, important aspects of the problem-solving process are conceptual. For example, process research chemists often start their search for potential synthetic routes by examining the literature and deriving possibilities from theory. In one project in the sample, the process research group initially identified 27 theoretically possible ways to synthesize the desired molecule. Before any physical experiments were run on these, further modeling and analysis were done to weed out the processes with serious problems (e.g., toxic by-products or excess complexity). Eventually, these 'thought experiments' narrowed the field down to four possible synthetic processes. Attempts were then made to run these processes in small-scale laboratory equipment. Two of the processes did not work at all. Of the two other viable processes, one appeared to have characteristics that would make it attractive for commercial development. This one was advanced to the pilot development phase.

Pilot development involves scaling up the process to some intermediate scale and selecting reaction parameters (such as timing, temperature, pressure) which optimize the efficiency of the process. Because scale can affect the behavior of both chemical and biochemical processes, pilot scale production serves to uncover process problems and typically triggers additional refinements of the process. In some cases, attempts to scale up run into insurmountable problems and a new round of process research is required to find a new basic process. In contrast to process research, the character of pilot development is much more empirical, as it relies heavily on data generated from actual pilot production runs (or analysis of the output of those runs), rather than on theory. In many companies, a different organization is responsible for pilot production and often the people in this organization have different backgrounds than in process research (e.g., chemical engineers instead of chemists). As a result, technical problems are often framed and solved in very different ways between process research and process development. In process research, problems and their solutions are framed in terms of the basic chemistry or biochemistry of the process. A problem with low yield, for example, might trigger a search for different synthetic routes (in chemicals) or a change in the host production cell (in biotechnology). In contrast, pilot development

tends to focus on physical or mechanical solutions (e.g., changing the flow rate of the process or altering the equipment design). These differences are partly due to differences in personnel background (e.g., scientist vs. engineers), but they also reflect the impact of different physical environments. In a good research laboratory, equipment is never supposed to be a constraint on experimentation; equipment is transparent. In contrast, equipment and other aspects of the operating environment are precisely what define a pilot development facility.[6]

Finally, *commercial start-up* involves not only scaling up the process to commercial scale, but also transferring it and adapting it to the plant where the product will be produced for commercial scale. Transfer procedures typically include documenting the process in detail and transferring these documents, along with development scientists, to manufacturing sites. This is also a phase where unanticipated problems often can and do arise. The transfer process is complete once the plant can make a set number of batches of materials which meet quality specifications. The commercial start-up phase can be challenging because it is where the world of process R&D meets (and often clashes with) the realities of the plant. Whereas process research emphasizes conceptual exploration, deepens fundamental knowledge, generates plausible alternatives to technical problems, and lays foundations for further development, commercial start-up revolves around the immediate and pragmatic problem of getting a process up-and-running within rigid timelines. How smoothly this phase goes depends on how well problem-solving during research and pilot development have integrated knowledge about the factory environment.

Organizations have choices about allocating effort (e.g. resources) across these three stages of process R&D. For example, some organizations tend to emphasize process research and invest a significant share of process R&D resources before they ever test the process in the pilot plant. The philosophy behind this strategy is that identifying and solving as many technical problems as possible up front leads to smoother and more rapid development in subsequent stages. Other organizations prefer to do a larger share of their development in the factory. These organizations generally subscribe to the philosophy that it is impossible to 'get it right the first time'. Only by going into the actual production environment can you discover critical process interactions. Each of these approaches represents different strategies for learning. As discussed earlier, each of these different strategies should be appropriate in different technological environments.

Although the development cycle can be described generically, there are critical differences in the nature of the technological environment between chemical development and biotechnology development. These differences stem from the respective maturities of the underlying scientific fields, the development of relevant scientific theory, and the availability of process engineering heuristics. Chemical process R&D utilizes chemistry and chemical engineering—disciplines which have existed in academia and industry since the eighteenth century (Haber, 1958). There is a long history of basic scientific research in both chemistry and chemical engineering conducted by universities and chemical-producing companies. Much of the relevant theoretical knowledge has been codified in scientific journals and textbooks. In searching for and selecting alternative chemical processes, the developer has at their disposal a wealth of scientific laws, principles, and models which describe the structure of relationships between different variables (e.g., pressure, volume, temperature). As noted earlier, process research chemists in pharmaceuticals often begin their work by deriving alternative feasible synthetic routes from theory. Perhaps just as importantly, there is a long history of practical experience with chemical processes. The chemical industry emerged in the eighteenth century, and chemical synthesis has been used to produce pharmaceuticals since the late 1800s. Through this experience, a large body of engineering heuristics have evolved which are widely used to guide process selection, scale-up, and plant design. In addition, computer-aided modeling is performed to simulate the impact of different process variables on yields, cost, throughput, and capacity. The knowledge base is not just technical, but also extends to organizational issues. Through cumulative experience, pharmaceutical companies have developed organizational routines and standard operating procedures. Well-established routines for quality assurance

[6] For a discussion of how the physical environment can impact problem-solving, see Tyre and Von Hippel (1993).

and process control, production scheduling, changeovers, maintenance, and other production activities define clear constraints about the feasibility of different process technologies within an actual production environment.

The characteristics of the knowledge base underlying biotechnology process development is quite different from those described for chemical synthesis. In comparison to chemical-based drugs, biotechnology is in its infancy. The major discovery triggering commercial R&D on therapeutic recombinant proteins was only made in 1973. The first commercial biotechnology enterprises were founded in the mid-1970s. Although there is extensive basic scientific research in molecular biology, cell biology, biochemistry, protein chemistry, and other relevant scientific disciplines, most of this work has been geared toward the problems of product discovery. Compared with the chemical world, very little basic research has been done on the problems associated with engineering larger-scale biotechnology processes. Not only do process developers in biotechnology have little theory to guide them in searching for and selecting alternatives, they also have very little practical experience. The first biotechnology-based pharmaceutical to be manufactured at commercial scale—recombinant insulin—was approved by regulatory authorities in 1982; and since that time, only a total of about 25 biotechnology-based therapeutics have been approved for marketing. Indeed, there was initially skepticism by some observers that recombinantly engineering processes could even be scaled up. Researchers interviewed during the study generally described biotechnology process development as involving 'more art than science'.

Compared with chemical synthesis, biotechnology process technology is a regime characterized by relatively immature theory and thin practical experience. Using Bohn and Jaikumar's (1992) terminology, bioprocessing technology can be considered at a lower 'stage of knowledge' than synthetic chemical process technology. The weaker knowledge base underlying biotechnology production means that the ability of laboratory research to bring together, integrate, and generate the relevant knowledge will be limited. It is difficult to characterize processes in the laboratory. Feedback from laboratory experiments is likely to be noisy. Process development performance is likely to hinge on the experiments conducted under conditions more closely resembling the final production environment. In contrast, the chemical process environment, with its rich base of theoretical and practical knowledge, provides better opportunities to explore options, characterize the process and make predictions about process performance in laboratory settings. In this environment, high process development performance is likely to hinge on exploiting opportunities for learning during process research.

EMPIRICAL ANALYSIS

Data and sample

The data used in the analysis are drawn from a larger study on process development performance in the pharmaceutical industry. Since the type of information required for the analysis is not publicly available, it was necessary to gain the cooperation from pharmaceutical companies willing to participate in the study. Because these data are highly proprietary, the names of the participating firms and details of specific projects, other than aggregate statistics, cannot be disclosed. Data for the present analysis were collected from 23 process development projects; 13 projects involved the development of traditional chemical processes and 10 involved new biotechnology-based processes. In total, 11 organizations participated in the study (five established drug companies, five relatively young biotechnology firms, and a biotechnology division of a major pharmaceutical firm). For each project, data were collected on the history and timing of critical project events, resources expended, and the details of approaches used to identify and solve problems. These data were obtained through a combination of in-depth interviews with project participants, questionnaires, and proprietary company documents. In total, the data collection process spanned 2 years, and involved close to 200 interviews with personnel from participating R&D sites and plants in the U.S.A. and Europe.

The nature of the data collection process is one reason the sample size is relatively small. A second factor limiting the sample size was the population of potential projects. Each process development project in the sample was associated

with the development of a new molecular entity—a relatively rare event in the pharmaceutical industry. The largest and most productive pharmaceutical firms rarely launch more than one new molecular entity in any given year, and many companies have gone several years without launching any. The situation for biotechnology-based drugs—an emerging area in pharmaceuticals—is even more constraining. Since 1982, only 25 biotechnology-based drugs have been developed and approved. Thus, although the sample of 23 is relatively small, it actually represents a significant share of the total number of projects by all companies completed during the time frame of the study. The small sample size obviously involves trade-offs. On one hand, it severely constrains the statistical analysis. On the other hand, it permits a very deep examination of individual projects which in turn provided insights into the development processes, the nature of the problem-solving, the types of variables to include in the statistical analyses, and the appropriate measures for such variables.

Dependent variable

In turbulent environments there is strategic value in being able to develop new capabilities rapidly. Given the strategy of using process development as a window into the broader phenomenon of organizational learning, the dependent variable is the elapsed *lead time* (in months) between the start of the process development project and its successful completion.[7] A process development project was considered to have started when the organization first began to explore ways of producing the molecule that would be feasible at a larger scale (i.e., the start of process research). A project was considered completed only after the process technology was successfully transferred to the commercial plant and could be operated consistently, within desired performance specifications. That is, the project was viewed as completed only when a fully operational production routine had been established.

[7] Occasionally, process development projects are temporarily halted or are idle because of exogenous factors such as a delay in clinical trials. These idle periods were subtracted out of our measure of process development lead time.

Independent variables

Learning strategy

Learning strategies in development are characterized by the allocation of effort to different phases of the project. As discussed earlier, process development projects go through three phases: (1) process research; (2) pilot development; and (3) commercial start-up. Using data on the number of person-hours invested in the project over different phases, two variables were constructed:

RESEARCH % = percentage of total project person-hours expended prior to the first pilot batch of production.

PILOT DEV % = percentage of total project person-hours expended between the first pilot batch of production and the start of technology transfer to the commercial plant.

A high percentage of project resources expended during the process research phase indicates that the organization is focusing its efforts on laboratory-based learning and small-scale experiments. Because the time between the start of the project and the first pilot batch can vary significantly across projects, and because this may have an influence on the resources expended, a third variable was created to control for the lead time before the first pilot batch:

PILOT-1 LEAD = the number of months elapsed between the beginning of the project and the first pilot batch of production.

Organizational structure

Prior research on development suggests that organizational structure will have an important influence on development performance in general, and lead times in particular. Specifically, more integrated structures have been shown to be associated with rapid development (Clark and Fujimoto, 1991; Iansiti, 1994). In the sample,

two types of organizational structures for process development were observed. In one set of projects, the 'upstream' research activities (such as defining the basic chemistry of the process) were performed in a different organizational subgroup from the 'downstream' development activities (such as process optimization and scale-up in the plant). The other set of projects were characterized by an integrated structure in which a single group was responsible for all phases, from initiating process research through transfer to and scale-up in the commercial manufacturing site. Based on the previous research, the integrated structure would be expected to have shorter development lead times:

INTEGR = 1 if the project used an integrated organizational structure
= 0 otherwise.

Technical content

A dummy variable, CHEM, was coded as 1 if the project involved the development of a traditional synthetic chemical process, 0 if it was a biotechnology project. In preliminary analyses, the effects of additional content control variables, such as number of chemical steps in the process, the scale of the output, and the therapeutic class of the drug, were examined. These other variables did not improve the statistical quality or insight of the models. Since their inclusion also did not impact the other effects examined in the model, they were dropped from further analysis and are not reported here.

Descriptive statistics

Table 1 presents means and standard deviations of the continuous variables, and a frequency distribution for the dummy variable INTEGR, for both the full sample and the chemical and biotechnology subsamples. These descriptive statistics provide a picture of some of the areas where differences exist between the chemical and biotechnology projects. With respect to overall process development lead times, there is a relatively large difference between the chemical projects and the biotechnology projects (80.15 months on average for the chemical projects and 41.40 months on average for the biotechnology projects). This difference of nearly 40 months is interesting in light of the fact that biotechnology is the newer process technology. The higher variance relative to the mean process development lead time for biotechnology projects indicates that biotechnology projects may be associated with greater uncertainty. To the extent that novelty is also associated with difficulty, biotechnology process development projects might have been expected to take longer than traditional chemical projects. The fact that this is not the case in the sample might be due to 'an entrepreneurial firm' effect. All but one of the biotechnology projects were undertaken by relatively smaller and younger entrants into the pharmaceutical business. By virtue of their smaller size and entrepreneurial structures and systems, biotechnology firms may have communication and integrated problem-solving capabilities supporting fast development. Unfortunately, the sample does not include enough variance between firm type and project type to test this hypothesis directly. Interestingly, however, in the one case in the sample where an established pharmaceutical firm undertook a biotechnology project, its lead time performance was actually superior to that of the biotechnology firms developing similar processes. The statistical analysis will provide further insights about the extent to which these differences can be attributed to particular development strategies and approaches, and whether those strategies and approaches tend to be associated with either chemical or biotechnology projects.

With respect to the percentage of resources expended during the research phase, the differences between the average chemical and biotechnology projects appears to be much smaller than the variance within each class. That is, while the biotechnology projects, on average, show a slightly more aggressive strategy for early investments in process research, this does not appear to be a biotechnology-specific attribute. There are some chemical projects that also show very aggressive 'front-loading' of resources early in the project. Likewise, there are a number of biotechnology projects where relatively few resources were invested in process research prior to the start of pilot production. In contrast, there appear to be some very significant differences during the pilot development phase of projects. In the chemical projects, a much greater share

Table 1. Descriptive statistics: Means and frequencies (standard deviations in parentheses)

	Full sample (n = 23)	Chemical projects (n = 13)	Biotech. projects (n = 10)
Percentage of total project hours invested in process research phase (RESEARCH %)	13.70 (13.41)	12.69 (15.60)	15.00 (10.59)
Percentage of total project hours invested in pilot development phase (PILOT DEV %)	42.26 (28.96)	55.15 (29.18)	25.50 (19.04)
No. of months between start of project and first pilot batch (PILOT-1 LEAD)	15.39 (11.28)	18.92 (12.58)	10.80 (7.64)
PROCESS DEVELOPMENT LEAD TIME	63.30 (27.07)	80.15 (20.88)	41.40 (16.34)
Frequency			
Integrated organizational structure	14	4	10

of total development resources were expended during pilot production than in the biotechnology projects (55% vs. 25%). This suggests that a much greater share of the process development in biotechnology projects is going on *after* the process is transferred into the plant. This strategy of doing development in the plant is consistent with the earlier discussion that the weaker knowledge structure characterizing biotechnology should require a greater emphasis on learning-by-doing. Finally, one of the most striking differences between the two subsamples is the distribution of organizational structures. All 10 of the biotechnology projects utilized a single integrated development group responsible from the start of process research through the final scale-up and validation in the plant. In contrast, only four out of the 13 chemical projects used such a structure.

The statistical analysis is done in two stages. Ordinary least squares (OLS) analysis is used to estimate the overall impact of the above variables on development lead times for the entire sample. These results provide a picture of the overall impact of different development strategies, but they do not indicate differential impacts across the chemical and biotechnology classes. A second set of models was therefore estimated separately for the two subsamples to test the hypothesis that process research will be more productive (in terms of reducing lead times) in chemical projects than in biotechnology projects.

OLS results for full sample

For the full sample of projects, the following model was estimated using OLS:

$$\text{LEAD TIME}_i = \beta_0 + \beta_1 \text{CHEM}_i + \beta_2 \text{RESEARCH \%}_i + \beta_3 \text{PILOT-1 LEAD}_i + \beta_4 \text{PILOT DEV \%}_i + \beta_5 \text{INTEGR}_i + e_i$$

where all variables are defined as before. Results of this analysis are shown in Table 2.

Three versions of the model were estimated. Model 1 represents the base case in which only technical class differences are controlled. The results confirm the earlier discussion about the

Table 2. Regression results, full sample: Process development lead time (standard deviations in parentheses)

	Model 1	Model 2	Model 3	Model 4
Constant	41.40*** (6.03)***	42.57*** (6.25)	36.96*** (7.39)	46.72*** (11.48)
CHEM	38.75 (8.02)	27.60*** (7.07)	22.98*** (7.71)	15.91 (9.98)
Percentage of total project hours invested in process research phase (RESEARCH %)		−0.886*** (0.270)	−0.776*** (0.277)	−0.776*** (0.275)
No. of months between start of project and first pilot batch (PILOT-1 LEAD)		1.121*** (0.344)	1.064*** (0.339)	1.034*** (0.338)
Percentage of total project hours invested in pilot development phase (PILOT DEV %)			0.180* (0.133)	0.197* (0.133)
Integrated organizational structure (INTEGR)				−9.858 (8.91)
Adj. R^2	0.504	0.69	0.70	0.707
F	23.34	17.35	14.04	11.62
p	< 0.01	< 0.01	< 0.01	< 0.01

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

lead time advantage of biotechnology-based projects. Model 2 examines the impact of concentrating resources on the research phase of the projects (e.g., the period before the process is first tested at pilot scale). After controlling for the duration of this period (PILOT-1 LEAD), the model shows that a greater concentration of resources during the research phase is associated with a shorter overall development lead time. The negative coefficient on RESEARCH % is significant at $p < 0.01$. Model 3 adds the variable on the percentage of resources expended *during* the pilot phase of development (PILOT DEV %). In this model, the estimated impact of RESEARCH % remains negative and statistically significant but the coefficient on PILOT DEV % is positive, although relatively weak.

Model 4 includes the effect of organizational structure. Consistent with the previous research on product development, the estimated sign of the coefficient on INTEGR is negative, suggesting that integrated process R&D organizations tended to be able to develop process more quickly. However, the standard error on the coefficient is relatively large and thus our confidence in this effect is limited. Part of the problem may be related to the level of detail captured in the variable. Previous studies of the product development have probed organizational structure at a much greater level of detail and have included metrics of team structures and types of managers associated with the project (see, for example, Clark and Fujimoto, 1991; Iansiti, 1994). By contrast, the metric used here is admittedly quite crude and does not capture many of the important underlying aspects of organizations affecting integration. Perhaps an even greater problem in interpreting this result is the high correlation between organizational structure and the technology class.

Another interesting aspect of the results is the impact of different development strategy variables on the dummy variable, CHEM. Model 1, which makes no adjustments for development strategy or organizational structure, indicates that the average chemical project took 38.75 months longer than the average biotechnology project. However, as shown in Table 2, the coefficient on the technology class dummy receded with the addition of each development strategy variable. It is difficult to interpret the coefficient on CHEM in Model 4 because of the high correlation with INTEGR. However, in Model 3, which does not include INTEGR, the coefficient on CHEM is 22.98 (vs. 38.75 in the completely unadjusted model). This suggests that some share

of the difference between the lead times in chemical and biotechnology projects can be explained by the specific approaches used on individual projects, rather than anything inherent in the technical requirements of the projects themselves. While the technical environment may create constraints and opportunities, what individual firms do and how individual projects are managed seems to matter a great deal.

The earlier discussion suggested that the technical environment may matter in other ways. The appropriateness of different practices and approaches may vary depending on characteristics of the knowledge environment. The analysis below examines the hypothesis that the leverage of research should be greater in the chemical segment due to the stronger knowledge base than in the biotechnology segment.

Analysis of the impact of research effort in chemical technology vs. biotechnology

From the earlier discussion, research is expected to have a greater pay-off (in terms of reduced overall lead times) in chemical projects than in biotechnology projects. To test this hypothesis, a version of the model was estimated with all variables, except RESEARCH %. The residuals from this model provide a measure of lead time adjusted for these other factors. They represent the variance in development lead times that cannot be explained by the length of the pilot period (PILOT-1 LEAD), the concentration of resources in the pilot development phase (PILOT DEV %), organizational structure (INTEGR), or technology class (CHEM). Adjusted lead time was then regressed against RESEARCH % separately for each subsample using a simple OLS model:

$$\text{Adjusted Lead Time}_i = \beta_0 + \beta_1 \text{RESEARCH \%}_i + e_i$$

Results from these regressions are shown in Table 3 and regression plots for each subsample are provided in Figures 1 and 2. The results are strongly consistent with the hypothesis. The coefficient on RESEARCH % for the chemical subsample is negative and highly significant. A greater share of resources expended during the research phase of chemical process development projects is associated with shorter lead times.

Table 3. Regression results, analysis of residual effects: Dependent variable = adjusted lead time[a] (standard errors shown in parentheses)

	Chemicals (n = 13)	Biotech. (n = 10)
Constant	10.238* (5.082)	−2.198 (7.221)
Percentage of total project hours invested in process research phase (RESEARCH %)	−0.807*** (0.259)	0.147 (0.400)
Adj. R^2	0.42	> 0.01
F	9.716	0.134
p	< 0.01	0.72

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

[a] Lead Time has been adjusted to control for the effects of differences in CHEM, PILOT-1 LEAD, PILOT DEV %, and INTEGR.

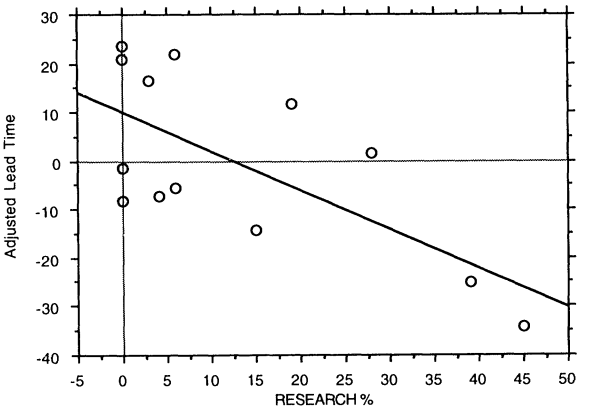

Figure 1. Regression for chemical subsample

Organizations undertaking chemical process development projects appear to be able to learn-*before*-doing. For the biotechnology subsample, in contrast, the regression is insignificant and the estimated coefficient is slightly positive. In biotechnology, additional focus on research does not appear to provide leverage for shortening lead times.

Further analysis of the biotechnology plot suggested an interesting pattern. Although the overall relationship is not statistically significant, three outliers (marked on the graph with shaded points), appear to be masking a relatively strong *positive* relationship between Adjusted Lead

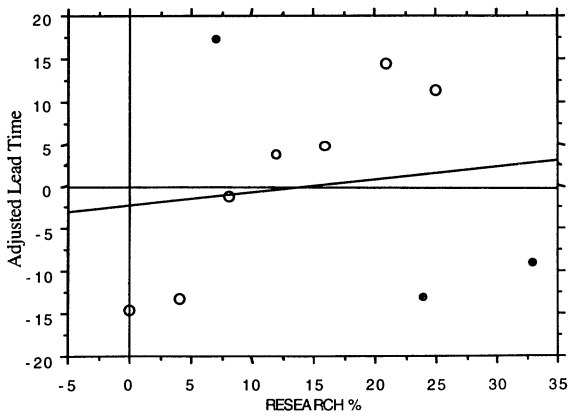

Figure 2. Regression for biotechnology subsample. Shaded points are 'outlier' points referred to in the text. Note: the estimated regression line shown in the plot is for all biotechnology cases, and does not exclude marked outliers

Time and RESEARCH % for most of the biotechnology cases. This leads to further investigation of the three outliers to identify any factors that might differentiate them from the other biotechnology projects. One factor appeared to stand out very strongly: all three outlier projects were undertaken by organizations with relatively more biotechnology process development experience than the others. Clearly, a few outliers do not constitute a trend and conclusions cannot be drawn at this time. However, this investigation suggests that the relationships between experience, firm-specific knowledge, and learning strategies may be worthy of further analysis. One plausible hypothesis is that experienced firms have accumulated deeper technical knowledge that can be tapped through research. A firm with little experience may be forced to 'learn-by-doing' until it accumulates enough understanding of the underlying technical parameters and interactions. It should be stressed that this is offered here as a plausible hypothesis for further investigation, rather than as a conclusion. Subsequent papers from this study will focus on these issues.

CONCLUSIONS AND IMPLICATIONS

The strategy in this paper has been to use process development as a vehicle to explore the broader phenomenon of organizational learning. While process development is but one of many possible activities that leads to the creation of new organizational knowledge, it has two characteristics that make it useful for this purpose. First, to the extent organizational knowledge is embodied in routines, the study of process development provides some insight into how such routines are created, implemented, and replicated. Second, going back to Schumpeter (1934), it has been well understood that organizational learning requires integration of new and existing knowledge. The integration required during development projects (Clark and Fujimoto, 1991; Iansiti, 1994; Henderson and Cockburn, 1994) is a microcosm of the learning processes within organizations (Iansiti and Clark, 1994).

The results of the analysis indicate that there is no one best approach to learning (learning-by-doing vs. learning-before-doing), but that it depends on the nature of the firm's knowledge environment. Deep knowledge of the effect of specific variables and their interactions increases the leverage of research and other forms of learning-before-doing. Learning-by-doing is required when organizations lack the underlying knowledge needed to simulate and predict effects 'off-line'.

Two caveats to the findings should be stressed. First, like most other studies of development, the small sample size has placed severe constraints on the power of the statistical analysis. The results presented here suggest some interesting patterns that can hopefully be further validated in other studies in other industry settings. Second, while process development in pharmaceuticals may be a useful window into broader issues, much more empirical analysis is required in other industry contexts and for other types of organizational activities to get a more complete picture. It might be useful to test the hypothesis as it relates to organizational innovations. Extrapolating the results of this paper, the most rapid approach to implementing an organizational innovation (such as a new product development process, a new way of handling customer complaints, a new incentive plan for the sales force, etc.) may depend on the structure of knowledge characterizing the specific organizational technology. For example, if the contemplated change lies in an area where there is well-developed and empirically validated theory and where the firm

has experience making similar changes in the past, then more effort in planning and organizational design might be valuable in accelerating implementation. In contrast, where organizational theory and practice are not well developed, detailed up-front planning may accomplish little. Instead, the organizations may need to experiment by implementing a specific change and observing how it works in practice.

The general conclusion that different approaches to learning may be required in different types of environments has some potentially interesting implications for strategy. Resource-based views of strategy emphasize the value of knowledge and organizational competencies as competitive assets (e.g. Winter, 1987). The framework and data presented here draw attention to the interaction between the knowledge base of the firm and its competencies. Qualitatively different types of organizational competencies are required to exploit different types of knowledge bases. The locus of strategically valuable resources and competencies may vary accordingly. For example, in environments characterized by rich scientific knowledge bases and detailed understanding of underlying causes and effects (higher stages of knowledge to use Bohn and Jaikumar's terminology), resources supporting research may be critical to competitive advantage. In contrast, in environments where technology is more art than science, resources that support learning-by-doing capabilities are likely to be very valuable. The discussion suggests that the appropriate characterization of the technical environment should go beyond the usual delineation by R&D intensity. Both of the environments included in this study—chemical process technology and bioprocessing—are highly R&D intensive. Yet they differ fundamentally in the degree to which process R&D is driven by theory and prior experience. Further research mapping characteristics of the environment (technical as well as competitive) into requirements for organizational processes is a fruitful area for further research.

Finally, some aspects of an organization's knowledge environment may be idiosyncratic. Thus, even within the same industry or same technology area, different firms may need to utilize different approaches to learning. Similarly, as firms gain experience in a technology through learning-by-doing, their knowledge base becomes deeper and they may have opportunities to be more proactive in their learning. Since technical environments are rarely static, and knowledge bases of individual organizations evolve as a matter of course, learning processes within a firm may need to change over time. Whether and why some organizations can adapt their internal processes more successfully than others are critical issues in understanding organizational learning and the ability of firms to sustain high performance over long periods of time.

ACKNOWLEDGEMENTS

This paper is part of a larger study on process development in pharmaceuticals. I am grateful to Steven Wheelwright for his collaboration on this project and to the companies who participated in the study. I would like to express my appreciation to the editors of this special issue for helpful comments on an earlier draft. I am also thankful to Kim Clark, Robert Hayes, Marco Iansiti, Richard Nelson, Richard Rosenbloom, and Eric von Hippel for helpful comments, and to Sharon Rossi for excellent research assistance. All errors and omissions remain entirely my responsibility. The financial support of the Harvard Business School Division of Research is gratefully acknowledged.

REFERENCES

Alchian, A. (1959). 'Costs and output'. In M. Abramowitz (ed.), *The Allocation of Economic Resources: Essays in Honor of B. F. Haley*. Stanford University Press, Stanford, CA, pp. 23–40.

Arrow, K. (April 1962). 'The economic implications of learning by doing', *Review of Economic Studies*, **29**, pp. 166–170.

Baily, M. N., E. Bartelsman and J. Haltiwanger (May 1994). 'Downsizing and productivity growth: Myth or reality?', Center for Economic Studies, Discussion Papers, U.S. Department of Commerce, Washington, DC.

R. Bohn and R. Jaikumar (1992). 'The structure of technological knowledge in manufacturing', Harvard Business School working paper.

Chandler, A. D., Jr. (1990). *Scale and Scope: The Dynamics of Industrial Competition*. Harvard University Press, Cambridge, MA.

Clark, K. and T. Fujimoto (1991). *Product Development Performance: Strategy, Organization, and Management in the World Auto Industry*. Harvard Business School Press, Boston, MA.

Dosi, G. and L. Marengo (1993). 'Some elements of an evolutionary theory of organizational competencies'. In R. W. England (ed.), *Evolutionary Concepts in Contemporary Economics*. University of Michigan Press, Ann Arbor, MI, pp. 234–274.

Garvin, D. (1988). *Managing Quality*. Free Press, New York.

Garvin, D. (1994). 'The processes of organization and management', Harvard Business School working paper #94–084.

Haber, L. (1958). *The Chemical Industry During the Nineteenth Century*. Oxford University Press, Oxford.

Hayes, R. H. and K. B. Clark (1986). 'Exploring the sources of productivity differences at the factory level'. In K. Clark, R. Hayes and C. Lorenz (eds.), *The Uneasy Alliance: Managing the Productivity–Technology Dilemma*. Harvard Business School Press, Boston, MA, pp. 151–188.

Hayes, R. H. and G. Pisano (January–February 1994). 'Beyond world-class: The new manufacturing strategy', *Harvard Business Review*, pp. 77–86.

Hayes, R., S. Wheelwright and K. Clark (1988). *Dynamic Manufacturing: Creating the Learning Organization*. Free Press, New York.

Henderson, R. M. and I. Cockburn (1994). 'Measuring competence? Exploring firm effects in pharmaceutical research', *Strategic Management Journal*, Winter Special Issue, **15**, pp. 63–84.

Hirsch, W. (1952). 'Manufacturing progress function', *Review of Economics and Statistics*, **34**, pp. 143–155.

Hirschmann, W. (1964). 'Profiting from the learning curve', *Harvard Business Review*, **42** (1), pp. 125–139.

Iansiti, M. (1994). 'Real-world R&D: Jumping the product generation gap', *Harvard Business Review*, **71** (3), pp. 138–149.

Iansiti, M. and K. B. Clark (1994). 'Integration and dynamic capability: Evidence from product development in automobiles and mainframe computers', *Industrial and Corporate Changes*, Special Issue (forthcoming).

Lieberman, M. (1984). 'The learning curve and pricing in the chemical processing industries', *Rand Journal of Economics*, **15**(2), pp. 213–228.

Nelson, R. (August 1982). 'The role of knowledge in R&D efficiency', *Quarterly Journal of Economics*, pp. 453–470.

Nelson, R. and S. Winter (1982). *An Evolutionary Theory of Economic Change*. Harvard University Press, Cambridge, MA.

Pralahad, C. K. and G. Hamel (May–June 1990). 'The core competence of the corporation', *Harvard Business Review*, pp. 79–91.

Rapping, L. (1965). 'Learning and the World War II production functions', *Review of Economics and Statistics*, **48**, pp. 98–112.

Schumpeter, J. (1934). *Theory of Economic Development*. Harvard University Press, Cambridge, MA.

Stobaugh, R. and P. Townsend (1975). 'Price forecasting and strategic planning: The case of petrochemicals', *Journal of Marketing Research*, **12**, pp. 19–29.

Teece, D. J. (1982). 'Towards an economic theory of the multi-product firm', *Journal of Economic Behavior and Organization*, **3**, pp. 39–63.

Tyre, M. J. and E. Von Hippel (1993). 'The situated nature of adaptive learning in organizations', Sloan School of Management Working Paper #BPS-3568-93.

Ulrich, K. T. and S. Eppinger (1995). *Product Design and Development*. McGraw-Hill, New York.

Von Hippel, E. and M. Tyre (1993). 'How the "learning by doing" is done: Problem identification in novel process equipment', Sloan School of Management Working Paper #BPS 3521–93.

Wernerfelt, B. (1984). 'A resource-based view of the firm', *Strategic Management Journal*, **5** (2), pp. 171–180.

Winter, S. G. (1987). 'Knowledge and competence as strategic assets'. In D. J. Teece (ed.), *The Competitive Challenge*. Ballinger, Cambridge, MA, pp. 159–184.

Wright, T. P. (February 1936). 'Factors affecting the cost of airplanes', *Journal of Aeronautical Science*, **3**, pp. 122–128.