5. DESIGN-SCIENCE RESEARCH

In this chapter we try to answer the questions: Can we build a certain innovation and how useful is a particular innovation? We can also ask, which kind a certain innovation ought to be, and how ought we to build a particular innovation? For example, how could we improve our human-computer interface that the number of errors will be diminished? If our research question contains the following verbs: build, change, improve, enhance, maintain, extend, correct, adjust, introduce, etc., our study might belong to the design-science research.

Research is normally divided to basic and applied research. The purpose of the *basic research* is to find out what is a part of reality. We considered different research approaches suitable for the basic research in Chapters 2, 3 and 4. In the *applied research* the knowledge of the basic research, the basic laws of the explanatory sciences are applied to. This view disregards the impressive body of knowledge developed by the design sciences (Dahlbom 1996). Referring to Simon (1981) van Aken (2004) describes that "the mission of a design science is to develop knowledge for the design and realization of, i.e. to solve *construction* problems, or to be used in the improvement of the performance of existing entities, i.e. to solve *improvement problems*", in other words, to implement some innovation. The utility of the new innovation is sooner or later evaluated.

Van Aken (2004) in a design-science study emphasizes both construction and improvement with an integrated outcome, called a technological rule, and one research approach. March and Smith (1995) and later Hevner et al. (2004) only underline construction with four outcomes (constructs, models, methods and instantiations) and two research approaches (build and evaluate). We shall in this chapter combine those two views by presenting first a technological rule, thereafter those four constructs and finally the build (Section 5.1) and evaluate (Section 5.2) approaches. At the end of this chapter (Section 5.3) we shall show how action research is the improvement approach in the sense of van Aken and the combination of the build and evaluation approaches proposed by the others.

Van Aken does not consider an instantiation as the outcome of the design-science study, but its ultimate mission is to develop design knowledge, i.e. knowledge that a professional can use in designing solutions to problems. "It is important to teach a civil engineer subjects like physics and mechanics, but in designing a bridge he or she needs the design knowledge developed by his or her discipline, like for instance the properties of different types of bridges. In the same way a medical doctor should have a working knowledge of physics and biology, but for medical problem solving he or she predominantly uses the results of the clinical research of his/her own discipline."

Design knowledge concerns "three designs: an *object-design*, the design of the intervention or of the artifact; a *realization-design*, i.e. the plan for the implementation of the intervention or for the actual building of the artifact; and a *process-design*, i.e. the professional's own plan for the problem solving cycle, or, put differently, the method to be used to design the solution to the problem. This design knowledge is general, i.e. valid for *classes* of cases. The problem of the professional, however, is always unique and specific. Therefore, general knowledge must be translated to the unique and specific case at hand."

"Within each of the three types of design knowledge, prescriptions are an important category. The logic of a prescription is 'if you want to achieve Y in situation Z, then perform action X'. There are algorithmic prescriptions, which operate like a recipe. However, many prescriptions in a design science are of a heuristic nature. They can rather be described as 'if you want to achieve Y in situation Z, then something like action X will help'. 'Something like action X' means that the prescription is to be used as a *design exemplar*. A design exemplar is a general prescription which has to be translated to the specific problem at hand; in solving that problem, one has to design a specific variant of that design exemplar." (van Aken, 2004)

"In the design sciences the research object is a '*mutandum*'; these sciences are not too much interested in what *is*, but more in what *can be*. The typical research product is the prescription discussed above or in terms of Bunge (1967b, p. 132) a technological rule: 'an instruction to perform a finite number of acts in a given order and with a given aim'. A *technological rule* is defined as *a chunk of general knowledge, linking an intervention or artifact with a desired outcome or performance in a certain field of application*. A major breakthrough occurred with the systematic testing of technological rules. The *tested* technological rule is one whose effectiveness has been systematically tested within the context of its intended use. The real breakthrough came when tested technological rules could be *grounded* on scientific knowledge (Bunge 1967b, 132), including law-like relationships from natural sciences. The typical research design to study and test technological rules is the multiple case: a series of problems of the same class is solved, each by applying the problem solving cycle. By borrowing concepts from software development one can say research on technological rules typically goes through a stage of *α-testing*, i.e. testing and further development by the originator of the rule, to be followed by a stage of *β-testing*, i.e. the testing of the rule by third parties."

March and Smith (1995) defined that "we *build* an artifact to perform a specific task. The basic question is, does it work? Building an artifact demonstrates feasibility. We build constructs, models, methods and instantiations." They gave a more concrete objective for research in the build activity: "It should be judged based on value or utility to a community of users". According to our wording, we should ask, did we achieve our stated goals in the construction? According to March and Smith and later Hevner et al. (2004) design science consists of two basic activities, build and evaluate. Building is a process of constructing an artifact/innovation for a specific purpose; evaluation is a process of determining how well the artifact performs.

According to March and Smith (1995), "*evaluate* refers to the development of criteria and assessment of artifact performance against those criteria. We (March and Smith) try determine if we have made any progress. The basic question is, how well does it work? Evaluation requires the development of metrics and the measurement of artifacts according to those metrics. Metrics define what we are trying to accomplish. They are used to assess the performance of an artifact". March and Smith (1995) wrote also that, if the artifact (i.e. construct, model, method or instantiation) is really novel, "actual performance evaluation is not required at this stage". To our mind, both building and use processes of an innovation must be evaluated. The evaluation of an innovation must cover both processes, actually the whole 'life' of an innovation, from an idea to the first realization and then use, and finally to its demolition must be evaluated. The final stage (demolition) of an artifact may mean either a transition from use of the old artifact to use of the new one or the finish of its use. We can present our object under consideration in Figure 5.1.

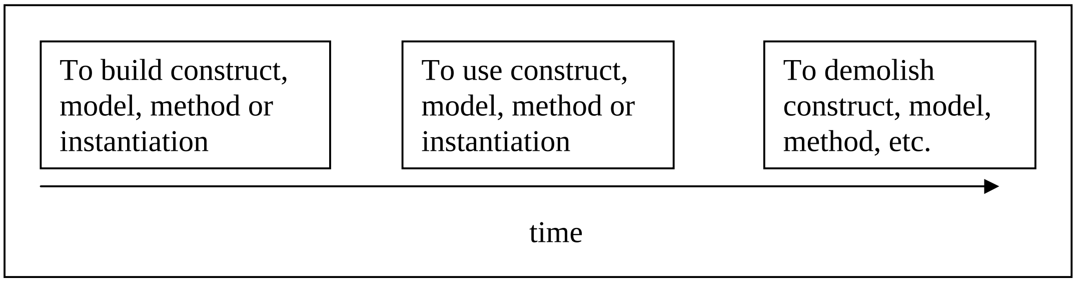


Figure 5.1 The sequential processes (build, use and demolish) of an artifact

The rest of this chapter is structured as follows: We shall firstly describe on the building process and its results (Section 5.1). The research tasks can then be divided to two groups depending on whether the innovation already exists or should it be developed. To estimate the innovation built we shall analyze the use process as a research object and give some recommendations for its evaluation in Section 5.2. We also enclose the demolition process and its assessment into that section. In action research (Section 5.3) the building and evaluation sub-processes form an organic whole.

5.1 The building process

We can consider the building process either after or before the actual realization. As said earlier multiple case studies are valid for the extracting and the developing case study. The *extracting* multiple case-study (van Aken 2004) "is a kind of best-practice research and is aimed at uncovering technological rules as already used in practice. A good example of such research is the classical study of Womack et al (1990) of the automotive industry and especially of Japanese practices. This research has produced, among other things, a number of very powerful technological rules, like the Kanban-system and Just-in-Time delivery for driving a supply chain.

In the *developing* multiple case study the technological rules are developed and tested by researcher(s) in close collaboration with the people in the field and often in the context of application. Such research is initiated by the researcher(s) interested in developing technological rules for a certain type of issue. Each individual case is primarily oriented at solving the local problem in close collaboration with the local people. Following the reflective cycle, after each case the researcher develops knowledge that can be transferred to similar contexts on the basis of reflection and cross-case analysis. –This development process can first go through a stage of α-testing, i.e. analysis of effectiveness of a certain rule in the original context. But invaluable insight can be gained by subsequent ‘β-testing’, i.e. translating the rule to other contexts, having third parties use it, assess its effectiveness and make final improvements. It is this β-testing, which can provide further insight into the indications and contra-indications for the rule and hence in its application domain.” Both the successes and the unsuccessful applications should be included into the scientific knowledge base of design science.

March and Smith (1995) and later Hevner et al. (2004) restricted their consideration on IT technical products. According to them design science products are of four types, constructs, models, methods, and instantiations. We use their definitions. *Constructs* or concepts form the vocabulary or language of a domain. A *model* is a set of propositions or statements expressing relationships among constructs. A *method* is a set of steps (an algorithm or guideline) used to perform a task. An *instantiation* is the realization of an artifact in its environment. As we described at the beginning of this chapter we have enlarged the artifact concept to the innovation concept. Instantiations operationalize constructs, models and methods. We shall below use all those four terms in our presentation. The motivation behind the building a new innovation is either lacking of that innovation or low quality of the outcomes achieved by old innovation. It is almost always possible to identify the starting point of an effort to construct something new and also the contemporary view on the desired state, e.g. the functioning artifact. The purpose of the construction process is to achieve a movement from the initial state to the goal state (Figure 5.2).

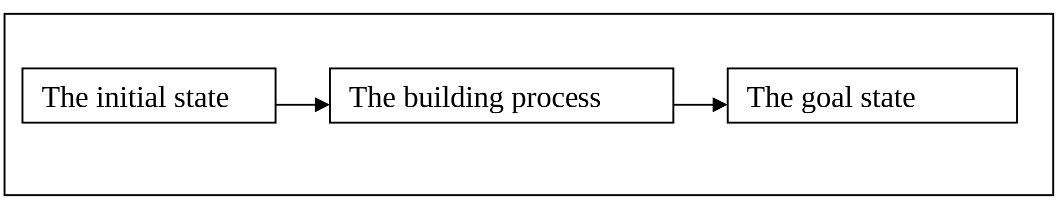


Figure 5.2 The building process

Some of the interested parties have perhaps considered the initial state to be problematic. The performance criteria of the old version of artifact or innovation may be below the stated objectives. Some party can also have an idea or a concept to apply or to use some resources (technological, human, data/ information/ knowledge, financial resources) in a new way in order to solve the problem. This concept resembles a business concept or business idea. In practice it can be a new theoretical or practical, e.g. technical invention.

To think construction task the goal state can be known or unknown. If it is known, the task of researchers as builders is to try to implement the desired change from the initial state to the goal one. If the goal state is unknown, we have at least two alternatives. We can firstly specify the goal state and then try to implement measures to achieve that state or we can in parallel realize both goal-seeking and implementation. Instead of implementing the totally new version of artifact by ourselves, we can also purchase a ready-made artifact, if such one exists and is for sale at a competitive price. The good hopes of builders will not always materialize, but the final state may differ from the goal state (Figure 5.3).

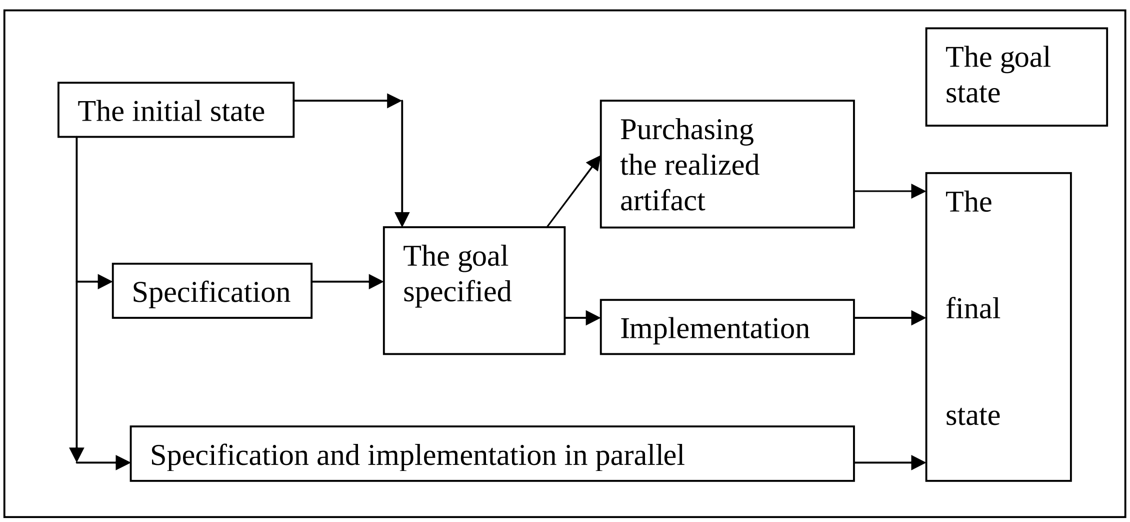


Figure 5.3 Different alternatives concerning the building process and its outcomes

March and Smith (1995) connect two models to two states, the first one to the initial state and the second one to the goal state, in such a way that the models represent situations as problem and solution statements. It means how things are at the beginning and how they ought to be at the goal state (a normative model). The (descriptive) model of the initial state may need to capture the structure of reality in order to be a useful representation. To emphasize the utility aspect motivating construction the (problematic) initial state is evaluated by using a certain utility metrics (or many), and the goal state is estimated to be better, more valuable, more desired with the same metrics. The model of the initial and/or goal state can (but need not) contain one or more new constructs.

According to Weber's (2003) recommendations the models of the initial and goal states contain both the lawful state and event spaces. The lawful state space is the set of combinations of construct values for which the model is expected to hold. We begin to specify the lawful state space of our model when we select the constructs to include in our model. The choice of constructs dictates the *things* in the world to which our model applies. The lawful event space is a set of changes of state of the constructs for which the model is expected to hold. In some cases, an event is unlawful because either the prior state or the subsequent state is unlawful. In some circumstances, however, both the prior state and the subsequent state are lawful but the transition between them is unlawful. In the construction task the model of the initial state (M/initial) contain both the lawful sub-state and sub-event spaces, and the model of the goal state (M/goal) both the lawful sub-state and sub-event spaces. If model M/initial differs from model M/goal, i.e. if something new will be constructed, then the lawful sub-state and sub-event spaces in model M/initial differs from the lawful sub-state and sub-event spaces in model M/goal. The differences between models M/initial and M/goal are intended to be stable.

According to March and Smith methods are based on a set of underlying constructs (language) and a representation (model) of the solution space. Methods are often used to translate from one model or representation (M/initial) to another (M/goal) in the course of solving a problem. To this end models and methods play a central role in our consideration. – To our mind constructs are in some sense subordinate for models and methods.

Our aim is to develop some criteria and measures to estimate the building methods, too. We therefore return to the views on methods taken by March and Smith who assume a well-structured building task and therefore emphasize the implementation process only. We cannot totally agree with them, because at the beginning of the specification and parallel processes there does not necessarily be any model representing the solution space (i.e. the building task is ill-defined). The task of the specification process is to find that model. The other path from the initial state to the final state goes via the parallel process, and its idea is to define the solution space model in the course of the process from the initial state to the final one.

The specification process

The purpose of this sub-section is to consider various sub-methods to derive and determine the goal state of the desired innovation or artifact under construction. There are normally at least two parties of professionals, say users and a researcher. The new artifact is intended to be used by users, and a researcher is assumed to best know the main idea, let it be technical, organizational, informational or combination of them, to be applied to the new artifact. The potential communication problem is based on division of labor between users and a researcher. Both parties have difficulties to understand each other, because their professional languages differ much or they may play the so called language game.

Users emphasize the business needs and utility of a new innovation. A researcher stresses new scientific knowledge and a real new innovation (as an opposite of routine design (see Hevner et al. 2004)) to be constructed. According to van Aken (2004) there are good chances to achieve a win-win situation for both parties. A researcher sometimes develops the first prototype herself and then brings it to the negotiations with users. Both parties can then have a more concrete basis for their discussion of the desired goal state.

The implementation process

In this sub-section the research question can be formulated as follows: Having the certain initial and goal states and particular resources, how could we build an artifact satisfying the given specifications? To solve this "how" problem we need some theoretical and/or empirical *concepts and ideas*. In consideration of information systems development Nunamaker et al. (1991) included the similar principle in the first stage of their methodology: Concept design is the adaptation and amalgamation of technological and theoretic advances into potentially practical applications. Those *technological and theoretic advances* utilized in building the artifact must be described in the study report. Our view is not restricted to information systems only but other potential artifacts and innovations can be constructed, and then *informational and organizational ideas* utilized in building the artifact must also be described in the study report.

Another important topic to be included into the study report is *design alternatives*. In order to find solution to the how-problem, a researcher can at least apply two different heuristics: problem reduction and state-transition. In the *problem reduction heuristics* a problem is divided into sub-problems, and the latter again into sub-sub-problems, etc. until the solvable sub…problem is found. This reduction can be performed by applying two principles: breadth first and depth first. By using the *breadth first* principle a researcher divides a problem into sub-problems and tries to solve all the sub-problems at the highest level first. With unsolvable problems she then continues to the next lower level, and so on. By using the *depth first* principle a researcher first seeks the most difficult sub-problem and tries to solve it, and after that she solves the next difficult problem and so on. If some sub-problem is not solvable, the construction process stops. In the reporting we recommend that a researcher first consider the biggest or largest or most demanding design task, develop some best alternatives for it, select the best alternative. In a report both alternatives and the best one with evidence are presented.

The *state-transition heuristics* means that the original problem definition, i.e. the initial state, is transformed to the goal state in the sequence of transitions. A researcher can report every transition from state i to state i+1. If there are too many transitions to be included into a report, the most significant transformations can be described.

Hevner et al. (2004) illustrated and evaluated the application of the design-science by selecting three articles as exemplars:

• Gavish and Gerdes (1998), which develops techniques for implementing anonymity in Group Decision Support Systems (GDSS) environments

• Aalst and Kumar (2003), which proposes a design for an eXchangeable Routing Language (XRL) to support electronic commerce workflows among trading partners

• Markus, Majchrzak, and Gasser (2002), which proposes a design theory for the development of information systems built to support emergent knowledge processes.

Purchasing a ready-made artifact

Buying a complete product is an alternative to make your own one. It is not reasonable to re-invent a bicycle, if the same or similar system already exists. In order to know what to buy you must determine and specify your goal state. The purchasing process itself consists of such phases as finding out the potential candidates, selection the best candidate, acceptance test and change-over or introduction of the new product. A researcher herself can try to find candidates or she can ask submissions of bids. The best candidates are then compared by using specifications, performance figures and use tests. Actually it is a comparative evaluation (Section 5.2) of candidates. A researcher can then enter into a contract of purchase with responsibilities and compensation clauses. If the delivery is not immediate, some acceptance test must be organized, when the new system or artifact moved into the destination. The new system is not necessarily ready to be used from the first moment, but users must be trained, some practices adjusted in a new way, etc.

The purchasing process above mainly concerns a real product, not any service. The artifact purchased from outside is normally materialized, not social. From outside we can also bring social innovations, for example, best practices into use, not as such but to be learned and adjusted into the local context (Wareham and Gerrits, 1999). This is example, how new services are developed. They can also be purchased from outside by using outsourcing (Loh and Venkatraman, 1992).

The merit of the purchasing process as a scientific effort can be small. It must be considered as an alternative of the building process, i.e. the same or similar artifact (to be purchased) does not yet exist. If some artifact is, however, purchased as a component of the larger system, the purchasing process must be shortly reported and evidence for the selected alternative given.

Parallel specification and implementation processes

The gradual development of the artifact is often based on fact that people have difficulties to imagine such that did not yet exist. To this end the definition of the goal state is difficult. Instead of that some sketches of the goal state are experimented and at the same moment also realized by prototypes. Winograd (1995) proposed some responsive prototyping media like rough hand sketches and scenarios, low-fidelity prototypes (wizard of oz), programmed facades and prototype-oriented languages, by which you build an application but it is too slow for productive use.

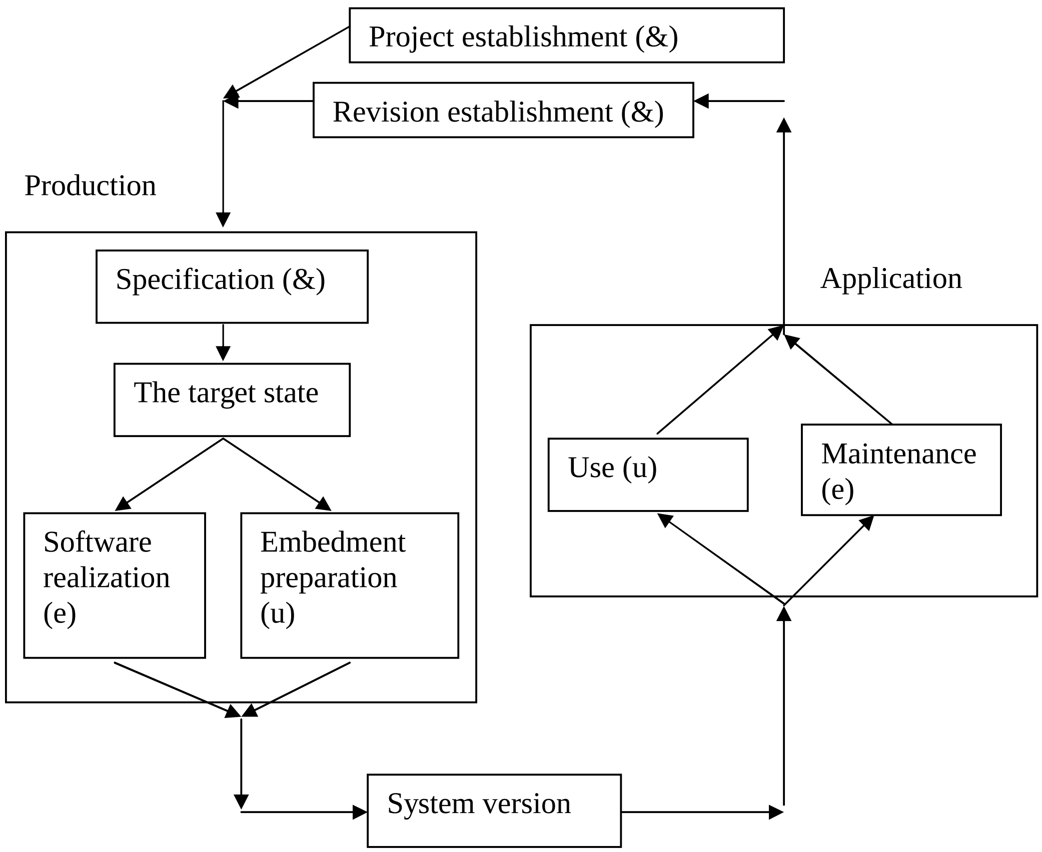


Figure 5.4 Parallel specification and implementation processes for constructing new versions

In Figure 5.4 we have used the following notations: u = user of innovation performs a task, e = expert performs a task, & = user and expert together perform the task.

Floyd and others (1989) have long developed software engineering method called STEPS (Software Technology for Evolutionary Participative System Development). It recommends that a new version is given to (experimenting) use, maintained during the use and because of the bigger change need the new version is realized (Figure 5.4). Floyd and other assume that future users cannot themselves implement a new software, but two professionals are needed. Instead of software terminology we in Figure 5.4 used a general terminology. Users learn at work, and it continually generates new ideas for development of the system. A researcher must sometimes stop either developing new versions or following the development cycles. She can report a sequence of the versions, their main ideas or solution concepts and impacts.

Engeström (1987, 323) developed the methodological cycle of expansive developmental research with 5 phases: 1) Activity 1 (phenomenology, delineation), 2) Analysis of activity (object-historical, theory-historical and actual-empirical ones), 3) Formation of new instruments (springboard, models, microcosm), 4) Practical application of new instruments to change the activity through strategic tasks, and 5) Activity 2 (reporting and evaluation).

On evaluation criteria of innovation

March and Smith (1995) wrote that “research in the build activity should be judged based on value or utility to a community of users”. They differentiate two cases concerning whether the construct, model, method, or instantiation already exists or is it totally lacking. For the latter case “building the *first* of virtually any set of constructs, model, method, or instantiation is deemed to be research, provided the artifact has utility for an important task. The research contribution lies on the novelty of the artifact and in the persuasiveness of the claims that it is effective. Actual performance evaluation is not required at this stage” (March and Smith 1995). To our mind, comparison of *the totally new* construct, model, method, or instantiation takes place with "nothing" or with some conceivable one, i.e. the potential importance of the new construct, model, method and instantiation should be evaluated. To this end, a researcher could ask, for example, whether the new *construct* better differentiates (and/or describes) the phenomenon, to which this construct refers, from other phenomena than any other conceivable construct. By emphasizing the utility aspect, a researcher could also ask the potential benefits of the new construct in use.

For the former case, the construct, model, method, or instantiation *in a certain form already exits*, March and Smith (1995) gave the recommendation: “The significance of research that builds subsequent constructs, models, methods, and instantiations addressing the same task is judged based on ‘significant improvement’, e.g. more comprehensive, better performance”. To apply the comparison idea a researcher could ask: Is the new construct, model, method or instantiation in some sense better than the old one? Phrase 'in some sense' means a certain assessment criterion used in comparison.

Compared those recommendations above with van Aken's (2004) views, we find some overlapping cases and some amendments. She emphasizes both construction and improvement with an integrated outcome, called a technological rule, and one research approach. Construction corresponds to the totally new construct, model, method, or instantiation of March and Smith, and improvement, as it says, to 'significant improvement'. According to van Aken in the developing multiple case study the technological rules are developed and tested by researcher(s) in close collaboration with the people in the field and often in the context of application. The extracting multiple case study is a kind of best-practice research and is aimed at uncovering technological rules as already used in practice. To our mind, the latter is new in the design-science literature, although March and Smith do not explicitly negate it. The emphasis of the technological rule is also new. We understand that the technological rule underlining prescriptions is close to the method describing how the big change (construction or improvement) was achieved. According to van Aken the technological rule must be tested and grounded. - To summarize, we collected the results into Table 5.1.

Table 5.1. Evaluation of the building process with the old and totally new outcomes

	March and Smith (1995)	Van Aken (2004)
The old outcome exists	Significant differences between the old construct, model, method or instantiation and new one, respectively	As March and Smith, plus a special emphasis of the tested and grounded technological rule in both the developing and extracting multiple case studies
The totally new research outcome	"Actual performance evaluation is not required at this stage."	The technological rule in the developing multiple case study with both positive (tested and grounded rules with both driving and blocking mechanisms) and negative results

The *practitioner* may construct the fine artifact. In order to get a scientific merit for it, she must describe the building process in detail, argue her selections and explain her decisions. The originality of the solution and its superiority to the known solutions must also be demonstrated.

According to Hevner et al. (2004) the purpose for establishing the seven guidelines is to assist researchers, reviewers, editors, and readers to understand the requirements for effective design-science research. Following Klein and Myers (1999), we advise against mandatory or rote use of the guidelines. We contend that each of these guidelines should be addressed in some manner for design-science research to be complete. Table A summarizes the seven guidelines.

Table A. Design-Science Research Guidelines

Guideline	**Description**
Guideline 1: Design as an artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a search process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Van Aken (2004) studying management research gave another list of guidelines:

1. *Descriptive relevance* or external validity: the raison d’être of a technological rule is its external validity as established by testing in multiple case-studies.

2. *Goal relevance* or the extent to which results refer to matters the practitioner wishes to influence: in a prescription-driven research program goal relevance is a key criterion for the choice of rules to be developed, tested and grounded.

3. *Operational validity* or the extent to which the practitioner is able to control the independent variables in the model: the very nature of a technological rule assures its operational validity.

4. *Non-obviousness*: because a technological rule is not forced into a reductionistic format as quantitative causal models are, there is little danger of overly obvious research results.

5. *Timeliness*: a practitioner need arising from the ‘incredible long periods of time’ required to adequately assess organizational phenomena and the scientists’ reluctance to make recommendations before all the facts are in: in this respect the technological rule has no advantage over the causal model; for classes of management problems for which timeliness is a real issue, the practitioner will have to deal with consultants rather than with academic researchers.

On structuring a report of the artifact building process

In this sub-section we shall present our proposals for table of contents used in a research report. We give two proposals when the outcome is a new normative method with prescriptions and one proposal for a new instantiation (3). A new method can be derived in two ways, deductively (1) or inductively (2).

(1) A new normative method (derived deductively)

1. Introduction (see Section 9.2)
2. Selection of the method (arguments, excluding other alternatives)
   Selection of data gathering techniques Section 7.1
3. Earlier methods and their explicit or implicit objectives and assumptions on reality
   (validity, realistic view)
4. Survey of general problem-solving methods
5. Construction of the new method
   (its [theoretical] idea or concept, alternatives selected, proof if possible, first test results)
6. Comparison of the new method with the old ones
7. Discussion (Section 9.2)

(2) A new normative method (derived inductively)

1. Introduction (see Section 9.2)
2. Selection of the method (arguments, excluding other alternatives)
   Selection of data gathering techniques Section 7.1
3. Earlier methods and their explicit or implicit objectives and assumptions on reality (validity, realistic view), earlier experiences achieved, reported negative results from the use of earlier methods, improvement ideas,
4. New technical, social and informational opportunities to new procedures
5. The objectives of the new method and the values behind them; construction of the new method and its theoretical and empirical evidence, the first evaluation in practice or as thought experiment)
6. Comparison of the new method with the old ones
7. Discussion (Section 9.2)

We propose that a researcher would describe her building process in detail enough in order to persuade a reader that she developed the best possible solution.

(3) Implementation of an innovation

1. Introduction (see Section 9.2) (contains description of a construction task)
2. Presentation of the new idea or concept (compared with alternative ideas and concepts)
3. Restrictions (concerning long term physical, human, financial and data resources)
4. The greatest design problem, its solution alternatives and the selected alternative with reasons
5. The next greatest design problem, … (as in Item 4)
6. The description of implementation and a preliminary evaluation (or if the implementation is stopped at the design level, the design must be performed as in detail as a reader can be sure that it can be sure that the plan is realizable)
7. A preliminary evaluation of the new innovation
8. Discussion (Section 9.2)

If the solution is yet implemented, the plan of implementation should be described as in detail as a reader can be convinced of the possible realization of the plan. Our proposed structure above is not exactly suitable for each project, but a researcher can in a creative way apply it to her case.

5.2 Evaluation of construction results

March and Smith (1995) describe that "research in the evaluate activity develops metrics and compares the performance of constructs, models, methods, and instantiations for specific tasks. Metrics define what a research area is trying to accomplish. Since 'the second' or subsequent constructs, models, methods, or instantiations for a given task must provide significant performance improvement, evaluation is the key activity for assessing such research." March and Smith (1995) considered all types of artifacts: constructs, models, methods and instantiations and proposed some metrics for them. We think that they strove to give as universal metrics as possible. We shall structure this section to two sub-sections. Firstly we present the universal metrics proposed by March and Smith. Secondly, we develop some new metrics.

On 'universal' metrics proposed by March and Smith

We shall follow the order: constructs, models, methods, and instantiations, and present the proposals of March and Smith and then comment the proposals. According to March and Smith "evaluation of *constructs* tends to involve completeness, simplicity, elegance, understandability, and ease of use". March and Smith did not give any rationale for their list. To our mind, the list above mainly concerns the use of an individual working alone. But during the building and use of an IT artifact stake-holders, e.g. systems analysts, users and managers, are mostly working together. The view on constructs, forming the vocabulary of a domain, also seems to emphasize communication, not only description.

Boland and Tenkasi (1995) analyzed knowledge-intensive firms with specialties and knowledge disciplines. They called a group of specialized knowledge workers with term "community of knowing". The authors argued that producing knowledge (including new constructs) requires the ability to make strong perspectives within a community, as well as the ability to take the perspective of another into account. Knowledge work of perspective making and perspective taking requires individual cognition and group communication. Boland and Tenkasi presented two models of language, communication (language game and conduit) and cognition (narratives and information processing) for amplifying our thinking. Those models can raise new criteria for evaluation of constructs.

Models are according to March and Smith "evaluated in terms of their fidelity with real world phenomena, completeness, level of detail, robustness, and internal consistency". The first criterion, the fidelity of the model with real world phenomenon, can be checked ex post, but at the beginning of the implementation process (ex ante) the model of the goal state describes a desire not yet realized. The next two criteria, completeness and level of detail, can related to the reality by following Smith (1985) who wrote that "every model deals with its subject matter at *some particular level of abstraction*, paying attention to certain details, throwing away others, grouping together similar aspects into common categories, and so fort. … Models *have* to ignore things exactly because they view the world at a level of abstraction. And it is good that they do: otherwise they would drown in the infinite richness of the embedding world". Hence, we cannot demand completeness of the model in relation to reality. The robustness criterion was not defined. One interpretation can be similar as Bunge did in consideration of classification. "By *classification* we can divide elements into classes or groups. One of the principles of correct classification (Bunge 1967a, 75) is that the characters or properties chosen for performing the grouping should stuck to throughout the work", i.e. the grouping principle(s) is robust. The internal consistency criterion is, to our mind, a natural requirement from research point of view.

By using differentiation (form and content) we can say that the criteria above only concern the content of the model. The models can be represented in many ways, physically, mathematically, pictorially etc. The correct content can be presented in the pictorial form that produces a wrong understanding (Järvinen 2001). We therefore propose that because the models as the constructs above are used in communicating ideas between stake-holders, the representation of the model should support communication. By using the conduit model of communication, we can say the meaning of the sender should be communicated with the model representation to the receiver without any change.

Yap and Bjørn-Andersen (1998) analyzed different media from three aspects: Richness of knowledge representation, level of interactivity and richness of perspective. They described four media as follows:

Picture: non-interactive, two-dimensional representation, static perspective

Video footage: non-interactive, three-dimensional representation, dynamic perspectives

3 D animation: interactive, three-dimensional representation, flexible multiple perspectives
Virtual reality: fully interactive, immersive representation, holistic/integral perspective
The different media above clearly give chances to improve a representation of a model, it will have an influence on understanding, learning and use a model.

Evaluation of *methods* according to March and Smith concerns "operationality (the ability to perform the intended task or the ability of humans to effectively use the method if it is algorithmic), efficiency, generality and ease of use".

Mathiassen (1981) alone and with Munk-Madsen (1986) demanded that for every new method its application domain must be defined, i.e. what is the smallest and the largest problem the method could solve. For technological rules van Aken (2004) proposed that both driving and blocking mechanisms should be reported.

Evaluation of *instantiations* according to March and Smith concerns "the efficiency and effectiveness of the artifact and its impacts on the environment and its users". To our mind, March and Smith are only considering the planned changes in efficiency and effectiveness of the artifact. But as Orlikowski (1995) in her paper "Evolving with Notes" showed there can also be emergent changes with positive and negative unanticipated outcomes that accompanied these changes.

Kling (1987) differentiated the discrete-entity models from the web models. The discrete-entity model focuses on relatively formal-rational conceptions of capabilities of information technologies and social settings in which they are developed and used. These conceptions focus on explicit economic, physical or information processing features of the technology. The second class of models, web models, are a form of 'resource dependence' models. They make explicit connections between a focal technology and the social, historical and political context in which it is developed and used. Computer systems, in this conception, are developed, operated and used by an interdependent network of producers and consumers and cannot be analyzed solely according to their discrete features and components. To our mind, Kling in his web model emphasizes the wider range and longer paths of the impacts of computing systems than done normally in discrete-entity models.

The inability to include many benefits and costs during ex-ante evaluation is seen as proof in the failure of traditional investment appraisal techniques, therefore, prompting management to consider the broader analysis of appraisal techniques and thus, supporting the development of a taxonomy that can be used as a frame of reference. To enable senior managers to understand the differences, characteristics and limitations that are inherent within many modes of appraisal, Irani and Love (2002) developed a selection of appraisal techniques within six taxonomies (Figure 5.5).

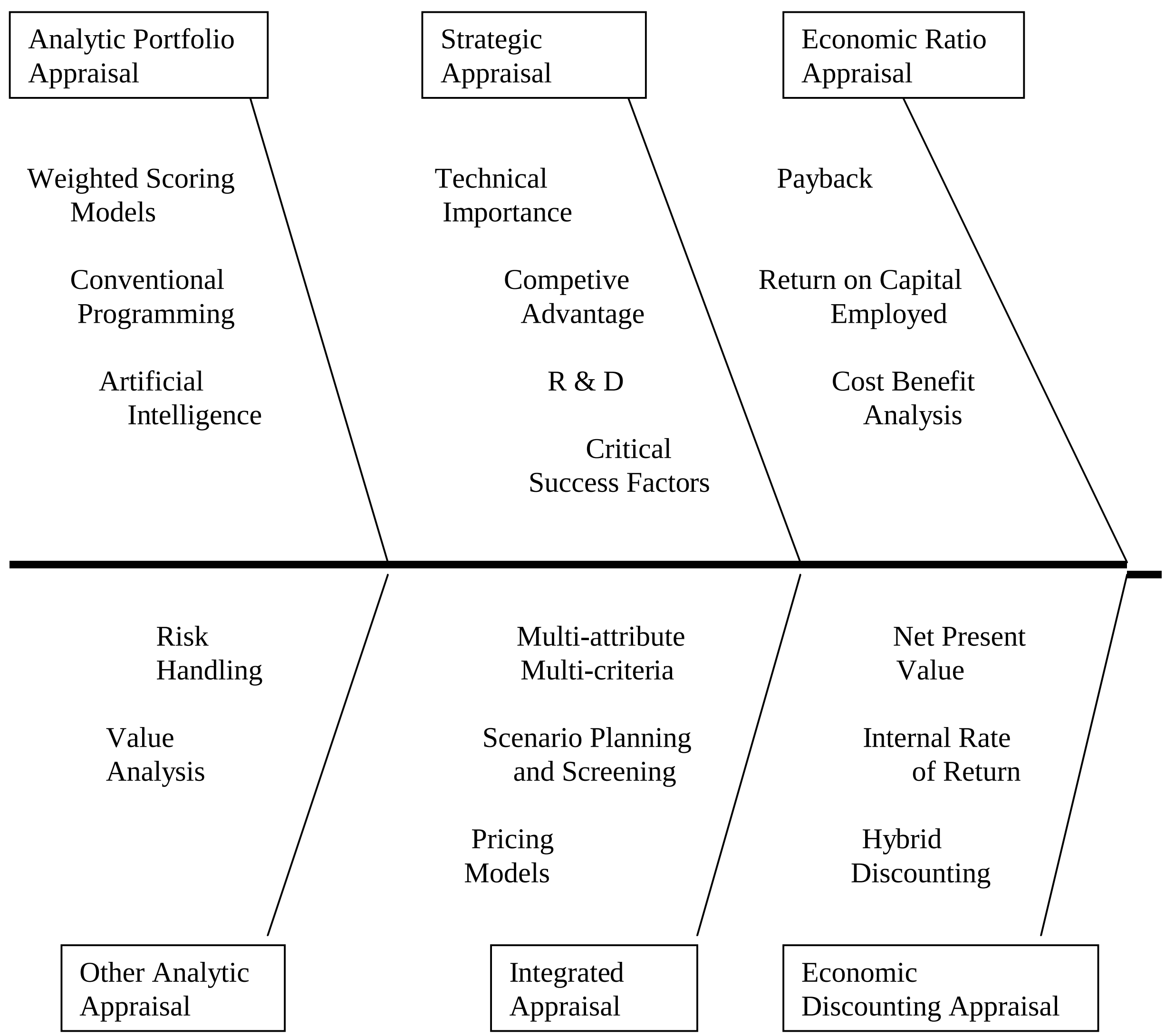


Figure 5.5. Taxonomy of investment appraisal techniques (references excluded, can be found in the article)

Some new metrics

March and Smith (1995) described their view on evaluation as follows: "Once metrics are developed, empirical work may be necessary to perform the evaluation. Constructs, models, methods and instantiations must be exercised within their environments. Often this means obtaining a subject group to do the exercising. Often multiple constructs, models, methods or instantiations are studied and compared. Issues that must be addressed include comparability, subject selection, training, time and tasks." To our mind, March and Smith prefer an experiment on one occasion. Many implications of a certain artifact will appear during the long use of the artifact. We would like to enlarge the set of metrics by considering the longer period, too.

Perhaps the most popular effectiveness measure, the *cost/benefit* of a certain artifact can be estimated during the longer period of use. The total costs and benefits cannot be known until at the end of the life cycle. Counting costs and benefits is not so simple. In fact, Virkkunen (1951) had 5 problems: 1. The range problem – which costs and benefits are included?, 2. The measurement problem – how to measure costs and benefits?, 3. The valuation problem – how to give value to costs and benefits?, 4. The division problem – how to divide costs and benefits to products and services?. The latter can be divided into two sub-problems: The allocation problem – how overhead costs are allocated to products and services?, and The periodization problem – how is a lot cost divided to periods?

In the use phase (after the building phase (b)) the system can be in operation (o) or out of operation, i.e. under maintenance (m).

```
|bbbbbbbb|ooo  oo  ooo oo    oo     oo |
         |    m   m    m  mm   mm   m|
```

Lientz (1983) summarized the 1977-1979 surveys on software maintenance. A first result of the exploratory survey was that maintenance and enhancement were found to consume approximately half of the system and programming personnel hours. A second finding was that approximately 60 percent of the maintenance/enhancement effort was for perfective maintenance. This finding was somewhat unexpected since the literature had supported the belief that fixing problems and keeping systems operational were the major concerns. A third finding was that problems of a managerial nature dominated those of a technical nature in the view of the respondents.

Lientz (1983) paid attention to how to measure a system while it is undergoing maintenance. To explore sources of potential change, the environment of an application system must be considered. The environment consists of four factors, each of which can affect a system: user-external environment, technological change, managerial factors and marketplace. Lano and Haughton (1992) identified four main forms of maintenance activity:

- Corrective maintenance: eliminating errors in the program functionality.
- Adaptive maintenance: modifying the application to meet new operational circumstances.
- Perfective maintenance: enhancement (new operations and refinements to old functions)
- Preventive maintenance: modifying a program improve its futile maintainability.

By referring to Foster (1990) Lano and Haughton informed that the costs of maintenance activity have been estimated as being as high as 80% of the long-term cost of developing and maintaining systems; and this proportion is rising. Hence, we propose that costs of maintenance, and the long-term cost of developing and maintaining systems should be included into the set of software metrics. We also note that our proposal integrates both building and use phases of an artifact, in this case software. We could not find any similar integration in March and Smith (1995).

Four evaluation generations

Evaluation can be compared and related by using Heiskanen's dissertation (1994, p. 166). He took different views on evaluation from the learning and teaching domain. “A most suitable starting point for understanding the subtleties of the construction of an assessment procedure of a social or an organizational system is provided by Guba and Lincoln (1987, 1989) who argue that the evaluation of social and political programs has undergone three generations, and the fourth one is emerging.” During the first generation the evaluator was a technician (Guba and Lincoln 1987, p. 203).

"The first generation began to take shape in the early part of this century... [It] may be characterized as a technical process. During this period evaluation meant little more than measurement - determining the status of individual pupils with respect to norms that had been established for certain standardized tests... Evaluation was seen as a means of determining whether pupils measured up to the "specifications" that the school had set - largely college preparatory specifications."

The second-generation evaluation can be characterized by description of patterns and strengths and weaknesses with respect to certain stated objectives. Information collected during the evaluation process was then utilized to guide refinements and revisions: formative evaluation (Guba and Lincoln 1987, p. 207). The role of the evaluator was that of a describer.

The perspectives in the two first generations appeared to be too limited. The evaluation was required to contain also judgements (Guba and Lincoln 1987, p. 206 and 207). In the third-generation evaluation the evaluator is seen as a judge who was chosen precisely because of his or her connoisseurship qualities.

In the fourth-generation evaluation the evaluator has the role of negotiator and change agent. The fourth-generation evaluation is guided by several principles that are generated by considering five more or less axiomatic concepts: value pluralism, stakeholder construction, fairness, merit and worth, and negotiation (Guba and Lincoln 1987, p. 209). The fourth-generation evaluation is organized by the claims, concerns, and issues of stakeholders. It utilizes a constructivist paradigm (Guba and Lincoln 1989, p. 71), or in other words, this fourth-generation evaluation is explicitly linked to an interpretive approach (Walsham 1993, p. 166). The evaluation scheme proposed by March and Smith seems to represent the first or the second generation above, and the higher generations need another view with different philosophical presuppositions.

Dahlbom and Mathiassen (1997) studied the future of computing profession. They then analyzed the professional focus of computing and found three foci: 1. "*Artifact focus*: The computing profession is concerned with technical, computer-based artifacts developed for individuals, organizations, and markets. Questions of quality address the artifacts themselves, and primarily their technical functionality. *Culture focus*: The computing profession is concerned with computer-based artifacts in practical context of their use. Questions of quality concern quality in use, the way artifacts fit organizational contexts, the way they influence and are influenced by, individual practice and organizational culture. *Power focus*: The computing profession is concerned with the role of computing in changing society and the lives of people. Questions of quality concern the impact of artifacts on the distribution of power, autonomy, integrity and democracy." We above mainly concentrated on the artifact focus. Dahlbom and Mathiassen (1997) recommend the wider range of implications of computer-based artifacts. They only give some preliminary impacts and hence more research is needed.

Table 5.2. Evaluation criteria and views for the use of research outcomes from different aspects

Re-search outcome	March and Smith and their ‘universal’ metrics	Our supplements for March and Smith	Our new metrics
Con-struct	completeness, simplicity, elegance, understandability, and ease of use	cf. Boland and Tenkasi (1995) communication (language game and conduit) and cognition (narratives and information processing)	
Model	their fidelity with real world phenomena, completeness, level of detail, robustness, and internal consistency	form and content (Järvinen 2001); richness of knowledge representation (Yap and Bjørn-Andersen 1998)	
Method	operationality (the ability to perform the intended task or the ability of humans to effectively use the method if it is algorithmic), efficiency, generality and ease of use	application domain (Mathiassen and Munk-Madsen 1986) driving and blocking mechanisms (van Aken 2004)	
Instan-tiation	the efficiency and effectiveness of the artifact and its impacts on the environment and its users	emergent changes with positive and negative unanticipated outcomes (Orlikowski 1995); in addition to economic, technical and physical impacts, also impacts on social, political and historical contexts (Kling 1987); investment appraisal techniques (Irani and Love 2002)	cost/benefit – range, measurement, valuation, allocation and periodization problems (Virkkunen 1951); corrective, adaptive, perfective and preventive maintenance (Lientz 1983, Lano and Houghton 1992); division of power (Dahlbom and Mathiassen 1997)

To summarize our considerations concerning the evaluation of the use of construct, model, method or instantiation we have collected our results to Table 5.2. We reserved one row for every research outcome. To compare with March and Smith we gave one column for their metrics and two other for our own. We present our supplements and our own metrics.

Demolition

This sub-section is based on assumption that every artifact has its birth, build, use and death, and also the last stage must be studied. As we above wrote there are two alternatives either a transition from use of the old artifact to use of the new one or the finish of the use, definitively. We shall consider these alternatives by analyzing different resources (technical, human and data), from which the expired artifact is composed of.

In computing field, hardware normally goes out of date fast. The outdated working procedures must unlearned and the new ones learned, if we have the transition alternative. If the difference between the old and new procedures is small, a human being can return to his old habits, which may cause troubles and failures. The change from the old organizational structure to the new one can have different forms as Van de Ven and Poole (1995) explained. The old data resources can be utilized in the transition alternative, and some conversion or reorganization might be necessary. But in the case that use of the old artifact is totally dropped out, we must take care of rejection of data in order to prevent them to fall into the hands of wrong people, for instance, we must take care of a privacy of the persons whose data are in our expired data base.

Table 5.3. Evaluation criteria for the demolition

	March and Smith	Our proposals
transition		Total costs and benefits, software components etc. can be reused (Frakes and Terry, 1996)
quit		Total costs and benefits, software components etc. can be reused (Frakes and Terry, 1996), deletion of the expired data may cause problems

The software developed for the old application can sometimes be reused. It often concerns software components (Mili et al. 1995), but Frakes and Terry (1996) showed that architectures, source code, data, designs, documentation, estimates (templates), human interfaces, plans, requirements and test cases could also be reused.

We have shown that in both alternatives (transition and quit) the old system or its components have some value and need some after-treatment. At the end of the old artifact its total *costs and benefits* during its whole life cycle can be calculated.

Discussion

Researchers can use our recommendations in the innovation projects under study. To our mind the researcher can then be in four different roles. Firstly, he can be a stakeholder or a player, in other words he is building a new innovation partly for himself. He can then apply a reflection-in-action approach (Heiskanen and Newman 1997, Heiskanen 1994 and Schön 1983). Secondly, a researcher can be solving the practical innovation problem together with managers and user in the firm, i.e. he is performing action research will be described in Section 5.3 (Cunningham 1997, Clark 1976). Thirdlly, Lanzara (1999) told that he as researcher did not participate in the project but gave his interpretation to project members. Fourthly, a researcher can also be outsider, he follows the innovation project as an observer, collects data and describes the results. The description can stay at the very high level, e.g. as citations from manager interviews presenting their hopes and desires (cf. Yap and Bjørn-Andersen 1998).

Our purpose is to continue this analysis by considering other approaches, too. Sweeney et al. (1993) proposed a usability evaluation framework for human-computer interfaces. They used the three approaches (user-based, theory-based and expert-based) in their framework. Their idea on expert-based evaluation might also need further research in connection with studies in the build activity.

March and Smith emphasized the ability to build a certain artifact, i.e. successful cases. To our mind an unsuccessful case can, however, give a new knowledge, and it should therefore be accepted to be an object under study.

On structuring a report of an evaluation study

Our tentative disposition is one simple alternative that a researcher can change according her will.

1. Introduction (see Section 9.2)
2. Selection of the object for evaluation - reasons
3. Development of an evaluation framework (dimensions of utility in relation to the innovation and its usage; selection of other criteria and excluding other ones)
4. Development and validation of measuring instruments (cf. Section 7.2)
5. Performance of evaluation and results
- ex ante (the innovation is not yet realized or used)
- the pilot use (realistic) in the other environment, references, evaluation of error space, simulation, thought experiment, testing boundaries (cf. testing the program)
- comparison of (two or more innovations) – as similar as possible (controlled) environments
- ex post (after the use of the innovation has reached an equilibrium, i.e. the learning phase is over) the criteria also contain the target state used in the development and/or in purchasing

6. Discussion (Section 9.2)

During the building and evaluation processes of the artifact different criteria might be applied, and this may lead exceptional results. This may change the disposition above.

5.3 Action research

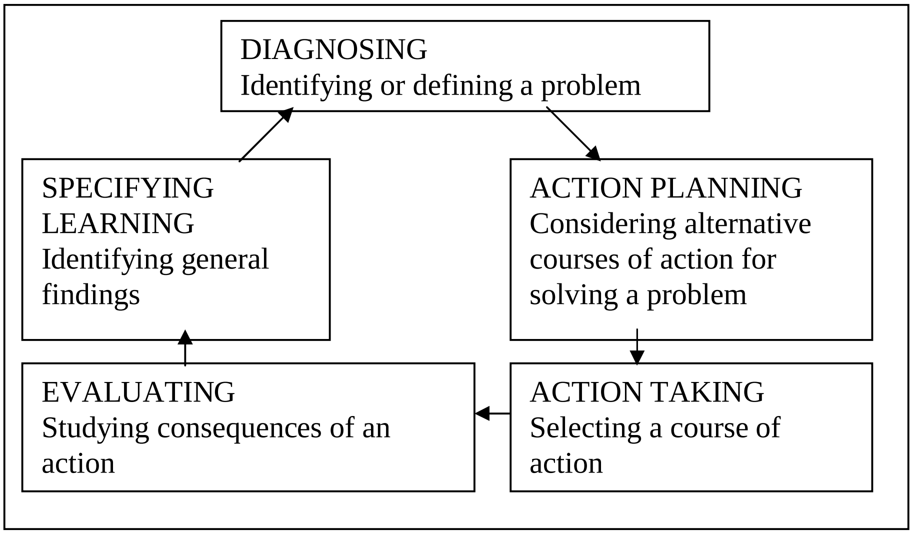
We recommended in the building section that the even totally new artifact should be evaluated, i.e. its utility to be considered. In this section our consideration differs from that, because we here assume that both building and evaluating sub-processes closely belong to the same process. The research method having this characteristic is called action research.

Rapoport (1970) identified four streams of action research development: 1. *The Tavistock* stream of experience brought together psychologists and social anthropologists with psychiatrists of a psychoanalytic orientation. During and immediately after the war (The Second World War) they conducted a number of successful experimental action-programmes in personal selection, treatment and rehabilitation of wartime neurosis casualties and of returning prisoners of war. 2. *The Operational Research* stream of work was of a multi-disciplinary mix emphasizing mathematics, engineering and the physical sciences rather than the psycho-biological sciences. It too, emerged in the wartime efforts, to which it made 'heroic' contributions, particularly in relation to logistical problems of various kinds. 3. *The Group Dynamics* stream emerged from the work of Kurt Lewin and his followers. They have generated a number of studies with emphases on leadership, power, group dynamics, stress and identity. 4. *The Applied Anthropology* stream was another development with an action research emphasis. It gained considerable impetus during the war in relation to problems of psychological warfare, intelligence and administration of occupied territories. – Rapoport also defined *action research* as the method which aims to contribute both to the practical concerns of people in an immediate problematic situation and to goals of social science by joint collaboration within mutually acceptable ethical framework.

Oquist (1978) analyzed the kind of knowledge action research produces and its relation to different schools of philosophy of science. Action research is the production of knowledge to guide practice, with the modification of a given reality occurring as part of the research process itself. Within action research, knowledge is produced and reality modified simultaneously, each occurring due to the other. Oquist presented the reasons why action research does not belong to Empiricism, Logical Positivism nor Structuralism, but it seems to belong to Pragmatism (cf. Goldkuhl 2004) and Dialectical Materialism (some of those schools are shortly described in Chapter 10).

Susman and Evered (1978) characterized action research (A.R.) with six properties: 1. A.R. is future oriented; 2. A.R. is collaborative; 3. A.R. implies system development; 4. A.R. generates theory grounded in action; 5. A.R. is agnostic and 6. A. R. is situational. According to Susman and Evered those six properties provide a corrective to the deficiencies of positivist science. Our classification scheme confirms the finding above, because the positivist science is classified into chapters 2 and 3 and gives answers to the question what is a part of reality? Action research belongs to Chapter 5 and gives answers to the question How to improve an artifact?

Susman and Evered (1978) described the cyclical process of action research (Figure below).



Hult and Lennung (1980) performed a wide literature survey and then formulated a new definition of actions research: *Action research* simultaneously assists in practical problem-solving and expands scientific knowledge … as well as enhances the competence of the respective actors … being performed collaboratively … in an immediate situation … using data feedback in a cyclical process … aiming at an increased understanding of the totality of a given social situation … primarily applicable for the understanding of change processes in social systems … undertaken within a mutually acceptable ethical framework. - The suspension points (…) differentiate the parts Hult and Lennung widely described. – Action research is sometimes also called action science (Argyris et al. 1987).

Kalleberg (1995) identified three research designs for starting action research. In each of them the primary focus should be on an existing system that could (feasibility) and should (desirability) be transformed:

1. *Inspection*. We may ask if there is something to learn from a comparable, existing unit, or a unit that has existed before. We try to learn from some variation in social reality (news from somewhere).

2. *Imagination*. We imagine a non-existing, "utopian", but feasible and desirable alternative (news from nowhere).

3. *Intervention*. We intervene together with others, in order to improve the unit at the same time as we study it (learning from doing).

Kalleberg presents three challenges for the second design (imagination). Firstly, the alternative sketched should be creative enough. Secondly, it should also be realistic enough, and thirdly, the new arrangement should be the desirable one in comparison with an existing one. To consider the last challenge Kalleberg proposes simulated discourses and mental experiments.

On structuring a report of action research

We propose that a researcher would describe her action research process in detail enough in order to help a reader to understand the situation.

1. Introduction (see Section 9.2) (description of the primary and potential secondary task)
2. Description of the research site / environment / organization / problem area / initiative
3. Description of research process (how many times and in which ways the cycle: Diagnose, plan, implement, evaluate and learn was repeated)
4. Collecting and presentation of findings in one section
5. Discussion (Section 9.2)

Some special properties of a particular action research project can be taken into account by giving an own section for that property.