Go Back

Musings about the effectiveness and evaluation of business simulators

Andreas Größler

Industrieseminar, Mannheim University
D - 68131 Mannheim
Germany

Phone: +49 621 181-15 83
Fax: +49 621 181-15 79
E-mail: agroe@is.bwl.uni-mannheim.de

Abstract

Evaluation research on business simulators is still in its infancy, both inside and outside the system dynamics community. Experiences made with other instructional media (for instance, interactive video) and with non-computer-based simulation games suggest that this is to a great deal caused by asking the wrong questions. The paper presents a conceptual model of human learning that can be used as a basis for evaluation research of business simulators. Furthermore, it is discussed why questions about the absolute efficacy of such simulation tools cannot be answered generally. In the same way, comparisons with other teaching methods are not fruitful. The only evaluation approach open for rigorous experimentation is testing of business simulators which are systematically varied in one feature. Issues concerning this approach are discussed at the end of this paper.

Key words: business simulator, evaluation, effectiveness, simulation, system dynamics

The effectiveness and efficiency of business simulators is still an unsettled issue of research. In order to build a comprehensive body of corresponding literature, common methods and research concepts are needed. For this purpose, different definitions of effectiveness and different forms of business simulators have to be discussed and evaluated.

In the past, the question of effectiveness has arisen with other teaching and training media as well. Technologies as interactive video offer examples of past discussions about reasonable goals for evaluation research and appropriate methods to conduct this research. A transfer of insights generated in these fields can help to indicate best practices and uncover issues relevant to evaluation research for business simulators. We can assume, for instance, that an evaluation of business simulators has to measure knowledge as well as performance. Another implication of previous research is that effectiveness questions cannot be answered if they are too general. In addition, numerous methodological problems can be identified, which have occurred in previous evaluation studies.

In the beginning of the paper, possible effects of business simulators are summarized from the literature. After that a model of human learning is described. The aim of the description of this model is not to explain human learning, but to find starting points for evaluation research. Next, the topic is discussed why questions about the absolute efficacy of simulators cannot be answered generally. Also, the fruitlessness of comparisons with other teaching methods is addressed. The implication is then presented, that the only evaluation approach that is apparently open for rigorous experimentation is that of testing business simulators while systematically varying exactly one of their characteristics. A description of suitable experimental designs follows. The paper finishes with a list of possible issues when conducting evaluation studies with business simulators.

Assumed effects of business simulators

Business simulators are computer programs that simulate socio-economic scenarios with the help of a formal model, for instance a system dynamics model.[1] Users of business simulators usually play the role of the management of an enterprise. Together the formal simulation model, a human-computer interaction component and further (gaming) functionality build the three basic aspects of a business simulator as depicted in Figure 1.[2] A major assumption when using business simulators is that they cause a learning effect. This is their main goal, however, other possible areas of use are conceivable as well (for instance, personnel selection, entertainment, etc.).

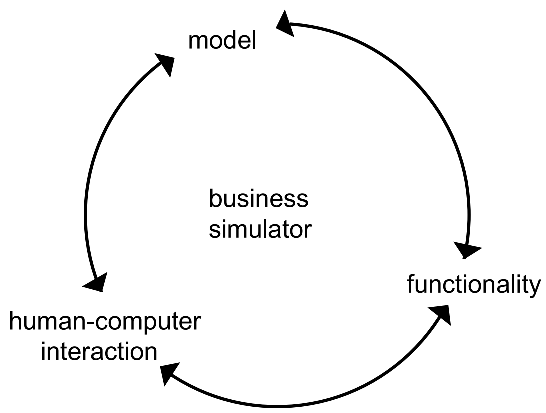


Figure 1: Three aspects of business simulators (Maier and Größler 2000)

Why business simulators could be seen as effective tools to make people learn about dynamic cause-effect-relations can be concluded with the help of two lines of argumentation. The first argument compares business simulators with (the disadvantages of) conventional teaching methods. The other argument discusses current strands in theoretical instructional research and their relation to the instruments actually in use for teaching.

Stein Greenblat (1988) identifies five disadvantages of conventional teaching (she takes a lecture with discussion as an example). These disadvantages are contrasted by characteristics of simulation games, which could help solve the problems identified before (adapted in Table 1).

In a similar way, Richmond (1993) contrasted a teacher-oriented style of teaching/ learning with student-oriented teaching/learning, which emphasizes the role of an active learner. In Richmond's view, business simulators are tools to support this paradigm of student-orientation.

Disadvantages of lectures (with discussion)	*Solutions by business simulators*
Learners are mostly passive	Learners have to be active ("learning by experience")
Material must be presented in a strict, sequential order that is determined by teacher	Only task determines order of material that can, therefore, mostly deliberately be chosen by the learner
Discussions are characterized by social hierarchies or what is socially desirable	Discussions are characterized by experiences made with simulator
Systemic, holistic points of view are difficult to mediate	Students are forced to achieve a systemic view in order to manage the simulators successfully
Verbal descriptions are interpreted differently	Terms are defined by their use in the simulation

Table 1: Comparison of conventional and teaching with business simulators

Following a more theoretical approach towards the supposed effectiveness of business simulators, Stein Greenblat (1981) lists six propositions of teaching theory that seem to illustrate the adequacy of simulation games for teaching complex issues:

1. The human cognitive apparatus is not seen as a container any more that can be filled with knowledge, but knowledge is constructed within the brain (constructivism; Duffy and Jonassen 1992).
2. More important than "static" knowledge is the ability and the motivation to learn.
3. Active exploration and self-controlled search for learning resources are considered crucial to avoid inert knowledge.
4. Learning has no value itself but must be goal-oriented.
5. Learning occurs when students are actively engaged with objects.
6. Finding and analyzing relevant information in an abundance of data is necessary to understand complex systems.

In this sense, business simulators can even be superior than reality because they provide possibilities for experiential learning with direct and immediate feedback about the consequences of decisions (Lane 1995). According to Goodyear et al. (1991, 274) these seemingly positive effects of business simulators can be summarized in a way that

„simulation-based learning is usually expected to motivate, to invite active and deep processing of subject matter, to allow for systematic exploration, for fruitful failure, and for unlimited practice, all of which should contribute to better learning outcomes, reduced learning time, or both.“

Learning as the basic goal of business simulators

Although the arguments laid down in the previous section are suggestive, they are no proof for the teaching effectiveness of business simulators. There is only little evidence in the literature that simulators are as effective as they are supposed to be (Lane 1995). Questions of the effectiveness and efficiency of business simulators are still an issue of research. Although there are some hints that these programs promote learning (Milling 1995) their validity has not been proven yet. Most papers about this topic only report on anecdotal data and not on empirical evidence. Those papers, which try to collect empirical evidence, often show methodological shortcomings.

However, in order to justify the further and extended use of business simulators, evidence is needed, if and how they can enhance decision making through learning (Bakken, Gould and Kim 1994). Even anecdotes and observations collected over decades can be misleading and are open to systematic errors (Doyle 1997). Accumulation of reliable results is a precondition of scientific progress (Richardson 1996). The basis of this endeavor are common research concepts, definitions, methods and procedures.

A major aim of the evaluation of business simulators is the investigation of their validity. A learning tool is valid if it is suitable and relevant for its predetermined purpose (Cronbach 1990). Two types of validity are distinguished: internal and external validity. A simulator is internally valid if its application shows sound and comprehensible behavior. A simulator is called externally valid if the insights generated from its application can be transferred to other systems, especially reality (Bortz and Döring 1995). External validity can be distinguished from fidelity which indicates an objective similarity between simulator and reality that is independent from the learning purpose of the simulator. Fidelity, however, affects external validity in such a way that too low fidelity as well as too high fidelity can hinder learning. The appropriate degree of fidelity of a business simulator depends on the type of learning that is to be accomplished (Prensky 2001).

Evaluation studies of the effectiveness of every training measure can be divided into two categories: formative and summative evaluation (Hays and Singer 1989). Formative evaluation should take place during the development of business simulators, in order to assure internal validity and an appropriate degree of fidelity. Summative evaluation, however, happens while actually employing a business simulator. In this case, the focus lies on the learning process, not the tool, and on external validity. Like in most studies, this latter form of evaluation is referred to in this paper.

The basic assumption of the work reported here is that business simulators are mostly used to promote learning. Generally speaking, this means that they are effective if users *learn something*. The term “learn” will be under discussion in the rest of this section but, firstly, a clarification of what can be learned from these tools seems necessary.

In most cases, it is not clearly defined what users are supposed to learn when they fly a simulator. This is true in two ways: it is neither made explicit for the users nor have the developers or facilitators of simulators clarified it for themselves.[3] In other words, usually business simulators do not have clear learning objectives. Nevertheless, it is often stated that users should learn about the domain and about general behavior modes of complex, dynamic systems (Funke, J. 1991). Similarly, users of business simulators often seem to be able to articulate general statements about what they have learned (Greenblat 1981). In terms of Andersen et al. (1990) these learning goals (and subjects' estimations of them), however, would only be "meta-purposes". Considering the task of evaluation, it can be agreed with Andersen et al. when they continue with an appeal to make "meta-purposes" more concrete. Without a clear statement of learning goals any evaluation becomes both, impracticable and superfluous. (For a contrary opinion see Jonassen 1992.) Thus, even though users do not have to be informed about the learning goals of a simulator, they implicitly have to follow them if learning success is to be measured. If the users arbitrarily or unconsciously try to aim at other goals than the learning goals intended, evaluation cannot be successful.

In a few studies, users' perceptions of the effectiveness of simulators are assessed for evaluation purposes (i. e., perceived, not actual effectiveness.) Such an assessment usually is rather easy to conduct (for instance, with post-gaming questionnaires or interviews). The idea behind this is that high perceived effectiveness might also lead to high actual effectiveness due to motivational reasons (Klein and Fleck 1990). Although this sounds persuasive a rigorous evaluation must not stop at this point. A study by Kaufman (1976), for instance, showed that positive attitudes towards a simulation game do not naturally lead to a better performance concerning learning goals. Objective data is necessary. To measure perceived effectiveness, therefore, can only be an additional element (or a first step) of evaluation research.[4] Thus, the effectiveness of business simulators for promoting learning has to be based on more rigorous foundations.

A clear definition of learning is one basis for evaluation research. However, such a definition, which is universally accepted and applicable, has not been given in the literature so far. Avoiding the purely behavioral as well as the purely cognitive approach (Lefrancois 1982) it is defined here as to include two processes and the result of these processes[5]:

1. a change in cognitive structure, and
2. a change in behavior,

which lead to

3. a change in performance.

Starting with the first point, learning can be defined as the increase or enrichment of the complexity of cognitive structure or knowledge (Figure 2). Following Ashby's Law of Requisite Variety (1970) this enrichment allows to succeed in complex and changing environments. Thus, learning is seen as a flow of information entities that are consistently interwoven with already existent knowledge to broaden the repertory of behavior.

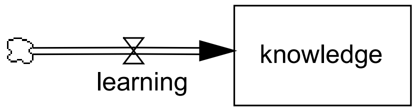


Figure 2: Learning as enrichment of cognitive structure[6]

An increased complexity of cognitive structure gives the potential to perform better. In business contexts, learning and the improvement of cognitive structures have no value per se. What is intended by learning is to induce better performance. However, performance has no direct link to the cognitive structure. It is mediated through behavior. This idea leads to Figure 3.

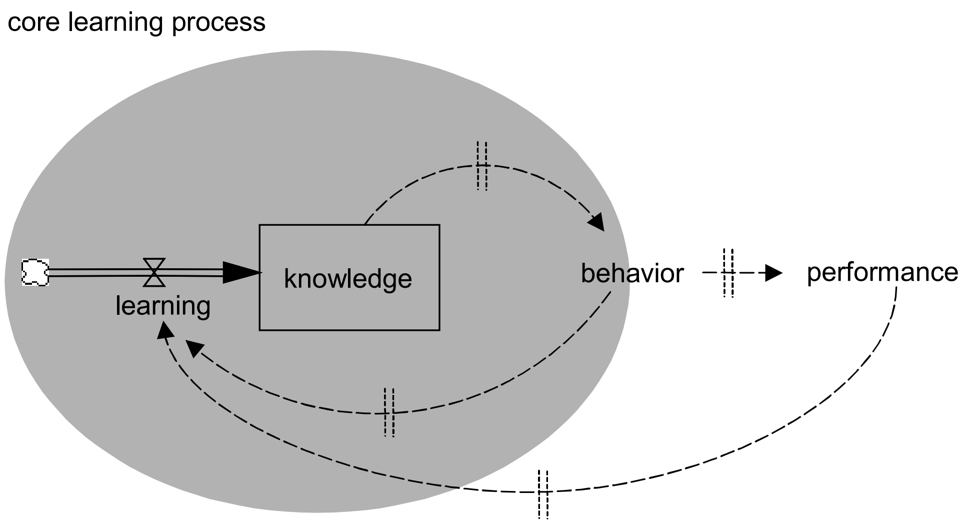


Figure 3: Three elements of learning

In the same way as the last figure, Figure 3 abstracts from important variables which certainly influence the learning process. For example, conceivable influences of intelligence, motivation or already acquired knowledge are not depicted (e. g., Beckmann and Guthke 1995). Also, the question whether knowledge can be mediated objectively (objectivism) or is constructed within the individual (constructivism; Duffy and Jonassen 1992) is not discussed here. Furthermore, the ability of an individual to observe the environment and effects of his or her behavior is not included in the model because it seems not relevant in an evaluation context. A more concise model of learning, however, would need to include a few or all of these variables.

Note that changes in cognitive structure and behavior are interdependent as shown in Figure 3. Cognitive preconditions influence how an individual behaves. But behavior just as well shapes the cognitive structure of an individual ("learning by doing"). Changed cognitive structures must not necessarily and not immediately cause changes in behavior and vice versa (Huber 1990; Argyris and Schön 1996). Different types of learning barriers can be identified (Kim 1993).

Similarly, the connection between the core learning process and performance is often distorted, fuzzy, and delayed (Broadbent and Aston 1978; Berry and Broadbent 1995). That means that not every learning which takes place has an (immediate) effect on performance. Although it is usually a strong indicator, not every change in performance is related to learning (for instance, there could be random environmental influences). The same is true vice versa: not every change in performance affects knowledge and, indirectly, behavior (because, maybe, the individual simply has not detected the changed performance).

Performance is an indicator, how well a certain goal is achieved, for example, the goal to manage a company. Performance can also be seen as the ability to solve problems. How behavior (action) is linked to performance (profit) is a key research issue in the management sciences (Epstein and Westbrook 2001). In the given definition of learning, performance is outside the core learning process. It is the outcome of both: indirectly (via behavior) of changes in cognitive structure and directly of changes in behavior. However, particularly in the business field, learning is then highly desirable if it yields an improvement in performance. A clear statement that learning has no value by itself but has to result in better performance is therefore important. (See Garvin 1993, for the same argument aiming at Organizational Learning.) This led to the inclusion of performance as the third element of learning. In this definition, behavior plays the role of the "middle man". It is neither the basis nor the result of learning but necessary as mediator.

If performance were the unbiased outcome of intra-individual processes, performance measurement would be a valid and complete estimation of the learning process (Bakken 1989). Neglecting any environmental influences and any delays or barriers in cognitive processing, a change in the learning process would have a direct effect on performance. This would be a convenient situation because usually performance measures are "naturally" given (for instance, profit of a company, marks in school, etc.) Such measures could therefore easily be obtained. In contrast, especially changes in the cognitive structure are only measurable through indirect measures with all the problems connected with that approach (Folley 1967). Nevertheless, evaluation needs to assess changes in cognitive structure because there is no direct link between cognition, behavior and performance.

In addition to this distorted relation between performance and learning, performance measures usually are—although ubiquitously used—not unimpeachable. For instance, management's performance cannot easily be quantified using cumulative net income. It has various facets and changes dynamically (Bakken, Gould and Kim 1994). A more secure way to measure learning, therefore, is to additionally evaluate the process of changing cognitive structures and behavior (Kirkpatrick 1967).

Issues which are only roughly discussed in this paper are questions of the duration and the consistency of changes in cognitive structure or behavior. It seems clear that a single, short-term change in behavior hardly can be assumed to be part of a learning process. Thus, learning is supposed to cause consistent changes in cognitive structure and behavior of an individual. This change has to last a certain time span.[7]

Applied to business simulators, learning means that the users' cognitive structure changes as well as their behavior and their performance do. The performance goal is to control the simulator successfully (neglecting, at this point, carry-over effects to other systems). Thus, performance can be directly defined as controlling ability. Usually, business simulators contain no external or random influences and users are the only ones who determine performance through their behavior. Therefore, performance is a valid, direct measure of behavior (if valid scores for performance are defined).

However, it is not enough to measure performance because the relation between cognitive structure and behavior (and therefore performance) is still equivocal as described above. To acquire knowledge is an objective of business simulators. This effect cannot be measured using only performance scores. Because of these problems it can be concluded that both, cognitive structure and performance, have to be assessed in order to account for a complete learning process (Spender 1998). To measure only performance or only cognitive changes is not enough (Doyle 1997). See Figure 4 for an updated learning process with business simulators.

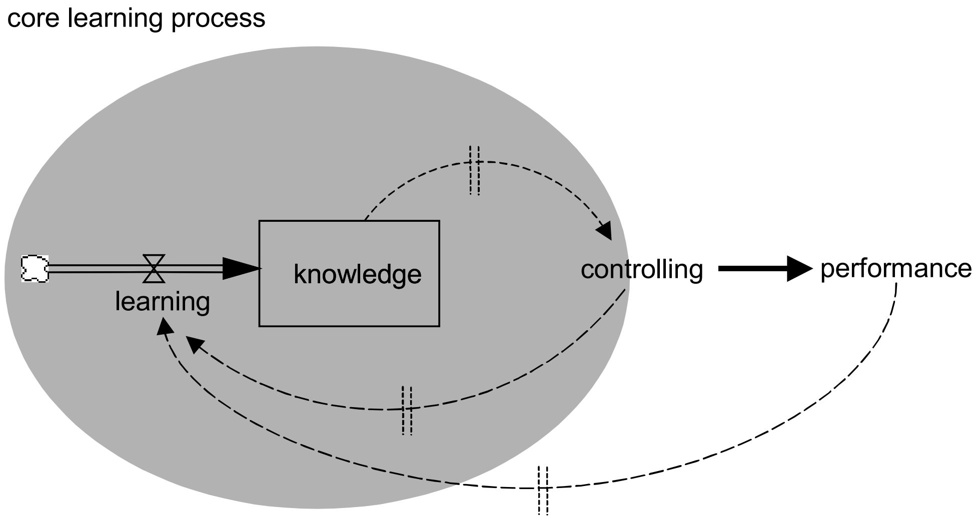


Figure 4: Learning with business simulators[8]

Different Forms of Effectiveness and Efficiency of Business Simulators

As far as users' gaming performance is investigated, business simulators are effective in the short term. If simulators are internally valid users can be supposed to perform better in controlling these instruments after a certain period of time using them. Thus, performance and behavior obviously have changed. But gains in explicable knowledge usually are difficult to prove because the connection between this form of knowledge and controlling ability often is delayed and distorted. This issue led to the discussion about implicit or tacit knowledge (Broadbent, FitzGerald and Broadbent 1986; against the assumption of implicit knowledge argues, for instance, Haider 1993). The term "tacit knowledge", however, indicates that some kind of knowledge acquisition occurs.

Usually, carry-over effects to other complex systems are intended; the transfer of insights is a learning goal. Users should not only be able to control a specific simulator better, but should also perform better in controlling a certain range of complex, dynamic situations. The question is, whether insights, which were acquired through simulation gaming, can be transferred to real life situations (Graham and Senge 1990). In training evaluation research, external validity is basically determined by this transferability (Hays and Singer 1989). Wolfe (1976, 434) describes external validity as "the ultimate test of any teaching method". Chances for a learning transfer to take place increase when the individual manages to create explicit, domain-independent knowledge. Whether, on the other hand, implicit knowledge can be transferred seems questionable. Because the acquisition of explicit (or even domain-independent) knowledge could not be proven in many cases, the capability of business simulators to stimulate a learning transfer has not been confirmed yet.

Connected with the question of transferability is whether users perform better in the long run.[9] There are only few studies regarding this because methodological (and ethical) problems are enormous (Funke, J. 1995a). For example, a better performance over ten years after using a business simulator can certainly not solely be accounted as an effect of this use. (See Norris and Snyder 1982 for an exploratory longitudinal study.)

As another important question remains whether business simulators are more effective than other training measures. For a lot of training methods it can be assumed that they influence people in the way intended. They make people learn. Nevertheless, it is obvious that some methods are better suited in certain situations (in particular, for specific learning objectives) than others. The construct of effectiveness therefore can be reduced to a relative meaning.

However, the question of the advantage of one training measure over another cannot be answered in general and, most of the time, even not for a specific learning situation. For example, let us assume three groups of students: one group is using a business simulator, the second is reading a textbook, and the third is attending a lecture about the very same topic. When conducting an experiment it can be assessed which training group learned the most in relation to the learning goals set. This comparison needs a simple pre and post test experimental design (Cook and Campbell 1979) and some statistical inference (Hayes 1994). Additionally, let us assume that the business simulator group performed better than the lecture group, and this group better than the reading group.[10] Can now be concluded that one of the methods applied is generally more effective than another?

It cannot, because too many disturbing and interfering variables simply cannot be controlled, for instance, the quality of the lecturer, the comprehensiveness of the book or the validity of the simulator (Schulmeister 1996). The only thing which can be concluded is that in this specific situation with the given training methods the business simulator yielded better results than the lecture or the book. Most probably, not all parameter that determine the specific methods used are known to the experimenter. Therefore, results can hardly be generalized to all uses of that simulator, book, and lecture. And it must not be generalized to all conceivable business simulators, books, and lectures.

The problem of such comparative studies is that the methods under observation consist of a sample of characteristics which can be varied in essentially infinite ways (Stolurow 1962).[11] The easiest way to obtain data about a comparison of different training methods—asking users about their opinions and impressions—can lead to objectively wrong and biased results (Marchionini 1989). Hawthorne-effects must be assumed for many comparison studies (Roethlisberger and Dickson 1939). For these reasons, today nearly no comparisons between different learning media are conducted any more (Weidenmann 1993).

Effectiveness can be evaluated independently from the use of resources, which are necessary to achieve a certain effect. In practice, however, the cost to reach a specific learning goal are of vital interest. That is, efficiency matters, too. For example, do people learn faster (with less cost) when using a business simulator instead of reading a book about a subject? The argumentation can be equivalently transformed applying the universal rule of efficiency: given the same cost of two learning measures, does one measure yield better results than the other?[12]

Back to the example, let us assume the students used the business simulator for two hours, read the book for half a day, and attended the lecture lasting 45 minutes. Given the same ranking of effectiveness as above, it is not clear which method is more efficient. Imagine that the simulator group outperformed the lecture group just one percent in relation to the learning goals. Does this small advantage justify the cost of a longer learning time? Statements about efficiency can only be made if equal cost for the three methods are supposed. That is, two hours usage of simulator costs as much as half a day reading and as much as attending a 45 minutes lecture.[13] In this case, the business simulator indeed would be more efficient than the lecture and the textbook.

However, because of the problems of measuring relative effectiveness, this statement is also limited in scope. Again, these results cannot be generalized in a way that the simulator is more efficient than the book or the lecture in any case. And, once more, it must not be generalized to all conceivable business simulators, books, and lectures.

In summary, the (absolute) effectiveness of business simulators is trivial in a short-term view, given that the ability to control the simulated system is the learning goal. The (relative) effectiveness and the efficiency of these tools in comparison to other training methods can hardly be measured out of methodological reasons (Clark 1983). The same holds true for longitudinal studies and studies on learning transfer. What remains is the evaluation of different features of business simulators or different ways of the use of simulators. In this case, forms of simulation tools are compared that differ in just one variable. This allows rigorous experimentation.

Systematic variations of exactly one characteristic of business simulators makes it possible to find effective characteristics by which business simulators can be improved. These characteristics are located in all three aspects of business simulators. (See Figure 1 above.) Design-guidelines for simulators can be tested using this experimental approach if they are operational. (See, for instance, Thurman 1993 for a list of design issues and implications.)

Following this approach, it becomes possible to identify features of business simulators which make them effective in a particular learning situation. In other words: we are not heading towards establishing a theory about the general effectiveness of business simulators, but it is—at least—possible to articulate hypotheses about characteristics of business simulators which cause effectiveness under certain circumstances. These hypotheses can be scrutinized using the evaluation approach of experimenting with business simulators that differ in one characteristic.

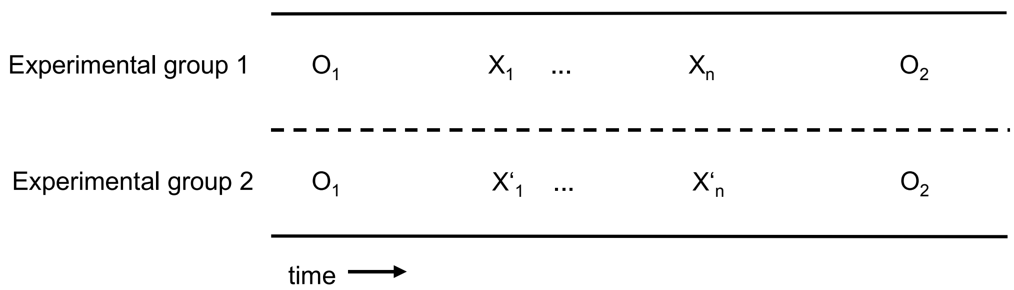
In conclusion, the search for characteristics of business simulators which are effective for specific groups of users in specific situations can replace the search for the absolute value of simulators. Because, as Bosco (1986, 15) pointed out in an analogy, "the question 'Are books effective in providing instruction' does not lead to a categorical answer. Rather, the answer to this question depends upon content of the book, the way it is being used, the objectives of instruction, etc." Similarly, consequences of the use of business simulators depend on various factors.[14]

Designs to Conduct Evaluation Experiments Using Business Simulators

A suitable experimental design to conduct evaluation research with business simulators that are varied in one characteristic is the ‘pre test / post test design with control group’ (Cook and Campbell 1979).[15] The principle structure of this kind of experimental design is depicted in Figure 5. In this figure, experimental group 1 can be seen as a control group that is compared to experimental group 2. The business simulator—which both groups use as an experimental treatment—differs in exactly one characteristic from one experimental group to the other.

Explicable changes in cognitive structure are measured by knowledge tests. Therefore, results for pre and post knowledge tests are compared. The comparison of the results of post knowledge tests between the experimental groups permits us to find differences between the groups that can be supposed to result from different treatments. In order to draw this conclusion, a pre knowledge test is necessary to assure differences do not already occur before the experiment. A pre knowledge test must be applied, whenever it cannot be assumed that subjects are randomly distributed in the experimental groups. Thus, it controls for unequal group characteristics and provides data to correct results statistically if necessary.

In addition to the pre knowledge test, also a biographical test can be conducted to assess other factors in group composition which might have an influence on experimental results, for instance, gender, age, education, motivation, and general intelligence of the subjects. With this data, relevant subject characteristics can be determined and problems of sampling can be avoided or—at least—are made transparent to the experimenter.



O_1: Observation 1 = Pre knowledge test, O_2: Observation 2 = Post knowledge test

X_1 ... X_n: Treatment = Playing with business simulator

X'_1 ... X'_n: Treatment = Playing with business simulator (modified in one characteristic)

Figure 5: Generic experimental design for evaluation studies with business simulators

Interpretation is easiest, if pre and post knowledge tests are equal in content. To mitigate memory effects, questions are often given in a different order in the two knowledge tests. However, sometimes it would not be possible for subjects to answer questions of a pre knowledge test before using the business simulator. This happens, for example, when completely novel knowledge is supposed to be mediated by simulator use. In this case, a pre test can be used which is supposed to correlate with results of the post test. It could, for instance, be constructed to measure only general knowledge about a domain. This modified experimental design is said to work with proxy pre test measures (Cook and Campbell 1979). In order to measure sustainability of acquired knowledge, the post test can be replicated after some time (Bredemeier and Greenblat 1981).

As the discussion above indicated, different forms of knowledge might exist. Therefore, knowledge tests should try to measure these different forms. They can be designed to consist of several parts. Causal diagrams can be used to assess explicable knowledge about system's structure. Questionnaires can help to discover declarative knowledge about facts. Prognosis tests can be used to measure how far this knowledge can be mobilized to estimate the future behavior of the system.

The type of knowledge test discussed so far does not necessarily yield good results concerning the assessment of implicit knowledge and transferability. It does not allow to measure carry-over effects. To do so, subjects are required to control a second (test) simulator, after working with the first (treatment) simulator. Thus, a second simulator can be a part of the pre and post experimental observation. With this modification, also dynamic aspects of knowledge can be measured. However, methodological problems concerning the applicability of one simulator as a valid knowledge test for another simulator can hardly be solved. In summary, the question of construct validity of the knowledge test is most critical to evaluation research (Greenblat 1981, 152: "...it is difficult to tell at this point whether the lack of evidence in support of the propositions stems from poor outcomes or poor measurements.").

Because of the problems when using a second simulator, changes in tacit knowledge are usually measured with the help of performance scores of the business simulator, which is used as a treatment in the experiment. A few independent runs of the simulation game are favorable, in order to achieve reliability estimations. Furthermore, obtaining performance scores from more than one use of a simulator allows to assess changes in performance during the experiment. These changes are supposed to be caused by a learning effect which is not explicable. Whether or not different scenarios (starting points) are advisable for the game runs is an open question: On the one hand, the same scenario in every use increases reliability measures, particularly of measures of internal consistency (for instance, Cronbach's alpha; see Cronbach 1990). On the other hand, different scenarios provide the possibility to obtain measures, which are more independent than pure repetitions of a single game run. In addition to that, this approach can also serve as a first estimation for the transferability of acquired knowledge. The effect that high scores occur after many uses of a business simulator are, however, primarily a sign of internal validity, not necessarily of a transferable learning effect, which took place.

A comparison between performance results of the experimental groups allows to detect systematic differences which can be assumed to result from the varied characteristic of the business simulator. Thus, business simulators are used as measures to mediate knowledge *and* as a method to assess this mediation process. The simulation tool has a double function in many experiments: it serves as treatment as well as being a means to measure the results of this treatment. Methodological problems concerning this point have not been discussed widely so far. See also Funke, J. (1993), for a criticism aiming in the same direction and the claim for a clear distinction between knowledge acquisition and knowledge application.

Issues of Business Simulators as Research Instruments

Business simulators have not only been used for teaching purposes. In a 1993 paper, Brehmer and Dörner describe business simulators as tools to conduct psychological research. In this case, simulators are used to investigate the decision making process of subjects in complex situations. This investigation is done in order to draw conclusions about the human mind, reasoning processes, and the ability of subjects to handle apparently chaotic situations. (For a review of experiments with simulators see, e. g., Funke, J. 1991)

The most prominent reason for using simulators in psychological research is "ecological validity" (Buchner 1995). Unlike in classical experiments, subjects can—with the help of simulators—be confronted with a real-world problem in a context which is as complex as reality. Because simulators are, however, only a "virtual reality", experiments can still be conducted within a laboratory (i. e. controlled) setting. Furthermore, simulators allow experimentation without being confronted with real world consequences. They make experimentation possible and useful, when in the real world situation such experimentation would be too costly or—for ethical reasons—not feasible; or where the decisions and their consequences are too broadly separated in time. Other reasons for the use of simulators are the possibility to replicate the initial situation, and the opportunity to investigate extreme conditions without risk (Pidd 1992; Milling 1996).

Based on the psychological literature, a list of unsettled methodological issues of business simulators can be identified (Funke, J. 1995b; Süß 1999; Keys and Wolfe 1990). It is supposed that—in order to achieve valid evaluations—some topics concerning simulators must be clarified. In the following list, also some brief ideas are presented, how problematic influences on the validity of evaluation studies might be lessened.

1. **Validity of model**. While business simulators are based on formal models, the real-world domain itself usually is not completely open to formal description. Therefore, the validity of the simulation often depends on the ability, knowledge, and experience of the modeler. For the congruence between simulator and reality, this is in particular true when qualitative relations are to be modeled (for instance, the relationship between image of a firm and its market share). Nevertheless, in some experimental contexts, fidelity is not necessary (and not desired, for example, in order to suppress influences of existing knowledge about a domain). In contrast to fidelity, internal validity of the simulator, is a prerequisite for any use as a research tool. While this is an issue that cannot be solved ultimately, it can, nevertheless, be mitigated by a careful validation process of the business simulator (Barlas 1996) and by thorough education of modelers and designers.
2. **Level of abstraction**. The right level of abstraction and detail of a simulation cannot formally be determined. Which level of detail, which information is necessary to understand a scenario, which is superfluous? This point is connected with not fully understood effects of information about the context of a scenario. Instructional design might provide answers or more detailed decision heuristics in the future concerning this basic question.
3. **Handling of time**. The "compression of time" in business simulators could affect users. For example, the process of planning and controlling a scenario might also be

compressed in comparison to real world decision making processes. (On the question of different types of "time" in business simulators, see Größler 1999). The validity of a score based on control performance, therefore, can be doubted. However, "compression and expansion of time" are major advantages of business simulators (Kim 1989, 327) which cause, on the other hand, problems with experimentation. Business simulators with adjustable time frames can be used to investigate this problem (Größler 1999).

4. **Difficulty of task**. Some simulators are so complex (regarding number and interconnectedness of variables, dynamic behavior, handling of user interface; Packer and Glass-Husain 1997) that subjects are not able to control them. They just use trial-and-error strategies while gaming. Performance scores are not valid in this case. Lane (1995) discusses the trade-off between fidelity (or ecological validity) and gaming character of simulators. See Hays and Singer (1989) for a definition of fidelity. The problem might diminish with advances in instructional design theory.
5. **Cul-de-sac situations**. Some simulators tend to get stuck in situations without the possibility to improve it, although subjects recognize their errors. See also Brehmer (1992). In contrast to that, some errors cannot be observed because the corresponding effects occur much later or in another area of the simulated system (Goodyear et al. 1991). This issue can be solved applying a rigorous and thorough testing of the business simulator using standard procedures from software development (see, for instance, Sommerville 2001).
6. **Different cognitive processes involved**. In different phases of using a simulator different cognitive processes could prevail (Funke, U. 1991; Reigeluth and Schwartz 1989). To take this into account, different measures for these different phases might be needed. But firstly, a common psychological theory of these cognitive processes has to be articulated. Then, different scores for these different processes can be implemented within business simulators.
7. **Ambiguity of (process) scores**. It is argued that static scores (outcome measures) hardly contain information about the process of flying the simulator (for instance, Bakken 1989). Process measures (like, e. g., the strategy that a subject followed), which are used to eliminate this disadvantage, are, however, open to multiple interpretations and are usually not unambiguous. Thus, if process scores are used their interpretation must be laid down in advance, not post hoc, and they must be quantifiable.
8. **Confounded user characteristics**. There might be a lot of relevant user characteristics which can hardly be controlled completely (e. g., pre usage knowledge about domain, motivation, expertise in working with computer, general intelligence etc.). Based on psychological theories those characteristics that might be confounded have to be controlled and examined during experimentation. This is, however, a basic issue for all evaluation studies and psychological experiments.
9. **No optimal solution**. It is a characteristic of complex problems that no optimal solution can be computed. In the same way, usually no optimal solution in business simulators does exist, which could be used to assess the performance of subjects using the simulator (no absolute benchmark). However, performance can be compared to other users, for example, to experts in the domain (relative benchmark). It has to be seen as a sign of ecological validity that the optimal solution is not known.

10. **Duration of game run**. Playing a scenario often takes a considerably long time (in some cases a few hours) but yields only one independent measure of the game score. Thus, reliability of data is often limited. Furthermore, users can get tired or bored. The single measurement provided by one game run, on the other hand, is usually accompanied by many observations (e. g., mouse clicks of users, windows observed by users) which leads to the problem of data reduction (Brehmer 1992). This is a basic issue because complex situations just need time to be understood and managed by users.
11. **Integration with other teaching**. It is often stated that business simulators should be used in connection with other training measures (for instance, after teaching basic, declarative knowledge about the domain). The final aim is to embed the simulation into a complete suite of teaching methods ("interactive learning environment"). How simulators can practically be combined with other instructional media and what characteristics these other media should have is, however, only rarely discussed. Which effects are due to the simulation, and which are due to the other measures, remains unclear. See also Kerres (1998) for the importance of research about the embedding of teaching media into a didactic context. Instructional design might have more answers in the future.

Besides these methodological issues, it must be considered that the actual **costs** and the opportunity costs of developing and experimenting with business simulators are high. Although there is no cost data for developing business simulators available in public, costs for the development of "conventional" computer-based training (CBT) programs are estimated to be about 10,000–30,000 Euro per hour training (Kerres 1998). In addition to that, working with business simulators often takes a considerably long time, in order to understand the complex content they convey. Thus, opportunity costs are high as well.

Thus, even when just aiming to compare single features of business simulators, still a number of issues remain. Some topics have to do with shortcomings of current business simulators and can therefore be mitigated through a careful design of future simulation tools (Größler 2000). Some items, however, stand as principal methodological issues of simulators. These problems have to be taken into account when conducting research with business simulators.

Conclusion

In this paper it is argued that the absolute and lasting success of teaching with business simulators cannot be proven scientifically. In the same way, comparison studies between business simulators and other teaching media lead to no definite answer about the superiority of simulators over other learning tools. However, when reviewing the literature about the evaluation of other teaching methods, this has become no surprise. The general effectiveness of no such method is proven; it can only be supposed according to experience (and most of the times for certain situational and personal circumstances).

If one follows this assumption an improvement of simulators and the finding of situations where simulators might be effective can be supported by comparative evaluation studies between simulators that are varied in just one feature. But also these studies require comprehensive knowledge in evaluation methodology and experimental procedures. Furthermore, there are still principal issues for all experiments with business simulators.

Advocates of business simulators should take this situation with both, modesty and self-confidence. They must be modest because the effectiveness of their favorite tool has not been proven, and will never be. They can be self-confident, however, because business simulators stand in one line with—when used within reason, very successful—media for teaching and learning: books, lectures, and audio/video. This paper is an attempt to structure the remaining task: to find sensible limits for the application of business simulators.

References

Andersen, D.F., I.J. Chung, G.P. Richardson and T.R. Stewart. 1990. Issues in Designing Interactive Games Based on System Dynamics Models. *Proceedings of the 1990 International System Dynamics Conference*, D.F. Andersen, G.P. Richardson and J.D. Sterman (eds.). Chestnut Hill; 31–45.

Argyris, C. and D.A. Schön. 1996. *Organizational Learning II. Theory, Method, and Practice*. Addison-Wesley: Reading, Massachusetts et al.

Ashby, W.R. 1956. *An Introduction to Cybernetics*. Champman & Hall: London.

Bakken, B.E. 1989. Learning in Dynamic Simulation Games; Using Performance as a Measure. *Computer-Based Management of Complex Systems: Collected Papers from the 1989 International System Dynamics Conference*, P.M. Milling and E.O.K. Zahn (eds.). Springer: Berlin et al.; 309–316.

Bakken, B.E., J. Gould and D. Kim. 1994. Experimentation in Learning Organizations: A Management Flight Simulator Approach. *Modeling for Learning Organizations*, ed. J.D.W. Morecroft and John D. Sterman. Productivity Press: Portland, Oregon; 243–266.

Beckmann, J.F. and J. Guthke. 1995. Complex Problem Solving, Intelligence, and Learning Ability. *Complex Problem Solving – The European Perspective*, P.A. Frensch and J. Funke (eds.). Lawrence Erlbaum: Hillsdale/Hove; 177–200.

Berry, D.C. and D.E. Broadbent. 1995. Implicit Learning in the Control of Complex Systems. *Complex Problem Solving – The European Perspective*, P.A. Frensch and J. Funke (eds.). Lawrence Erlbaum: Hillsdale/Hove; 131–150.

Bortz, J. and N. Döring. 1995. *Forschungsmethoden und Evaluation für Sozialwissenschaftler* [Research Methods and Evaluation for the Social Sciences], 2nd edition. Springer: Berlin et al.

Bosco, J. 1986. An Analysis of Evaluations of Interactive Video. *Educational Technology*, May 1986: 7–17.

Bredemeier, M.E. and C. Stein Greenblat. 1981. The Eductational Effectiveness of Simulation Games: a Synthesis of Findings. *Principles and Practices of Gaming-Simulation*, C. Stein Greenblat and R.D. Duke (eds.). Sage: Beverly Hills/London; 155–169.

Brehmer, B. 1992. Dynamic Decision Making: Human Control of Complex Systems. *Acta Psychologica* 81: 211–241.

Brehmer, B. and D. Dörner. 1993. Experiments with Computer-Simulated Microworlds: Escaping Both the Narrow Straits of the Laboratory and the Deep Blue Sea of the Field Study. *Computers in Human Behavior* 9: 171–184.

Broadbent, D.E. and B. Aston. 1978. Human Control of a Simulated Economic System. *Ergonomics* 21(12): 1035–1043.

Broadbent, D.E., P. FitzGerald and M.H.P. Broadbent. 1986. Implicit and Explicit Knowledge in the Control of Complex Systems. *British Journal of Psychology* 77: 33–50.

Clark, R.E. 1983. Reconsidering Research on Learning from Media. *Review of Educational Research* 53(4): 445–459.

Cook, T.D. and D.T. Campbell. 1979. *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. Houghton Mifflin: Boston et al.

Cronbach, L.J. 1990. *Essentials of Psychological Testing*, 5th edition. HaperCollings: New York.

Doyle, J.K. 1997. The Cognitive Psychology of Systems Thinking: An Agenda for Collaborative Research. *15th International System Dynamics Conference: Systems Approach to Learning and Education into the 21st Century*, Y. Barlas, V.G. Diker and S. Polat (eds.). Istanbul; 7–10.

Duffy, T.M. and D.H. Jonassen (eds.). 1992. *Constructivism and the Technology of Instruction – a Conversation*. Lawrence Erlbaum: Hillsdale/Hove/London.

Epstein, M.J. and R.A. Westbrook. 2001. Linking Actions to Profits in Strategic Decision Making. *MIT Sloan Management Review* 42(3): 39–49.

Folley, J.D. 1967. The Learning Process. *Training and Development Handbook*, R.L. Craig and L.R. Bittel (eds.). McGraw-Hill: New York et al.; 34–54.

Funke, J. 1991. Solving Complex Problems: Exploration and Control of Complex Systems. *Complex Problem Solving – Principles and Mechanisms*, R.J. Sternberg and P.A. Frensch (eds.). Lawrence Erlbaum: Hillsdale/Hove; 185–222.

Funke, J. 1993. Microworlds Based on Linear Equation Systems: A New Approach to Complex Problem Solving and Experimental Results. *The Cognitive Psychology of Knowledge*, G. Strube and K.F. Wender (eds.). Elsevier Science: Amsterdam; 313–330.

Funke, J. 1995a. Erforschung komplexen Problemlösens durch computerunterstützte Planspiele: Kritische Anmerkungen zur Forschungsmethodologie [Research on Complex Problem Solving with Computer-Based Planning Games: Critical Comments on the Research Methodology]. *Planspiele im Personal- und Organisationsmanagement*, T. Geilhardt and T. Mühlbradt (eds.). Verlag für Angewandte Psychologie: Göttingen; 205–216.

Funke, J. 1995b. Experimental Research on Complex Problem Solving. *Complex Problem Solving – The European Perspective*, P.A. Frensch and J. Funke (eds.). Lawrence Erlbaum: Hillsdale/Hove: 243–268.

Funke, U. 1991. Die Validität einer computergestützten Systemsimulation zur Diagnose von Problemlösekompetenz [On the Validity of a Computer-Based Simulation as a Diagnosis of Problem Solving Competence]. *Eignungsdiagnostik in Forschung und Praxis*, H. Schuler and U. Funke (eds.). Verlag für Angewandte Psychologie: Stuttgart; 114–122.

Garvin, D.A. 1993. Building a Learning Organization. *Harvard Business Review*, July-August 1993: 78–91.

Goodyear, P., M. Njoo, H. Hijne and J.J.A. van Berkum. 1991. Learning Processes, Learner Attributes and Simulations. *Education & Computing* 6: 263–304.

Graham, A.K. and P.M. Senge. 1990. Computer-based Case Studies and Learning Laboratory Projects. *System Dynamics Review* 6(1): 100–105.

Greenblat, C. Stein. 1981. Teaching with Simulation Games: a Review of Claims and Evidence. *Principles and Practices of Gaming-Simulation*, C. Stein Greenblat and R.D. Duke (eds.). Sage: Beverly Hills/London; 139–153.

Größler, A. 1999. The Influence of Decision Time on Performance in Use of a Business Simulator. *Systems Thinking for the Next Millenium – Proceedings of the 1999 Conference of the International System Dynamics Society*, R.Y. Cavana et al. (eds.). Wellington, New Zealand [CD-ROM Proceedings, Abstract p. 75].

Größler, A. 2000. Methodological Issues of Using Business Simulators in Teaching and Research. *Sustainability in the Third Millennium – Proceedings of the 18th International Conference of the System Dynamics Society*, P.I. Davidsen, D.N. Ford and A.N. Mashayekhi (eds.). Bergen, Norway [CD-ROM Proceedings, Abstract p. 85].

Haider, H. 1993. Was ist implizit am impliziten Wissen und Lernen? [What is Implicit with Implicit Knowledge and Learning?] *Sprache und Kognition* 12(1): 44–52.

Hayes, W.L. 1994. *Statistics*, 5th edition. Harcourt Brace: Fort Worth et al.

Hays, R.T. and M.J. Singer. 1989. *Simulation Fidelity in Training System Design*. Springer: New York.

Huber, G.P. 1990. Organizational Learning: The Contributing Processes and the Literatures. *Organization Science* 2(1): 88–115.

Jonassen, D.H. 1992. Evaluating Constructivistic Learning. *Constructivism and the Technology of Instruction – a Conversation*, T.M. Duffy and D.H. Jonassen (eds.). Lawrence Erlbaum: Hillsdale/Hove/London; 137–148.

Kaufman, F.L. 1976. An Empirical Study of the Usefulness of a Computer-based Business Game. *Journal of Educational Data Processing* 13: 13–22.

Kerres, M. 1998. *Multimediale und telemediale Lernumgebungen* [Multimedia and Telemedia Learning Environments]. R. Oldenbourg: München/Wien.

Keys, B. and J. Wolfe. 1996. The Role of Management Games and Simulations in Education and Research. *Yearly Review, Journal of Management* 16(2): 307–336.

Kim, D. H. 1993. The Link between Individual and Organizational Learning. *Sloan Management Review*, Fall 1993: 37–50.

Kirkpatrick, D.L. 1967. Evaluation of Training. *Training and Development Handbook*, R.L. Craig and L.R. Bittel (eds.). McGraw-Hill: New York et al.; 87–112.

Klein, R.D. and R.A. Fleck, Jr. 1990. International Business Simulation/Gaming: An Assessment and Review. *Simulation & Gaming* 21(2): 147–165.

Köbberling, A. 1971. *Effektiveres Lehren durch Programmierten Unterricht?* Julius Beltz: Weinheim/Basel/Berlin.

König, U. 1999. Ansätze zur Simulation des Wissenserwerbs mit System-Dynamics-Modellen [Approaches to the Simulation of Knowledge Acqusition Using System Dynamics Models]. *Intelligente Organisationen – Konzepte für turbulente Zeiten auf der Grundlage von Systemtheorie und Kybernetik,* M. Schwaninger (ed.). Duncker & Humblot: Berlin; 373–380.

Lane, D.C. 1995. On a Resurgence of Management Simulations and Games. *Journal of the Operational Research Society* 46: 604–625.

Lefrancois, G.R. 1982. *Psychological Theories and Human Learning*, 2nd edition. Brooks/Cole: Monterey.

Maier, F.H. and A. Größler. 2000. What Are We Talking About?—A Taxonomy of Computer Simulations to Support Learning. *System Dynamics Review* 16(2): 135–148.

Marchionini, G. 1989. Making the Transition from Print to Electronic Encyclopedias: Adaptation of Mental Models. *International Journal of Man-Machine Studies* 30(6): 591–618.

Milling, P. 1995. Organisationales Lernen und seine Unterstützung durch Managementsimulatoren [Organizational Learning and its Support through Management Simulations]. *Zeitschrift für Betriebswirtschaft, Ergänzungsheft 3/95: Lernende Unternehmen*: 93–112.

Milling, P. 1996. Simulationen in der Produktion [Simulation in Production]. *Handbuch der Produktionswirtschaft*, 2nd edition, W. Kern, H.-H. Schröder and J. Weber (eds.). Schäffer-Poeschel: Stuttgart; 1840–1852.

Norris, D.R. and C.A. Snyder. 1982. External Validation of Simulation Games. *Simulation & Games* 13(1): 73–85.

Packer, D.W. and W. Glass-Husain. 1997. Designing Interactive Multi-User Learning Laboratories. *15th International System Dynamics Conference – Systems Approach to Learning and Education into the 21st Century*, Y. Barlas, V.G. Diker and S. Polat (eds.). Boğaziçi University: Istanbul, Turkey; 79–86.

Pidd, M. 1993. *Computer Simulation in Management Science*, 3rd edition. John Wiley & Sons: Chichester et al.

Prensky, Marc. 2001. *Digital Game-Based Learning*. McGraw-Hill: New York et al.

Reigeluth, C.M. and E. Schwartz. 1989. An Instructional Theory for the Design of Computer-Based Simulations. *Journal of Computer-Based Instruction* 16(1): 1–10.

Richardson, G.P. 1996. Problems for the Future of System Dynamics. *System Dynamics Review* 12(2): 141–157.

Roethlisberger, F.J. and W.J. Dickson. 1939. *Management and the Worker*. Cambridge: Harvard University Press.

Schulmeister, R. 1996. *Grundlagen hypermedialer Lernsysteme: Theorie, Didaktik, Design* [Foundations of Hypermedia Training Systems: Theory, Didactics, Design]. Addison-Wesley: Bonn et al.

Sommerville, I. 2001. *Software Engineering*, 6th edition. Addison-Wesley: Harlow/Munich.

Spender, J.-C. 1998. The Dynamics of Individual and Organizational Knowledge. *Managerial and Organizational Cognition*, C. Eden and J.-C. Spender (eds.). Sage: London/Thousand Oaks/New Delhi; 13–39.

Stolurow, L.M. 1962. Implications of Current Research and Future Trends. *Journal of Educational Research* 55(9): 519–527.

Süß, H.-M. 1996. *Intelligenz, Wissen und Problemlösen* [Intelligence, Knowledge and Problem Solving]. Hogrefe: Göttingen et al.

Süß, H.-M. 1999. Intelligenz und komplexes Problemlösen [Intelligence and Complex Problem Solving]. *Psychologische Rundschau* 50(4): 220–228.

Thurman, R.A. 1993. Instructional Simulation from a Cognitive Psychology Viewpoint. *Educational Technology, Research and Development* 41(4): 75–89.

Weidenmann, B. 1993. Psychologie des Lernens mit Medien [The Psychology of Learning with Media], *Pädagogische Psychologie*, 3rd edition, B. Weidenmann, A. Krapp, M. Hofer, G.L. Huber and H. Mandl (eds.), Psychologie Verlags Union: Weinheim; 493–554.

Wolfe, J. 1976. Correlates and Measures of the External Validity of Computer-Based Business Policy Decision-Making Environments. *Simulation & Games* 7(4): 411–438.

Wolfe, J. 1985. The Teaching Effectiveness of Games in Collegiate Business Courses. *Simulation & Games* 16(3): 251–288.

Notes

1. Most of the statements in this paper can be applied to simulators from other domains as well.
2. Components of gaming functionality are discussed, for instance, in Prensky (2001) who identifies six key elements: rules, goals and objectives, outcomes and feedback, conflict/competition/challenge/opposition, interaction, representation or story (119).
3. That users should not always know about the point to be made using a simulator belongs to the idea behind these tools and, therefore, can be intended (Bredemeier and Greenblat 1981). If the designers and facilitators, however, do not know what the learning objectives are it can hardly be expected at all that users learn.
4. Bredemeier and Greenblat (1981, 163) name this concern: “students’ liking the experience may not necessarily mean they learned anything from it.”
5. The same three items plus the perception of users/trainees as a motivational aspect can be found in Kirkpatrick (1967) as steps for evaluation.

6. Figure 2 does only depict the raw learning process. It does not include numerous other important processes and feedbacks, like unlearning (forgetting) and the influence of cognitive structure on the learning rate. See König (1999) for a more complete level-rate-diagram.
7. Consistency and duration are preconditions of reliably measuring learning (Folley 1967).
8. Figure 4 is simplified (as is Figure 3) insofar, as additional influencing variables are not depicted. For example, a possible influence of intelligence or already available knowledge on the process of knowledge acquisition or controlling ability (and therefore on performance) are missing (Süß 1996; Beckmann and Guthke 1995). See also Figure 5.
9. Klein and Fleck (1990) argue that classical pre and post tests measure what they call the "recall effect", not a learning effect. However, this seems to be dependent on the validity of the tests used.
10. According to Clark (1983, 453) educational media are just "vehicles for instruction". What matters is content. Thus, any teaching measure can lead to any learning objective in principle. However, "certain elements" seem to facilitate learning. For example, gaming simulations seem to be well suited to teach dynamically changing matters (Reigeluth and Schwartz 1989).
11. Köbberling (1971) discusses the possibility to achieve reliable statements about the relative effectiveness of training methods in specific situations. This ought to be done through a substantial number of replicated "evaluative experiments."
12. Hays and Singer (1989) briefly describe different concepts for cost analysis of training measures.
13. Neglecting at this point different cost of developing the discussed measures of education.
14. I intentionally skip the point here that when conducting any kind of evaluation research not only characteristics of the learning tool business simulator have to be taken into account. What also needs to be considered are characteristics of their users and situational determinants (Funke, J. 1995b).
15. Other experimental designs that can be used in evaluation research with business simulators are: "the removed-treatment design with pretest and posttest", "the repeated-treatment design", and "the reversed-treatment nonequivalent control group design with pretest and posttest" (Cook and Campbell 1979). These are not described any further here.