

# The evaluation of a discipline: a framework for evaluating simulation games

Vincent Peters  
Hayke Everwijn  
Marleen van de Westelaken

**Abstract** Evaluation is considered as an important issue in relation to gaming and simulation. Many studies deal with this subject, but yet it is hard to get an overall picture of the strengths and the quality of gaming simulations at the aggregated level, also referred to as the discipline. In this chapter we look at the issue of evaluation of simulation games from a methodological perspective: the 'why', the 'what' and the 'how' of evaluation of simulation games. Concerning the 'why' nine concepts are distinguished in which evaluation can be connected to gaming simulations. For the 'what' we elaborate four levels on which the evaluation may focus (appreciation, cognition, behavior and performance). For the 'how' we point at the drawbacks of the experimental design for evaluation studies and we point at the strengths of theory based and qualitative evaluation in case studies. Throughout this chapter the central focus point is the question how the results of the evaluation of single simulation games can contribute to insight in the quality of gaming simulation as an instrument that may be used in varying situations.

**Key words** effectiveness; efficiency; evaluation; experiments; gaming as a discipline; simulations games; qualitative evaluation; theory based evaluation;

## 1 Introduction

One of the famous statements of Dick Duke about the effectiveness of simulation games is "It works, that is all we have" (Duke & Geurts, 2004, p. 211). This has been his answer for many years (in the eighties and nineties) to the question why simulation games do what they are supposed to do. His answer reflected the situation that there was a lack of empirical evidence for the effectiveness and efficiency of simulation games. Since then many evaluations of the use of simulation games have been reported, and also models for evaluating simulation games have been developed. Hense (2004) and Hense and Kriz (2008) have come to the conclusion that one way to improve the evaluation practice in the field of gaming and simulation is to let the evaluation of simulation games be explicitly driven by theoretical notions and expectations. They have developed an evaluation model and they refer to their approach as 'a theory-based evaluation approach'.

This approach is an important impetus for the practice of evaluating simulation games. However, we think that we may add something to the evaluation of simulation games by looking at the evaluation practice from a methodological perspective.

In this chapter we shortly discuss the concept of evaluation and its purposes, how the field (discipline) of gaming simulation can profit from evaluation studies, and next we elaborate on three themes that may help to improve the practice of the evaluation of simulation games:

- why what kind of questions do we want to be answered by means of evaluation?
- what what kind of variables could / should we involve in the evaluation of simulation games?
- how how can we design an evaluation study on simulation games, which research designs are fruitful?

## 2 The concept of evaluation

Evaluation is a concept that has many connotations and is used in different ways. A well-known definition of evaluation (in an educational context) is given by Cronbach (1983, p. 101-102): "...we may

define evaluation broadly as the collection and use of information to make decisions about an educational program. [...] Many types of decisions are to be made, and many varieties of information are useful. It becomes immediately apparent that evaluation is a diversified activity and that no one set of principles will suffice for all situations." Rossi and Freeman (1989) give a definition of evaluation, that can be considered as an elaboration of the terms 'collection and use of information' in Cronbach's definition: "Evaluation research is the systematic application of social research procedures for assessing the conceptualization, design, implementation, and utility of social intervention programs" (1989, p. 18).

These definitions refer to 'educational programs' and 'social intervention programs', but we are convinced that these terms may be replaced by the term 'simulation games' without doing any harm to the tenor of the definitions.

An important distinction that is made in relation to evaluation is the distinction between informal and formal evaluation. Douglass (1998, p. 2) contrasts these two types of evaluation in the following way:

<i>Informal</i>	<i>Formal</i>
Casual	Deliberate
Spontaneous	Designed and planned
Criteria are implicit	Criteria are explicit and clear
Evidence is quickly and insufficiently ascertained	Evidence is systematically collected
Judgment is hastily made	Judgment is carefully made

Earlier, Stufflebeam and Shinkfield (1998) have formulated three criteria that should be met in order to consider an evaluation as a formal evaluation:

- the criteria that are used for the evaluation have to be clear and unambiguous
- the methods used to come to a judgment are adequately selected and applied
- the evaluation process is reproducible and the criteria are public.

The combination of these definitions tells us something about the 'why' and the 'what' of evaluation: the objective of evaluating simulation games is to inform us about the quality of simulation games in order to enable us to make decisions. Investigating and assessing the conceptualization, design, implementation and utility of the simulation game accomplish this evaluation.

These definitions and distinctions also show us, that evaluation is more than just assessing the learning outcomes of using simulation games. In the next section we elaborate in what ways evaluation may serve our understanding of simulation games.

### **3 The purpose of evaluating simulation games**

When looking at the purpose of the evaluation of simulation games we have to distinguish two different purposes this evaluation may serve.

In the first place, evaluation may be interesting and important from the perspective of a specific simulation game in order to assess the quality of that specific instrument. Evaluation gives information to the game designer, the facilitator, the user (client) and the players whether the participants, by playing the simulation game, meet the objectives it was designed for. If a simulation game has proven repeatedly to be successful, one may come to the conclusion that this particular simulation game is an effective instrument (for the purposes it was designed for). In addition to the question of effectiveness, one may also be interested in the question of efficiency: is it worthwhile to use this instrument or are there other tools or instruments that may realize similar results against lower efforts or costs. This kind of evaluation questions help the designer and facilitator to understand whether they have an effective and efficient instrument, and it may convince potential users (clients) that playing the game will contribute to e.g. new behavior, skills or attitudes.

In the second place, evaluating simulation games may serve another purpose, especially if we look at the issue of evaluation from the perspective of 'simulation and gaming as a discipline' (cf. Crookall, 2010) or the gaming society. The gaming society is interested in assessing the quality of simulation games on a more general level: are simulation games in general effective and powerful instruments? For what purposes can they be deployed? And under what conditions can they be used successfully? The gaming society cannot do with the adage: "it works, that is all we have". There is need for more scientific evidence on simulation games as a tool for teaching, training, change, and research.

The responsibility for these efforts was put in the hands of academics: "It was a plea towards academics in the gaming discipline to do more research, share their results and jointly build up gaming/simulation as a true academic discipline" (Duke & Geurts, 2004, p. 211). In stimulating and coordinating these efforts there may be an important role for organizations like Isaga, Absel, Sagsaga, Saganet, Jasag and the other 'local saga's' to initiate and coordinate studies, in which the 'academics in the gaming discipline' will jointly try to answer these important research questions. However, since these organizations are rather informal and do not have the means (in terms of organization, research programs, personnel and money) to set up this kind of ambitious research projects, we as gaming society will have to rely on other sources of evidence, such as the accumulation of the findings of evaluation studies of single simulation games.

In recent years there have been many researchers who have conducted meta-analyses in order to compare and accumulate the findings of evaluation studies on single simulation games (see e.g., Greenlaw & Wyman, 1973; Pierfy, 1977; Wolfe, 1985; Dorn, 1989; Keys & Wolfe, 1990; Randel, Morris, Wetzal & Whitehal, 1992; Wolfe, 1997; Faria, 2001; Vogel, Vogel, Cannon-Bowers, Bowers, Muse and Wright, 2006; Sitzman, 2011). By accumulating the findings their overview articles try to reveal the strengths and weaknesses of simulation games as a category of instruments.

However, what hinders drawing conclusions from this kind of accumulating studies is the fact that the single evaluation studies are very different on many aspects like the theoretical background, the kind of variables involved, and the research methods used.

In order to provide a framework for comparing and accumulating evaluations of simulation games, we investigate in the next part of this article the 'why', 'what' and 'how' of the evaluation of simulation games.

## **4 The 'why' of evaluation: concepts of evaluation**

### **4.1 Purposes of evaluation**

Based on a comparison of several publications on evaluation and of articles in which simulation games are evaluated, we come to a distinction between nine different ways the concept of evaluation is used in the context of simulation and gaming. In Table 1 these concepts of evaluation are mentioned; after that they are described shortly. (cf. Mayer, Stegers-Jager & Bekebrede, 2007).

- I. *Evaluation as assessing the performance of participants in a simulation game.*
- II. *Evaluation by using simulation games.*
- III. *Evaluation as assessing the quality of the game design and the design process.*
- IV. *Evaluation as assessing the quality of the simulation game.*
  - a. *the quality of the game-in-the-box.*
  - b. *the quality of the game-in-use.*
- V. *Evaluation as assessing the effectiveness of the simulation game.*
  - a. *the degree of effectiveness of the simulation game.*
  - b. *how and when are simulation games effective.*
- VI. *Evaluation as assessing the efficiency of simulation games.*
- VII. *Evaluation as assessing the effects of a gaming based intervention.*

**Table 1 Concepts of evaluation in relation to simulation and gaming**

*I. Evaluation as assessing the performance of participants in a simulation game.*

The term evaluation refers in this context to evaluating the behavior / performance of the participants in the simulation game, not the game itself. The use of the word evaluation in this sense relates to identifying adequate performance indicators and to computing scores on each of these indicators within the simulation game (see e.g. Anderson & Lawton, 1992; Rising, 2004; Olson, Scheller, Larson, Lindeke and Edwardson, 2010).

*II. Evaluation by using simulation games.*

Simulation games are used as experimental conditions in which e.g. new policies are tested before they are actually implemented (ex ante evaluation). Alternatively they may be used as a setting in which hypotheses can be tested. Examples of this application are Keys and Wolfe (1990), Quanjel, Willems and Talen (1998), Mastik, Scalzo, Termeer and In 't Veld (1995), Kriz and Brandstätter (2002) and Kuit, Mayer and De Jong (2005). Bowen (1978) and Vissers, Heijne and Peters (1995) (among others) have published about this application of evaluation as a research method. If simulation games are used in this way, they have to match the rather strict criteria for validity and reliability that are formulated for all research methods and instruments.

*III. Evaluation as assessing the quality of the game design and the design process.*

This concept of evaluation is explicitly included in the design steps as formulated by Duke (1981), where it has to be assessed whether the game design is in accordance with the specifications. The evaluation may also concern the entire design process, i.e. the path and procedures to build the simulation game. An example of a study focusing on the evaluation of design principles can be found in Mayer, Bockstael-Blok and Valentin (2004).

*IV. Evaluation as assessing the quality of the simulation game.*

Vissers, Peters, Heijne and Geurts (1998) make a distinction between 'the-game-in-the-box' (the simulation game as it is designed) and 'the-game-in-use' (the simulation game as it is used).

*IVa. Evaluation as assessing the quality: the game-in-the-box.*

Several authors have defined criteria a simulation game should match in order to be considered a good simulation game. Dukes and Waller (1976), Thiagarajan and Stolovich (1979), Geurts and Joldersma (2001) and Hindle (2002) are examples of authors who come up with characteristics that simulation games should have. Feinstein and Cannon (2002) have investigated several publications and come to three criteria:

- fidelity                      the level of realism of the simulation game;

- verifiability is the model built in a proper way? Does the model work as it is supposed to work?
- validation did we build the correct model; are the conclusions drawn from the simulation similar to conclusions based on the 'real life situation'?

Peters, Vissers and Heijne (1998) and Vissers et al. (1998) have published on the validity of simulation games. In the context of computer-based games, the issue of fidelity has got a new dimension and new studies have been conducted (see e.g. Visschedijk, 2010; Lukosch, 2014).

#### *IVb. Evaluation as assessing the quality: the game-in-use.*

Even though we did not come across any references to this aspect in our set of publications, we think that this form of evaluation should not be forgotten in this overview. This concept of evaluation focuses on the conditions in which the simulation game is used, and addresses questions like: 'Is the simulation game played according to the directions for use?', 'Is it used for the proper group of participants', 'Is the debriefing done in an appropriate way, addressing the right themes?' and so on.

#### *V. Evaluation as assessing the effectiveness of the simulation game.*

Following the distinction Boocock and Schild (1968) make concerning the evaluation of social technologies in the 'engineering' and the 'science' approach, we see two different connotations for evaluation:

##### *Va. Evaluation as assessing the degree of effectiveness of the simulation game.*

From the perspective of the 'engineering approach' evaluation research tries to assess how successful simulation games are in bringing about desired changes in the behavior, skills, attitudes, knowledge, or emotions of participants.

Evaluating the effectiveness of simulation games implies that we have to ask and answer at least three questions:

- Is there a change in the target behavior (i.e. behavior, skills, attitudes, emotions, et cetera)?
- Is this a change in the desired direction?
- Is the simulation game the cause for this change?

The answers to these three questions give us insight in the effectiveness of the simulation game and provide us with information about the value of the instrument.

##### *Vb. Evaluation as assessing how and when simulation games are effective.*

In terms of a 'science approach' evaluation research focuses on the conditions that facilitate or hinder the simulation game to be successful in realizing effects. Getting insight in these factors and conditions focuses on issues like:

- the conditions for using the simulation game;
- characteristics of the participants;
- approaches for debriefing the results and experiences.

The study of Hense, Kriz and Wolfe (2009) is an example of a study where a large number of characteristics of the participants and the game environment are included in order to investigate what their role is in the way the simulation 'SIMGAME' operates.

#### *VI. Evaluation as assessing the efficiency of a simulation game.*

The next concept of evaluation in relation to simulation game has to do with assessing whether the simulation game is the most efficient instrument to bring about the desired changes. The simulation game may be effective, but perhaps the costs (in terms of efforts, money, time) to bring about these changes are higher than for other instruments that may realize the same or similar results. St. Germain and Laveault (1997) and Hindle (2002) stand up for this type of research. Especially when designing computer-based games the design costs may become very high.

Research questions focusing on the efficiency of simulation games are:

- Is our simulation game the best way to bring about these changes?
- Are there any side effects (positive or negative, wanted or unwanted)?

#### *VII. Evaluation as assessing the effects of a gaming based intervention.*

The last concept of evaluation is the one that is applied and reported the most: do participants have or show the desired skills, knowledge, attitudes, and/or behavior after playing the simulation game. In this case the evaluation mainly focuses at the question 'Do we see a change?' without drawing conclusions about the instrument itself (in contrast to the above mentioned concept *Va*, where one wants to draw conclusions about the simulation game). There are many publications describing the results of this kind of evaluation.

#### **4.2 Consequences for the practice of evaluating simulation games**

As the above enumeration demonstrates, evaluation studies of simulation games may concern very different subjects and serve various purposes. In the context of this article the first two concepts seem not very relevant in relation to our search for 'evaluation of the discipline' since they do not help us to get more insight in the quality of simulation games themselves.

The other concepts of evaluation can be seen as contributions to the 'design-in-the-small' as a part of the 'design science', as distinguished by Klabbers (2009).

The third concept focuses on the role of evaluation in the design process. On the one hand one may say that this kind of evaluation is very specific for a single simulation game (does the design match the specifications and objectives?). On the other hand, explicit descriptions of these evaluations and their findings may help the gaming society to improve the game design process and thus help to improve the image of the 'discipline'. Kaplan (1964) has made the distinction between the 'logic in use' and the 'reconstructed logic'. 'Logic in use' refers to the way a researcher (or game designer) operates in a concrete case, while 'reconstructed logic' refers to more generic methods or procedures that are based on a reflection on the 'logic in use' descriptions. Reconstructed logic often forms the basis for a methodology. The steps for game design, as described by Duke (1981) and Duke and Geurts (2004) can be considered as 'reconstructed logic'. If more game designers describe (in detail) the steps and procedures they have followed in the game design process (and the advantages, restrictions, pitfalls, required skills, and so on), it would be possible to integrate these experiences into a new or improved reconstructed logic. Therefore, from the perspective of the discipline, descriptions of the 'logic in use' are very important.

The evaluation of the 'game-in-the-box' (concept *IVa*) is interesting since it gives information on the quality of the simulation game as it was designed. However, additional evaluative information on the game-in-use (concept *IVb*) may inform users on the applicability of the simulation game in specific circumstances, different than the conditions where the simulation game was originally designed for; e.g. in relation to other themes, other groups of participants, different time frame, et cetera.

The next three concepts of evaluation (*Va*, *Vb* en *VI*) may help us to get insight in the quality of simulation games in terms of the effects, the conditions that enhance or hinder the effectiveness, and in terms of the efficiency. This insight is important, as well from the perspective of a specific simulation game, as from the perspective of the instrument of simulation game in general, i.e. the discipline.

Evaluation according to concept *VII* is only interesting for the gaming society if the results of this research can be accumulated in meta-studies. In order to accumulate the information provided by evaluation studies it is a prerequisite that the evaluators explicitly indicate which off the above-mentioned concepts of evaluation form(s) the basis for their study.

### **5 The 'what' of evaluation: variables in the evaluation**

When we want to judge whether a simulation game is successful or effective, there is a great diversity of variables that can be investigated. The selection of these target variables is dictated by the objectives of the simulation game. Dependent on the specific objective of the simulation game, the target variables will also vary in the type of effects that are measured: behavior, skills, knowledge, attitudes, emotions, interactions between people, et cetera. From the perspective of accumulation of the results of single evaluations this variety of target variables may seem disastrous: if each evalu-

ation study has its own set of target variables, how can we then compare these studies and how can we draw conclusion about simulation games at the level of the discipline?

A solution is that we do not look at the specific variables when comparing or accumulating separate evaluation studies, but we look at a more abstract level e.g. to the directions or the strengths of the effects that were measured, regardless of the specific variables.

There is another dimension on which the target variables may differ that may be useful to streamline the evaluations. Kirkpatrick (1967, 1998) distinguishes four levels on which an evaluation can focus. These four levels are:

1. reaction	indicating "...how well the trainees liked a particular training program" (1967, p. 88);
2. learning	referring to "...the principles, facts and techniques which were understood and absorbed by the conferees" (1967, p. 96);
3. behavior	which "can be defined as the extent to which change in behavior has occurred because the participants attended the training program" (1998, p. 20);
4. results	i.e. "the final results that occurred because the participants attended the program" (1998, p. 23).

Phillips (1997) has added a fifth level to the four levels of Kirkpatrick:

5. ROI	return on investment, to quantify the monetary value of training investments, "what do I get back for my investment".
--------	---

Since this fifth level focuses exclusively at monetary aspects, we will discard it in the context of this article.

The levels are hierarchically ordered and effects on a previous level can be seen as preconditions for the effects on the next level. If a simulation game is deployed to contribute to the solution of a problem, we are in the end interested in the question whether the problem does not exist anymore and whether this change can be attributed to the simulation game (level 4). Before we can answer this question we have to assess whether the 'message' that was in the simulation game is understood and absorbed by the participants (level 2) and whether they were able to apply the newly acquired knowledge in practice, i.e. whether they behaved accordingly to the 'lessons' of the simulation game (level 3). The first level focuses on the extent to which the participants consider the used tool (the simulation game) as a nice and supporting tool. We believe that for by far most of the situations where simulation games are deployed, the intended effects are aiming at the fourth level.

We will elaborate the four levels of Kirkpatrick in the following sections. For each of the levels, which we will give a slightly different name than Kirkpatrick, we will describe the focus of the evaluation and we address some methodological considerations.

### 5.1 Level 1: Appreciation

The first level is the level of 'appreciation'. When people have participated in a simulation game, the facilitator is interested in the extent that participants liked the simulation game. This concept of 'appreciation' (Kirkpatrick refers to the term 'customer satisfaction') involves more than only the 'liking of' the game. The game designer or the facilitator may also want to have information on other aspects of the simulation game or of the conditions it was used in. The following questions are examples of research questions that may be formulated and investigated at this level:

- Do the participants like the simulation game; do they appreciate it as an instrument for training or intervention? Did they have fun? Did they feel challenged by the simulation game?

- Do the participants understand / appreciate the metaphor of the simulation game? Does this metaphor facilitate or hinder them in getting insight in the real life situation? Do they see a relation between the situation and processes represented in the simulation game and the real life situation and processes this simulation game refers to?
- How do they value specific aspects of the simulation game, such as the scenario, the materials, the role descriptions, the time pressure, the complexity of the assignments, the symbols used, the level of abstraction?
- How do they value the debriefing? Does it help them to understand and interpret their own experiences in the simulation game, and to translate experiences in the simulation game to the real life situation? Does it help them to draw conclusions about their own performance?
- Do the participants feel secure in the simulation games? Do they see it as a safe environment?

Answering this kind of research questions serves two purposes: a) they inform the game designer and facilitator about the quality (as perceived by the participants), and b) it gives information that can be important to interpret the results of the evaluation on the next levels (cf. the aspect of the psychological validity of the game; Peters et al., 1998).

This type of evaluation is very valuable for the evaluation of single simulation games since it helps game designers and facilitators to improve the instrument. Since the kind of variables used in these evaluations varies and since the attention is narrowly focused on the specific simulation game and the specific context in which it was used, it is hard to draw general conclusions on the quality of simulation games in general from the kind of variables we referred to above.

In the last subsection we will come back to the evaluation from the perspective of the discipline.

## 5.2 Level 2: Cognition

In simulation games that are used to effectuate changes in 'behavior'<sup>1</sup> there is a cognitive component involved. This cognitive component, or the 'message' may include several types of information, such as the content of the desired 'behavior', arguments for adopting this 'behavior', procedures involved, criteria for the 'behavior', et cetera. The desired 'behavior' is described, explained and discussed, at the very least in the debriefing phase. At the level of cognitions evaluation focuses on the question whether the participants have understood the message that was embedded in the simulation game and whether they have made themselves familiar with this message.

Examples of specific research questions for this level are:

- Do the participants get the message? Do they learn what they are supposed to learn?
- Do they understand what they have learned? Do they know the importance of the new 'behavior' in relation to the objective of the simulation game?
- Are they able to distinguish the essentials from the details?
- Do they draw the right conclusions?
- Are there other elements they have picked up from playing the simulation game, wanted or unwanted?

The information yielded at this cognitive level helps the game designer to improve specific elements in the simulation game (e.g. to articulate specific cognitive elements in the scenario or other gaming elements), and it helps the facilitator to decide to put more or less emphasize on specific elements in the briefing or the debriefing. Looking at the desired outcomes of the simulation game, the information at this level is primarily important for understanding and interpreting the effects (or lack of effects) measured at the subsequent levels.

---

<sup>1</sup> We use the term 'behavior' here in a very broad sense; it includes knowledge, skills, attitudes, habits, emotions, interactions between people, et cetera. It may concern new 'behavior', but also unlearning old routines, knowledge, et cetera. To distinguish it from actual behavior we put the term between quotes.



### 5.3 Level 3: Behavior

In order to improve a situation or to solve a problem, it is not sufficient that the people involved know what is happening, why it should be changed or what should be done (the cognitive level). They will have to apply this new information in their activities and routines in their real life situations. Only if we can assess that participants apply the new 'behavior' (or do not use the routines they have to unlearn) we will be able to draw conclusions about the relation between changes in the real life situation and the simulation game. Therefore at the level of behavior the evaluation focuses at the following types of questions:

- Do the participants apply the 'behavior' they acquired during the simulation game in their real life situation the simulation game referred to?
- Do they apply this behavior in a proper way?
- Are they able to abandon old behavior and replace it by new behavior?

Measuring these effects has some pitfalls. A rather easy way to measure participants 'behavior' in the real life situation is sending them a questionnaire in which they are asked to report on their 'behavior'. Because of a possible discrepancy between reported or intended behavior and the actual behavior it is questionable if this method yields valid results. A more valid procedure for assessing effects at this level is to work with so-called unobtrusive measures, i.e. measuring without affecting the system or process under investigation. Possible ways to accomplish this are: assessing documents that have been produced, asking other people to describe the behavior of the person under investigation, and observations.

Another way to collect data on the behavior of participants in their real life situation is to have in-depth interviews with them. An interview gives the opportunity to ask rather detailed questions if the participants apply the newly learned 'behavior' in the real life situation, but it also gives opportunities to check for socially desirable answers e.g. by addressing an aspect once again from another angle.

However awkward evaluation research at this level is, it is interesting and necessary from the perspective of transfer of knowledge and skills from a simulation game to a real life situation (see e.g. Percival, Lodge and Saunders, 1993).

### 5.4 Level 4: Performance

The fourth level, finally, is the level where the evaluator tries to assess whether the aims of the simulation game are met. Simulation games are often deployed in situations where persons or organizations have to overcome problems, perform better, cooperate more, et cetera. So the final test for the effectiveness of using the simulation games is to assess whether the performance in the real life situation has improved after playing the simulation game. Evaluation at this level focuses on research questions like:

- Has the real life situation changed?
- Has it changed in the desired direction?
- Can these changes be attributed to the simulation game?

This type of evaluation is even more awkward than the evaluation on the third level. The methodological issues mentioned in the previous section are also valid for the evaluation at this level. But there may be complicating factors.

An extra complicating factor is that there must be enough time between the simulation game and the measurement of effects to allow the real life situation to change. If we teach people new procedures in a simulation game it will take some time before the effects of these new procedure are noticed in the real life situation. So we should wait long enough before measuring the effects. However, the longer we wait, the harder it will be to control for all kind of factors that may influence the performance, making it hard to assess the extent in which the effects may be attributed to the simulation game that was deployed.

### 5.5 The 'what' of evaluation from the perspective of the discipline

Specifically from the perspective of the discipline there are other important evaluation issues to be investigated, than mere the effects or outcomes. It is interesting to investigate whether the results that are measured at each of the four levels depends on the context where the simulation game is used or on characteristics of the participants. Wellington and Faria (1995) have distilled from a large number of studies factors that may affect performance in business simulations. They have identified the following factors (p. 205): personality characteristics of team members, previous academic achievement, ethnic origin of team members, team size, previous business experience, team organizational structure, degree of instructor explanation, and simulation grade weighting. This list may be expanded with other concepts and variables, like gender, age, learning style, education, occupation, experiences with simulation games, degree to which participants feel secure in the simulation game, do people participate as individual or as member of a group, et cetera. In the analysis of the results of a single simulation game it will be difficult to look for influences of these type of variables (because of the relatively small numbers of participants), but especially at the aggregated level these analyses will be possible by combining the data of single evaluation studies. This type of evaluation can be seen as evaluation of what we called type *Vb* in section 4.1, investigating the conditions that influence the effects of simulation games.

## 6 The 'how' of evaluation: research designs

In case of the evaluation of an instrument like simulation games the research focuses on the question whether playing the simulation game has contributed to changes in the desired direction on the level of cognition, behavior and performance (in terms of Kirkpatrick: levels 2, 3 and 4). The predominant research design for this kind of evaluation questions is the (quasi) experimental design: a pretest, the simulation game, and a posttest; the analysis focuses on determining the difference between the pretest and posttest. A control group may be added in order to improve the interpretation of differences between pretest and posttest.

How powerful the experimental research design may seem in analyzing differences between pretest and posttest and between separate groups (conditions), there is one big disadvantage of this research design: the experimental conditions are considered as black boxes. We assess the situation before, we assess the situation afterward and we look at the differences. However, this design does not give information about what happens within the simulation game that causes the changes. Other possible difficulties may threaten the quality of experimental studies for the evaluation of simulation games; we only mention them here briefly:

- the lack of control over the experiment, especially in a natural context
- the necessary elapse of time between the simulation game and the posttest
- changes in behavior and performance (level 3 and 4) have to be assessed in the real life situation, and therefore it is hard to design standard situations for the posttest, a prerequisite of the experimental design.

If we only have experimental studies to rely on for the evaluation of simulation games we will in the end still have the statement "It works, that is all we have"; however, we will have more proof for this statement. But if we want to know why simulation games work, we need other efforts that help to look inside the 'black box' and try to explain what happens in simulation games, in terms of input – throughput – output – outcome.

One example of an effort to look inside the 'black box' is the book 'Why do games work' (De Caluwé, Hofstede and Peters, 2008). Several authors, each from a different theoretical perspective, have investigated (theoretically) what is the 'active substance' in simulation games that facilitates people to learn or change.

Another way to get more insight into what makes simulation games work is to perform more in-depth research in the way simulation games are used and in the effects they have on participants.

The before mentioned approach for theory driven evaluation as proposed by Hense and Kriz is an example of such a study, in which the effects of simulation games are investigated in relation to a multiplicity of variables.

Another way of in-depth investigation of (the evaluation of) simulation games is by designing the research as a case study (Yin, 2003). Mallon and Webb (2006) give a good example of this kind of research, referring to it by using the term phenomenological approach. Studies such as Wilson et al. (2009) and Pavlas, Bedwell, Wooten, Heyne & Salas (2009), investigating the relation between attributes of simulation games and learning outcomes, may also be very helpful.

What these in-depth studies have in common is that they are rather laborious and time consuming. But in the end they will help answering the question why simulation games do what they are supposed to do.

## 7 Conclusions

In this article we have looked at the issue of evaluating simulation games from a methodological perspective. The importance of evaluation for the designer, facilitator, users or players of a specific simulation game differs from the importance and interest of evaluation from the perspective of the gaming society. As long as there are no large scale evaluation projects that try to assess the quality of simulation games at the level of the discipline, the small scale evaluation studies, focusing on the effects of a single simulation game, are the most important pieces of evidence for general conclusions about the strong and weak points, the effectiveness and the efficiency of simulation games, and about the conditions under which simulation games may be applied.

Given the diversity of objectives of evaluation studies we should not try to prescribe one methodology for the evaluation of simulation games. In stead we should put effort in setting up the evaluation studies and in describing the results in such a way, that they are better comparable and that it will be easier to accumulate or aggregate the results of the separate evaluation studies.

Based on what we investigated and described in this article we formulate five recommendations for studies on the quality of simulation games:

- If the gaming society wants to learn from the results of the evaluation of simulation games, it is a prerequisite that these evaluation studies are set up as formal evaluations; this implies that these studies are designed in a well-considered way and that the procedures, variables, criteria and conclusions are transparent.
- Evaluation of simulation games may serve different purposes; evidence about the effectiveness and efficiency of simulation games will help to present simulation games as a powerful tool. But also insight in the conditions that influence the effectiveness of simulation games is essential.
- In addition, systematic studies on the design process, building up the 'reconstructed logic', will help game designers in the design process of their simulation game and in the end it will contribute to the image of the tool simulation game.
- The most interesting information about the value of simulation games will be found in evaluations on the third and fourth level (behavior and performance); evaluation on the first two levels, the type of evaluation that is most frequently done, is especially useful for the evaluation of single simulation games, but for assessing the quality of simulation games from the perspective of the discipline, evaluations in the higher levels are also necessary, however laborious and awkward these studies may be.
- There is need for evaluation research that tries to look inside the black box, i.e. that tries to understand and explain which characteristics and elements of simulation games contribute to the effects that are assessed. This implies that in addition to experimental studies there is a need for studies that incorporate many explaining variables and/or in studies that are set up as small-scale case studies that try to reconstruct what happened during the simulation games and what caused the effects.

We have stressed the importance of the single evaluation studies for building up a total view of the quality of simulation games. But this should not imply that we put the responsibility for the evaluation of simulation games with the evaluators of single games. Their primary interest is to get information about that specific simulation game in order to improve it or to sell it, not to devise research questions and research strategies that will give more generic information. Therefore, there is a need for support and coordination from the level of the discipline. We think that should be considered as an important and challenging task for the "saga's", organizations that represent the community of gamers.

## References

- Anderson, P., & Lawton, L. (1992). A Survey of methods used for evaluating student performance on business simulations. *Simulation & Gaming*, 23 (4), pp. 490-498.
- Boocock, S., & Schild, E. (1968). *Simulation games in learning*, California: Sage Publications.
- Bowen, K. (1978). *Research games: an approach to the study of decision processes*. London: Taylor & Francis.
- Caluwé, L. de, Hofstede, G.J., & Peters, V. (2008). *Why do games work. In search of the active substance*. Deventer: Kluwer.
- Cronbach, L. J. (1983). Course improvement through evaluation. In: Madhaus, G. F., Scriven, M., & Stufflebeam, D. L. (eds.): *Evaluation models. Viewpoints on educational and human social services evaluation*. Boston/The Hague, Kluwer/Nijhoff, pp. 101-115.
- Crookall, D. (2010). Serious games, debriefing, and simulation / Gaming as a discipline. *Simulation & Gaming*, 41, pp. 898-920
- Dorn, D. (1989). Simulation games: one more tool on the pedagogical shelf. *Teaching Sociology*, 17, pp. 1-18.
- Douglah, M. (1998). Developing a concept of extension program evaluation. University of Wisconsin. (To be found at [cecommerce.uwex.edu/pdfs/g3658\\_7.pdf](http://cecommerce.uwex.edu/pdfs/g3658_7.pdf)).
- Duke, R. (1981). A Paradigm for game design, In: Greenblat, C., & Duke, R. (eds.). *Principles and practices of gaming-simulation*. London: Sage Publications: pp. 63-72.
- Duke, R., & Geurts, J. (2004). *Policy games for strategic management*. Amsterdam: Dutch University Press.
- Dukes, R., & Waller, S. (1976). Towards a general evaluation model for simulation games. *Simulation & Games*, 7, pp. 75-96.
- Faria, A. (2001). The changing nature of business simulation / gaming research: a brief history. *Simulation & Gaming*, 32, pp. 97-110.
- Feinstein, A., & Cannon, H. (2002). Fidelity, verifiability and validation of simulation: constructs for evaluation. *Simulation & Gaming*, 33, pp. 425-440.
- Geurts, J., & Joldersma, C. (2001). Methodology for participatory policy analysis. *European Journal of Operational Research*, 128, pp. 300-310.
- Greenlaw, P., & Wyman, F. (1973). The teaching effectiveness of games in collegiate business courses. *Simulation & Games*, 4, pp. 259-294.
- Hense, J. (2004). 'Theory-oriented evaluation of simulation games – The case of simgame', In: Kriz W., & Eberle, T. (eds.) *Bridging the gap: transforming knowledge into action through gaming and simulation*. Munich: Sagsaga.
- Hense, J., & Kriz, W. (2008). Making simulation games an even more powerful tool. Introducing the theory-based evaluation approach. In: Caluwé, L. de, Hofstede, G.J., & Peters, V. (2008). *Why do games work. In search of the active substance*. Deventer: Kluwer, pp. 211-218.
- Hense, J., Kriz, W., & Wolfe, J. (2009). Putting theory-oriented evaluation into practice: a logic model approach for evaluating SIMGAME. *Simulation & Gaming*, 40, pp. 110-133.
- Hindle, K. (2002). A grounded theory for teaching entrepreneurship using simulation games. *Simulation & Gaming*, 33, pp. 236-241.
- Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. San Francisco: Chandler Pub. Co.
- Keys, B., & Wolfe, J. (1990). The role of management games and simulations in education and research. *Journal of Management*, 16, pp. 307-336.
- Kirkpatrick, D. (1967). Evaluation of training. In: Craig, R., & Bittel, R. *Training and development handbook*. New York: McGraw Book Company, pp. 87-112.
- Kirkpatrick, D. (1998). *Evaluating training programs. The four levels*. San Francisco: Berrett-Koehler Publishers, 2<sup>nd</sup> edition.

- Klabbers, J. (2009). *The magic circle: principles of gaming & simulation*. Rotterdam: Sense Publishers. Third and revised edition.
- Kriz, W., & Brandstätter, E. (2002). Evaluation of a training program for systems/thinking and teamwork skills with gaming and simulation', In: F. Percival et al. (eds.) *The international simulation and gaming yearbook. Volume 11*, pp. 243-247.
- Kuit, M., Mayer, I., & Jong, M. de (2005). The INFRASTRATEGO game: an evaluation of strategic behavior and regulatory regimes in a liberalizing electricity market. *Simulation & Gaming*, 36, pp. 58-74.
- Lukosch, H., (2014). Designing simulation games to create shared situational awareness in safety critical work environments. *Human-Computer Interaction*, to be published.
- Mallon, B., & Webb, B. (2006). Applying a phenomenological approach to games analysis: A case study. *Simulation & Gaming*, 37 (2), pp. 209-225.
- Mastik, H., Scalzo, R., Termeer, C., & Veld, R. in 't (1995). *Simulatie van wetgeving, een verkenning van gebruiksmogelijkheden van spelsimulatie voor ex-ante evaluatie van wetgeving*. Den Haag: SDU. pp. 160. [Simulation of legislation, an exploration of uses of simulation games for ex-ante evaluation of legislation]
- Mayer, I., Bockstael-Blok, W., & Valentin, E. (2004). A building block approach to simulation: an evaluation using containers adrift. *Simulation & Gaming*, 35(1), pp. 29-52.
- Mayer, I., Stegers-Jager, K., & Bekebrede, G. (2007). *Spelend leren in virtuele werelden. Bouwstenen voor online gaming in het hoger onderwijs*. Groningen: Noordhoff. [Playful learning in virtual worlds. Building blocks for online gaming in higher education].
- Olson, D., Scheller, A., Larson, S., Lindeke, L., & Edwardson, S. (2010). Using gaming simulation to evaluate bioterrorism and emergency readiness education. *Public Health Rep.*, 125(3), pp. 468-477.
- Pavlas, D., Bedwell W., Wooten, S., Heyne, K., & Salas, E. (2009). Investigating the attributes in serious games that contribute to learning. *Proceedings of the Human Factors and Ergonomics Society. Annual Meeting*, 53 (27) pp. 1999-2003.
- Percival, F., Lodge, S., & Saunders, D. (1993). *The simulation and gaming yearbook 1993: developing transferable skills in education and training*. London: Kogan Page.
- Peters, V., Vissers, G., & Heijne, G. (1998). The validity of games. *Simulation & Gaming*, 29 (1), pp. 20-30.
- Phillips, J. (1997). *Handbook of training evaluation and measurement methods*, 3rd Ed. Texas: Gulf Publishing Company.
- Pierfy, D. (1977). Comparative simulation game research: stumbling blocks and stepping stones. *Simulation & Games*, 8 (2), pp. 225-268.
- Quanjel, M, Willems, R., & Talen, A. (1998). CRISISLAB: Evaluation and improvement of crisis management through gaming / simulation. In: Geurts, J. Joldersma, C., & Roelofs, E. (eds.). *Gaming / Simulation for policy development and organizational change*. Tilburg University Press: Tilburg.
- Randel, J., Morris, B., Wetzal, C., & Whitehal, B. (1992). The effectiveness of games for educational purposes: a review of recent research. *Simulation & Gaming*, 23(3), pp. 261-276.
- Rising, B. (2004). Evaluation in the context of collaborative / cooperative learning. In: Kriz, W., & Eberle, T. (eds.). *Bridging the gap: transforming knowledge into Action through Gaming and Simulation*. Munchen: Sagsaga.
- Rossi, P., & Freeman, H. (1989). *Evaluation: a systematic approach*. London: Sage Publications.
- Sitzmann, T., (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, 64 (2), pp. 489-528
- St. Germain, M, & Laveault, D. (1997). Factors of success of simulations and games: a systemic approach to the evaluation of an organizations impact on the user. *Simulation & Gaming*, 28 (3).
- Stufflebeam, D, & Shinkfield, A. (1998). *Systematic evaluation*. Dordrecht: Kluwer-Nijhoff.
- Thiagarajan, S. & Stolovitch, H. (1979). Frame games: an evaluation. *Simulation & Gaming*, 10 (3).
- Visschedijk, G. (2010). *The issue of fidelity: what is needed in 3D military serious games?* Master Thesis, University of Twente/TNO, Soesterberg.
- Vissers, G., Heijne, G., & Peters, V. (1995). Spelsimulatie in bestuurskundig onderzoek.[Simulation games in research on public administration]. *Bestuurskunde*, 4 (4), pp. 178-187.
- Vissers, G., Peters, V., Heijne, G., & Geurts, J. (1998). Validity of games/simulations: a constructive view. In: Geurts, J., Joldersma, C., & Roelofs, E. (eds). *Gaming/simulation for policy development and organizational change*. Tilburg: Tilburg University Press, pp. 353-360.
- Vogel, J., Vogel, D., Cannon-Bowers, J., Bowers, C., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: a meta-analysis. *Journal of Educational Computing Research*, 34(3), pp. 229-243.

- Wellington, W., & Faria, A. (1995). The effect of time pressure, team formation and planning on simulation/game performance. In: Crookal, D., & Arai, K. (eds.). *Simulation and gaming across disciplines and cultures*. Londen: Sage Publications.
- Wilson, K., Bedwell, W., Lazzara, E., Salas, E., Burke, C., Estock, J., Orvis, K., & Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation Gaming*, 40 (2), pp. 217-266.
- Wolfe, J. (1985). The teaching effectiveness of games in collegiate business courses. *Simulation & Gaming*, 16 (3), pp. 251-288.
- Wolfe, J. (1997). The effectiveness of business games in strategic management course work. *Simulation & Gaming*, 28 (4), pp. 360-376.
- Yin, R. (2003). *Case study research: design and methods*. Thousand Oakes: Sage (3<sup>rd</sup> edition).