# GROUP MODEL BUILDING
## A review of assessment studies

**Etiënne A.J.A. Rouwette**
**Jac A.M. Vennix**
**Theo van Mullekom**

Methodology Department
University of Nijmegen
The Netherlands

PO Box 9104
6500 HE Nijmegen, The Netherlands
tel +31 24 3612377
fax +31 24 3612351
E.Rouwette@maw.kun.nl

**Abstract**

This paper reports on work in progress on a review of group model building assessment studies. Previously results were reported at the 1998 System Dynamics Conference. In this contribution we intend to build on and add to the results presented there.

Over the last decades system dynamicists have experimented with approaches to achieve more involvement of their clients in the model building process. The concept of group model building has been coined to describe these approaches. Andersen, Richardson and Vennix (1997) have pointed out that group model building currently is more art than science. There are no standard protocols for conducting a group model building study. Neither do we have a set of robust insights drawn from a systematic, empirical research program attempting to assess the effectiveness of group model building or systems thinking interventions. Assessment studies of group model building interventions do not follow a 'standard approach'. Most researchers follow their own idiosyncratic approach. This is understandable since group model building and studies to assess its effectiveness are a recent phenomenon. However, in order to make substantial progress in the field, the time has come to develop a research program and standards to conduct empirical assessment studies. A logical first step in this process would be to review existing assessment studies and indicate similarities and differences as well as some robust findings.

## A short review of the literature

Since the inception of system dynamics in the second half of the 1950s, implementation of results and system improvement have been its foremost goals. Client involvement has often been considered a useful way of ensuring implementation. The literature on client involvement in system dynamics model building can usefully be ordered around five themes. First, the literature offers a small number of more or less standard approaches for client involvement, detailing how and when clients can best participate in modeling. At least eight approaches can be found in the literature. The oldest is the Reference Group approach (Randers, 1977). Another well-known approach is the Strategic Forum (Richmond, 1987; 1997). The stepwise approach (Wolstenholme, 1992), participative policy modeling (Verburgh, 1994; Vennix, 1996) and modeling as learning (Lane, 1992) build quantitative as well as qualitative models. From the field of soft Operations Research some qualitative techniques are used, e.g. approaches incorporating elements of Soft Systems Methodology (e.g. Sancar, 1987; Bentham and De Visscher, 1994) or cognitive mapping (e.g. White, Ackroyd, and Blakeborough, 1994). Lastly, there is group model building (Richardson and Andersen, 1995; Vennix, 1996; Huz, Andersen, Richardson and Boothroyd, 1997).

Another line of work does not so much provide a step-by-step description of how to involve clients, but goes into the more general requirements for effective involvement. Examples of these studies are the work of Senge (1987), who describes the prerequisites for successful systems thinking interventions, and De Geus (1988) and Morecroft (1988) who describe models as vehicles for learning. Over the last decade, a consensus has grown that the modeling process is an effective tool for fostering both ownership of the model and insight into the problem.

The development of standard approaches for client participation goes hand in hand with a third set of publications: descriptions of single participative modeling projects. These studies are the primary ‘raw data’ for this review, and common features and differences between studies will be the subject of the following sections.

Related to these is the fourth set of publications in which two or more modeling projects are described and contrasted. Roberts (1978a) compares two modeling efforts (at Badger Meter and Sprague Electric) and concludes that differences in implementation of results can be contributed to factors such as the seriousness of the problem that is addressed, and the opportunity for improvement that the problem offers. Frequently, these studies use the cases as a starting point for formulating guidelines. Roberts (1978a; 1978b) for example advises the model builder to communicate extensively with the client, in order to arrive at a correct understanding of the problem. Weil (1980) discusses three modeling studies and arrives at a similar conclusion. In his early projects the emphasis was on the model, the modellers worked more or less independently of the client and interaction was mostly with staff people. The end product of the project generally was a report. In contrast, in the most recent projects the client, who now consists of line managers rather than staff people, was actively involved in the model building process. In addition the project is not concluded with a report but followed-up by transfer of know-how to create an ‘ongoing analytical capability’ in the client organization. In his PhD research, Akkermans (1995, see also Akkermans and Vennix, 1997) compares six cases and tests a number of relationships between organization, problem, process, model, and implementation characteristics. Other recent examples of multiple case comparisons are Lane (1992) and Lyneis (1999). In this review, we will use these multiple case studies in two ways. Project descriptions are added to the database of modeling studies, while the guidelines or conclusions formed on the basis of the comparison of studies will be used to formulate expectations about what elements of involvement are most important for creating implementation. What is important to note about studies of single and multiple cases, however, is that reports are mostly of a qualitative nature. Studies employing questionnaires or other means of quantifying outcome results are scarce.

Another set of studies looks at the process of modeling in more detail. This takes the form of investigating which knowledge acquisition and representation techniques are used by system dynamicists (Fey and Trimble, 1992; Vennix, Andersen, Richardson and Rohrbaugh, 1992; Andersen and Richardson, 1997), the different roles in modeling (Richardson and Andersen, 1995), and guidelines for knowledge elicitation from experts (Ford and Sterman, 1998).

A final group of studies are those in what might be called ‘neighboring disciplines’. Within system dynamics, ‘systems thinking interventions’ seem to comprise both participative modeling and management flight simulators (MFS) or microworlds. Although MFS generally involve the manipulation of a full-blown model, and do not allow for participation in the actual building of the model, they are closely related and seem to be aiming for the same goal: improving mental models and decision making (Hsiao and Richardson, 1999). Two rigorous studies on improvement of mental models show only modest changes in mental models. Vennix (1990) has found that knowledge of structure and dynamics increases after participation in a computer simulation. However, the control group using a traditional approach had comparable results. Doyle, Radzicki and Scott Trees (1998) find evidence that subjects add concepts and relationships learned in a simulation to their existing naïve mental models. After the intervention, the authors still find that models are simple: ‘The most obvious feature of these diagrams, both pre and post, is how greatly simplified they are compared with expert explanations of economic systems and the long wave’ (Doyle, Radzicki and Scott Trees, 1998). On the basis of these and other studies Andersen, Richardson and Vennix (1997: 188) conclude that ‘even after extensive training in modeling, although individual learning occurred, no real improvement of participants’ mental models, in terms of entertaining more feedback loops or more elaborate causal relationships, could be established.’
Other neighboring fields are located outside of the system dynamics community, such as soft Operations Research (Rosenhead, 1989) and electronic technology for group decision support (McGrath and Hollingshead, 1994). Although these fields deserve a closer look and offer numerous insight that are relevant to participative modeling, we might summarize the situation here by stating that the problems of a very diverse set of studies resulting in an unstructured research program, which do not permit conclusive results, are encountered here as well.

In conclusion, there is a growing body of literature on how to involve clients in the modeling process, resulting in an increase in reported projects. By and large, these reports provide qualitative data on implementation results. Quantitative data are found on MFS, which are a related, but different type of intervention. In general, results on

MFS do not seem to offer much hope for effective participative modeling approaches. Both the absence of a structured comparison of results of modeling projects (extending beyond 2 to 6 cases), the lack of a standard approach to assessment, and the inconclusive or even disappointing results of related interventions make a review of existing participative modeling studies a logical step.

## Research questions

In this paper, we will give an overview of results of participation in modeling broadly, incorporating all kinds of approaches somehow involving a client in the model building process. For simplicity, we will use the term 'group model building' for all of the approaches mentioned, i.e. system dynamics modeling which involves the client at least in the phase of conceptualization of his problem (Richardson and Pugh, 1981). Please note that systems thinking interventions that involve a client in the manipulation of a full-blown model, such as in flight simulators or microworlds, are not included in this review. The reasoning here is that the client should be able to change the structure of the model before we can speak of participation in model building.

Most of the authors mentioned above provide explicit grounds for designing their approaches the way they do. Pivotal to all of these is the urge to maximize the opportunity for clients to contribute their ideas and learning from the process of modeling. In this way the model is an aid in the communication about the problem, and hopefully brings the participants to adopt more of each other's ideas and thereby creating more alignment of insight. Many more specific ideas on intervention – outcome relations can be found in the literature. There is for example the discussion on qualitative versus quantitative models, with some authors being content with limiting their modeling to the conceptualization phase only, and others setting out to achieve full quantification. Another theme that resurfaces from time to time is the discussion on the size of models. Senge (1987) sees a number of benefits in small models, whereas for example Lyneis (1999: 52) agrees on the ability of small models to foster insight, but still feels that 'As surprising as it may seem, the selling of results (as opposed to understanding) is easier to accomplish with a detailed, calibrated model than with a small model.' Other relations between elements of process and outcome are described by Akkermans (1995). In effect, research into the effective elements of the intervention comes down to breaking down results into smaller categories of projects, and comparing results between categories. This is an important element of any review of group model building projects, which is mentioned in literature on evaluation research and in reviews of group decision support methods. Pawson and Tilley (1997) argue that a realistic cumulation of evaluation studies comes down to discovering which combinations of mechanism and context lead to which outcomes. McGrath and Hollingshead (1994: 78) start their review of results of group decision support technology with the statement that outcomes will inevitably be a joint function of contextual and intervention characteristics. For group model building the same can be expected. So although there are important reasons for looking at patterns in results, in this paper we will limit ourselves to outcomes in general, i.e. over all categories of projects. In a follow-up study, we will try to identify meaningful patterns in the studies reported.

In this review, we will be concerned with only a limited number of outcomes. From the previous discussion on implementation, the picture emerges that modeling is more and more seen as an aid in communication about the problem, which stimulates participants' learning about problem structure. It is expected that if clients learn about each other's points of view, and use models to transfer and combine these insights, a shared view or consensus will eventually emerge. These results at the individual and group level are expected to contribute to the most important goal of system dynamics modeling: improvement at the system level. In an extensive assessment study, Huz, Andersen, Richardson and Boothroyd (1997) cover all of these levels. Using a questionnaire, they assess perceptions of the intervention, shifts in goal structure, change strategies, alignment of mental models and understanding of the system. In addition they include measurable system changes and results. In our review, we focused specifically on the following outcomes: participants' reaction to the intervention, insight, commitment, behavioral changes, communication, creation of a common language, consensus, system changes and results of system changes, and further use of modeling.

In the following we will give an estimation of robust results. Before we turn to results we will describe the selection of cases, the database formed on the basis of these cases, and general characteristics of the studies in the database.

## Selection of cases

We conducted a literature research for publications on group model building in the International System Dynamics Conference Proceedings (1981 to 1997), System Dynamics Review (1985 to Winter 1997), and publications on system dynamics by Productivity Press. These sources were reviewed by one author each, who subsequently selected relevant publications. Publications were deemed relevant if they described a system dynamics modeling project, involving client participation in at least the stage of conceptualization (following the definition of group model building in the above), and empirical results were described. Although the procedure for gathering cases makes it possible that specific studies are overlooked (either because they were erroneously excluded, or because they were reported in other sources than the ones used here), we feel that there is no reason to expect the collection of studies to lead to systematic biases. More troubling is the fact that the sources from which cases were gathered can be expected to be biased in certain respects. Of all the projects carried out, only a limited number is described in official journals and proceedings; e.g. projects conducted by consulting firms might be prevented from publication for reasons of confidentiality. This effect can be checked by comparing results for different classes of modeling projects. However, this paper focuses on description of general results and the identification of patterns in results will be the next step in the analysis of data. Another cause for concern is the limited number of studies reporting negative results, indicating a bias towards positive studies. There is less reason to expect this bias in studies with a focus towards research, in which the outcomes to be measured are defined in advance. The validity of case reports will be taken up again in the section on evaluation.
Each of the publications was scanned for separate group model building projects, which then formed the cases for our review. For example, Lane (1992) describes four

modeling projects, two of which are described elsewhere. In total, we catalogued over 75 publications describing 81 cases.

## Database

The data we gathered on each modeling project can be ordered in five general categories:

- background characteristics;
- client organization;
- problem to be modeled;
- intervention;
- evaluation of modeling impacts.

Background characteristics of each publication that were noted are authors, title, and source (journal or proceedings) of publications. As stated above, if a modeling project was reported in more than one publication, the information from all publications found was combined.

The following characteristics of the organizations involved were stored in the database: sort (profit, nonprofit or governmental), sector (e.g. energy or financial services), name and size of the client organization, and name of the consultant organization.

Problem elements that were thought important to include in the analysis are background (motive for starting the intervention), the research question that the modeling effort focused on and its type, the importance of the problem modeled as judged by participants, and whether or not implementation of results was expected from the outset of the project. The types of research questions are borrowed from a typology of fundamental and applied research (Swanborn, 1987). An open question aimed at uncovering elements related to the subject of the study, was marked as 'exploratory'. If the aim was to identify facts or delineate a state of affairs, the modeling question was categorized as 'descriptive'. If the model was used to identify the causes or reasons for a situation or development, it was termed 'explanatory'. A 'prescriptive' focus was one in which a concrete action to bring about change was sought after.

The following aspects of the intervention itself were coded. First, we recorded a more or less open description of the different techniques or phases employed for building the model. The model itself was characterized as qualitative or quantitative, by size, whether a preliminary model was used or the project started from scratch, which phases were followed, and in what phases the client participated. In addition, we recorded the number and function of participants involved. Finally, the database contains the sources of information for building the model (apart from persons or groups, these could be documents, real life situations, or models – system dynamics or otherwise), the software used, other materials, and the total time span of the intervention.

The most elaborate category in the database is the evaluation of modeling results. A first entry stored the content of the evaluation, or the variables that were reported. The design of the evaluation was recorded as follows (Swanborn, 1987; Cook and Campbell, 1979):

- an experiment, using pretests and posttests, a control group and random assignment of individual subjects to the experimental or control group;
- a field experiment, using pretests and posttests, a control group but no randomization;
- a one-group pretest-posttest, identical to the above but without a control group;
- a survey, involving only a posttest;
- a case study, if the description of the modeling process is focused on the project and its setting.

Data collection methods for deriving conclusions about the intervention's effects, were coded as observation, content analysis, questionnaires, individual interviews, or group interviews. The subject (number and function of persons involved in the evaluation) and object of the evaluation (who conducted the evaluation) and the time span covered, were recorded. Record was also kept of the way in which the client received feedback on the results of the modeling project, e.g. in a written report or presentation.

The conclusions of the evaluation, or the results of the intervention, were first recorded in a general form, as much as possible in the original wordings of the authors, including references to pages. Subsequently, results were summarized as in the following table. The keywords used in the table are used as a representation of the most important results of group model building, as formulated in the section on research questions.

| | |
|---|---|
| Individual | [positive reaction] or [negative reaction] personal evaluations of the intervention or model (e.g. ownership, discomfort, trust)<br>[insight] or [no insight] learning<br>[commitment] or [no commitment] a decision or commitment to results<br>[behavior] or [no behavior] changes in individual behavior or implementation of conclusions |
| Group | [communication] or [no communication] exchange of viewpoints<br>[consensus] or [no consensus] a shared view of the problem or actions<br>[common language] or [no common language] understanding of other participants |
| Organization | [system changes] or [no system changes] organizational or physical changes (e.g. production lines, personnel policies)<br>[positive results] or [negative results] results of these system changes (e.g. for profit or morale) |
| Method | [further use] or [no further use] further use of system dynamics methods<br>[efficiency] or [no efficiency] intervention elements or contextual factors that fostered or hampered the effectiveness of the intervention |

An entry was made in the categories above only when an author reported on this aspect. If, for example nothing is said about system changes, this category is left open; [no system change] is only recorded if the author explicitly states that no changes at the system level were implemented. The decision to use a 'bipolar' coding followed from the general nature of the studies collected. A large number of studies revert to rich descriptions of the phases and conclusions reached in the modeling process. Only a small subset uses quantitative measures of results such as Likert-type questionnaires. These quantitative results are recorded separately after each keyword. For example, in the study of Sancar (1987) [positive reaction] is recorded as a keyword, after which the score on ‘enjoyed participating’ (5.75) is noted also. In the analysis of results, qualitative and quantitative results will be contrasted.

We used the following coding procedure to determine the entries in the database. First, one author read the complete publication and entered his conclusions in all fields described in the above. Second, one of the other authors reread the publication and checked the entries, with specific attention for the problem, intervention, and evaluation categories. The entries under results were filled out after readers agreed on how to code a specific statement. For example, this included the decision to file 'star shells of insight burst around us' (Hickson, 1978: 482) under [insight].

## General characteristics of the cases

*Background and organization*

The first empirical study into the effects of client involvement in system dynamics modeling that we found, dates back to 1963. From our literature review it appears that only one other case was published before 1970. In the 1970s, a total of four cases appeared, growing to 14 in the 1980s. From 1990 to 1998 between four and 12 cases appeared in print each year, indicating a fast growth in the number of publications on this issue. About half of all group-modeling studies were conducted in profit organizations, one quarter in non-profit settings, and another quarter in governmental institutions. Cases in profit organizations are less likely to be published than cases in government and non-profit organizations, because of proprietary rights. Non-profit organizations were mainly universities (six), other education institutes (four), research groups in defense (two), energy (one), and regional development (one), and a charity institution. Profit organizations can be grouped into production organizations, distribution organizations, and services. Eleven modeling projects were done in oil producing companies, five in chemical companies, two in shipbuilding, one in food products, one in biotechnology, and another seven in various other production companies. Four studies were conducted in distribution. Service companies were working in insurance (seven), software (five), finance (three), hotel services (one), and sports (one). The size of client companies ranges from a few members to (divisions of) large multinationals with revenues in the area of 100s millions of dollars.

*Problems*

Problems that are not perceived as particularly pressing by participants in a modeling project, are more often initiated as a training or demo exercise, or alternatively start because champions working within the organization want to improve a situation that is not urgent in itself. The latter comes about either because the initiators have been recently exposed to system dynamics tools in a training, or because internal consultants suggest these as tools for continuous improvement. In one case a project that started out as quite urgent, addressing a matter of considerable importance to the client organization, became less urgent because a more serious matter (a merger) developed in another place of the organization (Verburgh, 1994). What is striking is that participants are never expected to be informants only. In all cases there is an additional focus on participants’ learning and commitment. In the study by Bronkhorst, Wiersma and Truin (1991) a scenario study on health care developments is initiated by the Dutch Department of Public Health. A scenario committee is formed in which dentists, dental hygienists, economists, and members of dental health care institutions take place. Although the participants are expected primarily to contribute to valid scenarios, their learning and commitment to the results is treated as an outcome in itself and a support for the conclusions of the study. Group model building studies in which no implementation of results is expected are, not surprisingly, also mostly conducted in a training environment.

About 75 percent of the modeling studies are aimed at finding an implementable solution. The only non-training studies that are not prescriptive in orientation, are aimed at discovering relevant environmental developments (scenario-studies by Genta, Kreutzer, Anderson, Hinote, Hood and McMillan, 1994; Morecroft and Van der Heijden, 1992; Rufat-Latre, 1994), or policy impacts (Rohrbaugh, Andersen, Richardson and Zagonel-Santos, 1997). As can be expected, all projects aim to explain situations or developments over time. System dynamics is sometimes credited for its free format, in which models are usually started with a 'blank paper' instead of pre-fixed notions on elements that have to be included (Coyle, 1998). However, only five studies can be said to have an explicit exploratory orientation from the outset. Most studies depart from specific hypotheses on the causes of problematic behavior, and add additional structure when needed.

*Elements and scripts*

In a previous section the wide variety of group modeling techniques was described. The empirical studies gathered here are a clear reflection of this. Among the techniques used are hexagons, brainstorming, nominal group technique (Delbecq et al., 1975), Delphi, groupware, and elements of soft systems modeling or cognitive mapping. A number of visual representations serve as group memory: causal loop diagrams, stock&flows diagrams, and graphs of developments over time. In building a model, participants usually perform three types of cognitive tasks (Vennix, Andersen, Richardson and Rohrbaugh, 1992): elicitation of information, exploring courses of action of convergent tasks, and evaluation. From the modeling studies it appears that the elicitation phase is supported by using individual techniques such as interviews, cognitive mapping, nominal group technique, or workbooks. Alternatively, elicitation of information is done in small subgroups (Andersen and Richardson, 1997). The elicitation phase might be started after the problem to be addressed is agreed upon, but can also consist of discussing and adaptation of a preliminary model. This pre-made model can be qualitative (e.g. an archetype in causal loop format) or quantitative, ranging from several variables and loops to a model of substantial size (e.g. Verburgh, 1994). In the latter case, the group modeling part is reduced to commenting an already existing structure, and the difference with using a management flight simulator becomes small. In about one quarter of the studies gathered here, a preliminary model was used. In the majority of cases, this was a quantitative model (of 14 studies that described the preliminary model in detail, 12 used a quantitative model). In two instances a qualitative model was presented at the start of the project, both representing an archetype.

Convergent tasks are those in which participants make a choice between problem formulations, model structure, or policy options. These tasks require the input and confrontations of the group of participants as a whole (Andersen and Richardson, 1997; Vennix, 1996). This phase mostly takes the form of a face-to-face discussion, although the Delphi method (Vennix, Gubbels, Post and Poppen, 1990) and

GroupSystems (Rouwette, Vennix and Thijssen, 1997) are used also. The evaluation phase also requires the group as a whole to discuss and agree on issues, although individuals and subgroups are used to prioritize issues (Andersen and Richardson, 1997).

*Number of participants and time investment*

The number of participants involved in face-to-face interaction is mostly between 5 and 12, and seldom larger than 22. If more people are involved, they mostly work in subgroups that meet at regular intervals to present findings to each other. In some cases a group of 30 to 70 participants is involved using hexagon brainstorming, GroupSystems, or a management flight simulator. In total, there are more than 660 participants involved in the group model building projects reported (in reality there are more as some studies do not report on the exact number of participants). Apart from students building models for training purposes, most participants are line managers. Members of staff or other experts participate in some projects. The time between start of the project and handover of final results varies between two full days and five years. Of the 53 studies provide detail on the duration, nine are between 2 days to three weeks, another nine are three to eight weeks, seven are between eight to sixteen weeks, and 23 last from 16 weeks up to five years.

Most projects take the form of two to four workshops, with intermediate feedback and reports, for example in the form of a workbook. Workshops may be an intensive full-day meeting, or consist of two to three hours of model building. The hours the client is involved in building the model is specified in only a few studies. Verburgh (1994) provides an estimate: over a period of six months, the client group participated in four sessions of about two hours each, and another five hours on answering questions in workbooks (a total of 13 hours per person). The Andersen and Richardson (1997) intensive two-day design involves the client for about 18 hours. Naturally, the longer the project duration, the more hours a client is involved; e.g. Hines and Johnson (1994) involve participants in 12 full-day sessions. The qualitative modeling facilitated by Vennix (1995) took three sessions of three hours each, and three workbooks on which participants spent about one hour each. We might tentatively conclude that a minimum client involvement of 13 – 18 hours is necessary for building a full-blown quantitative model.

*Involvement in stages of the model building process*

This raises the issue of the phases of model building a client is involved in. Although a number of studies discussing the elicitation of variables and building of the conceptual model have been mentioned, the literature offers few guidelines for the involvement of the client in the quantitative part of modeling. Morecroft (1992: 13) proposes to use friendly algebra to enable a client without modeling expertise to understand and contribute to this phase. Ford and Sterman (1998) ask experts on the problem to draw graphs with problematic behavior over time and estimate graph functions. Of the studies collected here, 63 use fully quantified models, while in 18 studies a qualitative model is build. Of 52 quantitative model building projects with non-students, only 31 explicitly mention client involvement in the formalization phase. In eight cases, members of the client organization possessed modeling expertise themselves and in effect built the complete model. In another nine cases, the consultants developed the formal model and presented the algebraic formulations to participants, who could then suggest changes. Participants in 12 studies contributed to the formalization phase by estimation of parameters, sketching variables over time, or other forms of data gathering. In two instances the intervention lasted over a period of several years and involved almost the complete organization in data gathering and estimation of parameters (Cooper, 1980; Weil and Etherton, 1990). Apart from the client, other information sources such as documents and observation of real life situations are reported in a small number of cases.

*Size of the model*

There is some discussion in the literature about the appropriate lower and upper bounds on the size of system dynamics models. Senge (1987: 875) discusses the benefits of very simple formal models 'involving only one stock variable and virtually no significant feedback loops' in the direct interaction with clients. Lyneis (1999: 45) feels that a model 'would probably need a minimum of several feedback loops and 20-30 equations' but also states that small Pugh-Roberts models contain 200- 400 equations. He assumes that the appropriate size partly depends on the experience of the modeling team. With more experience, insight can be gained from large models more quickly. In addition, the different lower bounds mentioned by different authors might be due to the fact that Senge refers to models used in direct interaction with clients, while Lyneis seems to discuss the use of models for giving insight to a (more or less) experienced modeling team. As an upper bound, Morecroft (1985: 16) suggests 100-200 equations and Hines and Johnson (1994) 200- 400 equations. The size of models built falls between an upper range of several thousand (two models), and a lower range of 5 to 19 variables (three models). Most models are either 20 to 50 variables (17) or 50 to 200 (14). Another three models are anywhere between 200 to 1000 variables.

## Research designs of studies collected

Before the results of the group modeling studies can be addressed, the way in which results are measured needs to be discussed. The criteria that need to be considered in assessing the value of research results can be grouped into criteria on internal and external validity (Yin, 1994; Hutjes and Van Buuren, 1992). Internal validity concerns the following:

- Concept validity: are concepts in the theoretical framework an accurate representation of the concepts in reality?
- Internal validity: are the relations in the theoretical framework an accurate representation of relations in reality?
- Reliability: to what extent are the results influenced by variations in measurement instruments or subjects?

External validity refers to the extent in which results can be generalized to a larger population. The field studies collected for this study are for the greater part 'real life' applications of group model building, which facilitates the generalization of results to this class of applications. However, some authors express doubts about the internal validity of studies into systems interventions.

In more than three quarters of the studies collected here, claims about the effects of the interventions are largely based upon reports of individual clients or on group discussions. There are several reasons to doubt whether this procedure results in an accurate estimation of results. Weil (1980: 273) remarks that mutual face-saving operations could lead both client and consultant to play down the importance of negative outcomes. Doyle (1997) and Doyle, Radzicki and Scott Trees (1998) criticize the common practice of evaluating systems thinking interventions by using retrospective self-reports or group sessions. Although field studies have a high degree of external validity, experimental control is low, meaning that any changes after the modeling process might be due to other factors (e.g. external developments in the problem, information gathered outside the modeling sessions) as well. Most modeling projects are aimed at improving participants' insight into a problem, which might preclude an accurate pretest of knowledge about the problem. If for example participants are interviewed using causal modeling, this elicits as well as structures (and therefore changes) mental models of interviewees (Morecroft, 1992). Doyle, Radzicki and Scott Trees (1998) also address the dangers of using group sessions as evaluation instruments: domination of individuals, compliance, the project leader steering the discussion, and the difficulty of addressing long-term memory in a group session could bias results. The authors also underline the necessity of measuring change instead of reported change (self-reports might be biased in the direction of what subjects think the researcher wants to know), exact specification of expected changes (preferably quantitative), and the use of a sufficient number of subjects to obtain statistical power.

Although the above presents a set of serious objections to relying on participants' retrospective self-assessment, we feel there are some considerations that increase the value of these studies as sources of information on group modeling results. Before throwing out the baby with the bath water, we need to consider the content of these modeling studies in more detail. Most accounts of modeling studies describe the interaction of the client group, organization and problem elements, leading them to design the modeling project in a certain way, and the conclusions on the basis of the modeling results. This is close to the description of a phenomenon in its natural context, which is a common definition of a case study (Yin, 1994; Hutjes and Van Buuren, 1992). Although individual case studies run most of the risks addressed by Doyle, Radzicki and Scott Trees (1998), there are several measures that increase their value as accurate sources of information. We will discuss the elements of internal validity briefly.

Concept validity requires finding a set of indicators to measure a concept and a careful operationalization. Hutjes and Van Buuren (1992) claim that in a qualitative case study the meaning of a concept might change (due to a continuous development of the theoretical framework), but at least at the conclusion of the study a precise definition of a concept should be given.
Most modeling studies collected here do provide a base for their respective claims, e.g. by providing quotes of participants' reactions, session reports, or otherwise. So, in general, the conclusion 'the client has gained insight', is accompanied by a specific statement such as 'the impact of the reward system in use on quality of work was shown by the model, and the client had never realized the extent of this impact before'. In the database, these specific statements have been recorded also, which makes it possible to check on the inference made by the author. Although this does not preclude different authors from using different definitions of for example 'insight', it does provide a way of going back to the data evaluative claims are based on.

Internal validity is concerned with the question of whether the theoretical relationships correspond to the empirical relationships, and the most important aspect is the relationship between the intervention (as one total 'package') and the outcome. Claims about the intervention - outcome relationship are strengthened if an effect is not found in a control group. In case study research (Yin, 1994; Hutjes and Van Buuren, 1992) one way of checking the validity of relationships a researcher claims to exist, is by careful explication of the reasoning that lead the researcher to her claims. Internal validity in this design is enhanced through the transparency and coherence of the web of reasoning behind the claims. In addition, the reasoning process can be checked by fellow researchers (peer debriefing) and participants in the researched setting (member check).

As stated before, in the model building studies collected, most attention is given to the process of modeling which is thought to provide a consistent and clear way of addressing a client's problem. There is much less theory on the link between modeling results and a client's change in cognitions and behavior, which causes authors to apply idiosyncratic labels for results, and define results after the fact instead of operationalizing them in advance. However, certain aspects of the reports collected facilitate assessment and accumulation of results. The specific question that initiated the modeling project does provide a direction in which results are expected before the intervention takes place. In a considerable number of studies (33) measurable system changes or results of system changes (19) are reported. In 21 cases changes in participants' behaviors that are in principle observable by others, are reported. Also, a number of publications (11) are co-authored by one or more clients in the modeling project. Although this of course leaves open the possibility of biases due to mutual face-saving, it does present one way of member check.

Reliability concerns the absence of variation in the results due to chance. Often this is translated to the expectation that an intervention should have a certain outcome *consistently*. If a certain technique, method or approach is shown to have a consistent outcome, this provides a strong reason for expecting a causal relation between intervention and outcome. However, Pawson and Tilley (1996) question the usefulness of this practice in social research. If one and the same intervention is applied in subtly varying circumstances, its results might not be consistent. Following the latter reasoning we would then have to conclude that the intervention has no effects. According to Pawson and Tilley a better way is to specify the context in which certain outcomes of certain interventions are expected, and build up the knowledge about an intervention by further and further specifying of CMO (context - mechanism - outcome) configurations. However, they explicitly state not to take this as a ground for bypassing experimental control, control groups and careful operationalization, as these offer the evidence on which to base claims about effects. Because a case study can seldom be conducted in the same manner again, Yin (1994) urges researchers to keep a close guard on the raw data on the basis of which they draw their conclusions. These should be kept separately, and, if possible, quantitative data should be used to back up statements.

In this study, the variety of context and mechanism combinations is very large. We therefore feel that the reliability of conclusions mainly depends on the identification of a recurring pattern in results - combinations of problems and clients (context) and modeling projects (mechanisms) that consistently yield certain results. As stated before, in this paper we will limit ourselves to a description of the data at a general level. In addition, it is useful to investigate whether results are dependent on the measurement method. We will therefore contrast the qualitative results of case studies with quantitative evaluations of studies using surveys and pre and posttests.

From the perspective of CMO configurations, reliability and external validity have a close resemblance. If the first is about finding consistent results for consistent combinations of mechanisms and contexts, the second concerns the question to which CMO configuration results can be generalized. In case study research, instead of striving for comparable research cases, cases that are dissimilar with respect to certain characteristics are sometimes used to test claims about generalizations. Using a 'most unlikely case' (Yin, 1994) we could for example test the use of group model building in a highly politicized organization. If the approach has beneficial results even in that context, the conclusion is strengthened that the politicality of the situation does not preclude the use of group modeling. The use of multiple cases is especially advanced as a way to test the external validity of case studies.
In this study we will draw conclusions on the basis of a comparison of several cases of modeling, and in effect are thereby conducting a multiple case study. Again, the identification of CMO configurations provides the base for any inferences, as the identification of critical elements in the context precedes the identification of any 'most unlikely case'.

## Results

In this section we will compare the results of all the studies collected in the database so far. We will group results into the following categories:

- reaction;
- insight;
- commitment and behavior;
- communication;
- common language;
- consensus or mental model alignment;
- system changes, results, and further use of modeling.

Before starting a comparison of assessment studies, three remarks have to be made. First, the authors of most of these studies did not set out to assess their modeling projects on all aspects contained in this review. Authors might not report certain outcomes because they are deemed of less importance, e.g. reactions. For the purpose of this review, this means that most reports are incomplete. However, we feel that in particular change in mental models, commitment, behavior, and system changes represent common goals for system dynamics modellers, which makes it likely that authors would report on just these aspects. These represent goals that group modeling approaches set out to influence, and are therefore likely to be included in a case report.
Second, the limitations of the data have been discussed earlier but need to be emphasized again. More than three quarters of the projects (67 in total) collected can be described as case studies, which establish results at the individual and group level in a qualitative manner. All of the case studies use observation for data gathering, five studies include individual assessment interviews and one uses a group interview. We found 14 studies that use a quantitative estimation of results. In 12 studies a survey design is used, and results are established using a questionnaire after the intervention.

In these cases results are measured using clients' subjective self-assessments. Only two studies (Verburgh, 1994; Huz, Andersen, Richardson and Boothroyd, 1997) use an objective measure of insight and consensus. Verburgh uses a one group pretest-posttest design, and Huz, Andersen, Richardson and Boothroyd a field experiment.
Third, after describing general results per category, we will compare the results of quantitative and qualitative assessment studies and discuss similarities and differences between results. If different research methods arrive at different conclusions, differences between operationalization of concepts or measurement method can be looked at in more depth. Still, we feel that these results should not be interpreted as more than an indication, and any statement on recurrent outcomes of modeling should be based on more in-depth studies using more elaborate designs.

The following table provides an overview of all concepts measured in quantitative assessment studies. Note that the results in italics are established qualitatively.

| **Case** | **Concepts measured** |
|---|---|
| 1. Akkermans case 2 (software services A) | insight, commitment<br>communication, consensus<br>*no system changes* |
| 2. Bentham and De Visscher (Shell) | positive reaction, insight, behavior<br>consensus<br>*system changes, results*<br>*further use* |
| 3. Berkvens and Neomagus (Shell) | insight, commitment<br>communication, consensus<br>efficiency |
| 4. Cavaleri and Sterman | insight, behavior<br>system changes |
| 5. Huz, Andersen, Richardson, and Boothroyd | positive reaction, insight, behavior<br>communication, consensus<br>system changes, results |
| 6. Rouwette, Vennix, and Thijssen | positive reaction, no insight, commitment<br>communication, consensus<br>efficiency |
| 7. Sancar case 1 (Door County) | positive reaction, insight, commitment, no behavior<br>communication, consensus<br>*efficiency* |
| 8. Sancar case 2 (Janesville) | positive reaction, insight, commitment<br>communication, consensus |
| 9. Vennix (DGSM) | insight, behavior<br>communication, consensus<br>*system changes* |
| 10. Vennix, Gubbels, Post, Poppen (health care) | insight |
| 11. Vennix, Scheper, Willems case 1 (Department of Transport and Public Works A) | insight, commitment<br>communication, consensus<br>efficiency |
| 12. Vennix, Scheper, Willems case 3 (Department of Transport and Public Works B) | insight, commitment<br>communication, consensus<br>efficiency |
| 13. Verburgh (health care insurance) | insight<br>no consensus<br>efficiency |
| 14. Wallace and Sancar | positive reaction, insight, commitment |

The following table presents an overview of the results for the qualitative and quantitative assessment studies. Note that the figures in the table are estimated by counting all studies which mention a specific result, e.g. in total 21 studies report on reactions, of which all 21 are positive and none is negative, and 60 studies do not mention reactions.

| **Comparison of qualitative and quantitative assessment of outcomes** | | | | | | |
|---|---|---|---|---|---|---|
| | *Qualitative assessment (n = 67)* | | *Quantitative assessment (n = 14)* | | *Total (n = 81)* | |
| reaction | positive<br>negative | 16<br>0 | positive<br>negative | 5<br>0 | positive<br>negative | 21<br>0 |
| insight | insight<br>no insight | 54<br>3 | insight<br>no insight | 13<br>1 | insight<br>no insight | 67<br>4 |
| commitment | commitment<br>no commitment | 20<br>3 | commitment<br>no commitment | 7<br>0 | commitment<br>no commitment | 27<br>3 |
| communication | communication<br>no communication | 22<br>1 | communication<br>no communication | 9<br>0 | communication<br>no communication | 31<br>1 |
| consensus | consensus<br>no consensus | 18<br>2 | consensus<br>no consensus | 9<br>1 | consensus<br>no consensus | 37<br>3 |

*Reaction*

As is shown in the previous table, in only about one quarter of studies are positive reactions of clients towards the results of the modeling process noted. In none of the studies a negative reaction is noted.

In 16 of 67 qualitative studies and 5 of 14 quantitative assessment studies, positive reactions of clients towards the results of the modeling process are noted. Considering the small number of quantitative studies included, we take these figures to indicate comparable results. On the basis of these data we conclude that in a minority of cases, group model building results in a client's positive reaction (this result is probably a low estimate as reactions are less likely to be included in any assessment than are the other outcomes).

*Insight*

A total of 71 out of 81 cases contain an assessment of the amount of insight gained. The fact that an outcome is included in almost all studies, strengthens the case for any claim about robustness. In 67 cases the result is positive, indicating that group model building resulted in an increase in insight. In four cases no insight was gained. Two of the four cases in which no insight resulted are projects in which models are built with students. In one case, the aggregation level of the model did not correspond with the mental models of students. The model was too abstract (Ginsberg and Morecroft, 1995). In the modeling course facilitated by Rouwette, Vennix and Thijssen (1997), participants gained only moderate insight into the problem, and no insight in each other's assumptions. This is explained by the focus on document analysis for data gathering and the lack of discussion between students about the problem. In the two cases in which models around real life-problems did not lead to insight, the model was too big to understand (Fey, 1978) or the issue was politically sensitive and too broad to achieve focus (Akkermans, 1995 case 4). Broadly stated, these four studies share a mismatch between the level of abstraction of the system dynamics model and the clients' mental models, and the modeling techniques used do not match the project's circumstances (unstructured discussion on a political sensitive issue, individual data collection hampering learning about others' opinions).

In conclusion, if the level of abstraction is adequate and techniques are matched to the objectives of the study, group model building studies generally result in increases in insight. On the basis of the data gathered, the amount of increase is difficult to determine. The issue is further confused by the difficulty of establishing what counts as a 'large' or 'sufficient' increase in insight. In cases aimed at finding implementable solutions from the outset, the increase in insight is probably best considered in relation to behavioral and systemic changes: if the clients succeeded in finding a solution to his or her problem, we assume that the insight gained was adequate and sufficient. However, the space of this paper prevents an in-depth discussion of this matter.

Insight is reported in 54 of 81 qualitative cases and 13 of 14 quantitative assessment studies. Only in three studies in the first group and one in the second did modeling not lead to insight. This result presents a contrast to the outcomes of management flight simulators discussed earlier, for which only moderate increases in insight were found. There are at least two explanations for this difference. The majority of the changes in insight found in this review are based on qualitative data or on self-reports. Had the same changes been measured in a more objective manner, they might have been reduced or even have disappeared altogether. In addition, a simple estimation of 'more insight' clouds the issue of how large an increase in reality is. However, the two studies applying objective assessments of insight (Verburgh, 1994; Huz, Andersen, Richardson and Boothroyd, 1997) show a significant increase in insight after the modeling sessions. A second explanation is that flight simulators present a different sort of interventions than group modeling projects, e.g. because of their lower time-involvement and lower relevance to urgent problems. The difference found in insight could therefore also represent a real difference in the effect of both kinds of systems thinking interventions on mental models. The large differences between interventions incline us to interpret these results as indications of real differences.

*Commitment and behavior*

In only 30 cases an influence on commitment is reported. In the majority of cases (27) commitment to the results of the modeling effort is created, but in three instances clients indicate that they do not feel committed to the study's outcomes. This may seem disappointing at first sight. However, we have to consider the fact that only 63 studies focus on implementable results from the start of the project. In 12 cases models are built for training or educational purposes, in which no implementation of conclusions is expected. In addition, even if clients in the modeling process are managers working on their own problem, a change in behavior is not always the objective. The conclusions of the modeling process might indicate that no behavioral changes are needed, or that adaptations on other ('systemic') levels of the organization are preferred. Also, some of the descriptions are made immediately after the project which might be too early to reach conclusions about implementation. In the instances where authors report that the project did not lead to commitment, we find the political issue again (Akkermans, 1995 case 4). In two other cases, management agreed on the analysis of the problem, but decided not to back up conclusions nevertheless (Watts and Wolstenholme, 1990; Raynolds and Raynolds, 1992).

For behavioral changes the results are comparable. In 21 studies projects are followed by changes in behavior, and in one instance modellers report not to have affected behavior. In this last study participants in general agree to the statement 'the diagrams imply solutions', but the management (who did not participate in building the model) does not implement conclusions because the model did not include all relevant aspects.

In conclusion, about half of the studies aiming at implementable solutions result in commitment and behavioral change. Although only in few projects clients state that they are not committed to results or do not rush to implement conclusions, the effect of group model building on individual behavior seems to be surprisingly small. Potential reasons for the gap between the number of studies focused at implementation and the reports containing behavioral changes, are that results at other organizational levels are aimed for, or that the evaluation was conducted too soon after the project for assessing any behavioral changes.

In 20 of 67 qualitative studies and 7 of 14 quantitative studies, commitment to the results of the modeling effort is created. In a very small number of cases (3 and 0 respectively), clients indicate that they are not committed to the study's outcomes. That quantitative studies find far more cases in which commitment is created than qualitative studies, is a possible indication that the concept is interpreted differently in both fields of study. In qualitative studies statements such as 'the client agreed on implementing result A' is categorized under commitment, whereas for example the questionnaire used by Vennix, Scheper and Willems (1993) also contains questions such as 'If I, with people from my organization, were to use the same approach in planning and in dealing with problems, all persons would loyally follow this plan to its natural conclusions.'
It seems clear that commitment to results found with regard to one specific problem is a much narrower definition than commitment as a result of group model building in a more general sense.

*Communication*
The results for communication are as follows: in 32 out of 81 cases an influence on communication is reported, of which 31 resulted in an increase in the quality of communication and one instance quality of communication is reported not to have increased (the unsuccessful case mentioned in the above, Akkermans, 1995 case 4). Most striking about this result is the low number of reports on communication, and the overall positive outcome.

In group model building stakeholders in a problem are brought together and a tool for decision making is applied that is new to most people involved. If model building is applied in a training environment this also presents a new way of dealing with a problem. A change in communication between participants in modeling therefore seems natural. Increasing the quality of communication is a central goal of most participative modeling approaches, which makes it seem unlikely that this result suffers from being overlooked. In addition, applying a new tool such as model building can be expected to immediately effect communication, which makes it unlikely that an assessment study fails to note its impact because it is limited to short-term outcomes. A possible explanation is that in applying a new tool for decision making, changes in communication are unavoidable and therefore are not detailed, in order to avoid 'stating the obvious'. The fact that in 15 cases in which nothing on communication is reported, consensus does result (in two cases no consensus resulted), points in the same direction. That exchange of viewpoints or communication is a necessary condition for consensus seems logical, and this assumption is made in the literature on decision support as well (e.g. Scheper, 1991). In 16 cases, results are reported at the individual level only.

The comparison of quantitative and qualitative assessments of communication resembles that for commitment: in quantitative studies more cases resulting in high quality of communication are found than in qualitative studies (9 out of 14 versus 22 out of 33). In one qualitative assessment study modeling had no influence on communication. Instead of a difference in conceptualization, this suggests to us a difference in focus between qualitative and quantitative assessment. Most questionnaire-based assessment studies contain items on communication, whereas in only a minority of qualitative studies an influence on communication is mentioned.

Overlooking these results, we tend to weigh the small number of reports less heavy than the overall positive result in cases were an effect on communication is reported. In 31 out of 32 reported effects on communication, quality of communication is increased. Our tentative conclusion is that in general, group model building leads to an increase in the quality of communication between participants.

*Common language*

An effect on common language is reported in only a minority of studies: 8 out of the total of 81 studies. In six cases this effect is positive, and in two cases it is explicitly mentioned that no common language resulted. One of the two cases in which no common language was reported, was the political sensitive issue described by Akkermans (1995 case 4). The other case (Zazara and Fisher, 1996) reports on the development of cross-curricular models with teachers of pre-college students. After three weeks of training, 70% of teachers uses modeling in their classes, but all of them use models specific to their own discipline. The interdisciplinary models were not used.

The low number of reports and their mixed character, makes it hard to draw any conclusions about this outcome. In the system dynamics literature, the expectation can be found that system dynamics serves as a sort of uniform platform for communication (e.g. Richmond, 1987: 132). On the basis of the studies gathered here, this expectation can neither be confirmed nor disconfirmed.

*Consensus or mental model alignment*

In 40 of 81 studies an influence on consensus is reported. In 37 cases a consensus view has been created, but in three instances clients indicate that there is no consensus on the conclusions of the modeling project. Two of the studies reporting no consensus have been discussed before (Akkermans, 1995 case 4; Ginsberg and Morecroft, 1995). Verburgh's (1997) objective assessment of mental model alignment shows no significant increase between pretest and posttest. This again presents a picture of a limited number of reports of an overall positive character. We feel that these results are in part due to the lack of a clear definition of consensus.

In 18 of 67 qualitative assessments studies, and in 9 of 14 quantitative studies consensus is reported. In two and one study, respectively, no consensus could be reached. Again, we would expect the conceptualization of this variable to have a large influence on this difference. For example, consensus could be used to refer to the problem, the actions for alleviating the problem, or both. If the first definition is used, agreement on a model representing the problem would already constitute consensus. If consensus on actions is referred to, the concept is close to commitment. In addition, the concept of consensus refers ‘complete agreement’ for some, while some authors would term an increase in convergence of ideas ‘consensus’ as well (Scheper, 1991). As mentioned before, one study using an objective assessment of mental model alignment shows no significant effect (Verburgh, 1997). Huz, Andersen, Richardson and Boothroyd (1997) report more alignment in perceptions of systems goals, but no significant increase in perceptions of strategies for change.

We therefore conclude that the studies collected provide some support for the influence of group model building on consensus (in only 3 out of 40 reports the result is negative). A more careful look at the exact definitions used, especially the difference between consensus on problem analysis and consensus on actions, would enable a more accurate assessment of the effect on consensus.

*System changes, results and further use of modeling*
In 33 modeling projects, changes at the system level are implemented. In three cases modeling results do not lead to changes at the system level. In two instances the model suggested changes in the reward system in the client organization, which the management was not rushing to implement (Roberts, Abrams and Weil, 1978; Akkermans, 1995 case 2). The third study is the political sensitive issue mentioned before (Akkermans, 1995 case 4). In 19 studies positive results at the systems level are reported. In 30 of 81 cases modeling continues to be used after the initial project is over.

These results should be compared against the number of studies that set out to find implementable solutions. We found 64 projects modeling ‘real world problems’, which suggests that in about half of the relevant cases, changes are implemented. More than half of these changes led to positive results. As a considerable number of reports collected here is written immediately after the project, limiting measurement to a short-term outcomes, this number might be a low estimate.

## Conclusion and discussion

As we stated in the introduction to this review, this paper represents very much work in progress. Numerous reasons for a careful interpretation of results were mentioned. However, we hope to have presented enough reasons also for concluding that a careful review of existing quantitative and quantitative studies enables us to gain insights into results of client participation in modeling. Some notable results so far were the following:

- most group model building efforts result in an increase in insight;
- a substantial number of studies leads to commitment and behavioral change;
- no support is found for the claim that group model building provides a common language;
- in the small number of studies measuring this outcome, quality of communication in general increased;
- in a number of studies an effect on consensus can be identified, but the operationalization of this concept requires careful consideration;

- group model building led to changes on a systems level in about half of the studies collected here.

In the near future we intend to add to the work presented here, both by including more cases in the database, and by looking into the possibility of gathering more data for each case. In addition, we feel that on closer examination, a number of meaningful categories of studies can be identified. We expect to be able to link these categories to the discussions in the literature, e.g. on quantitative versus qualitative, or small versus large models. Although collection and comparison of modeling studies can logically never end, we hope that some stage can be reached in which this work forms a base for formulating better research questions and thereby specifying context-mechanism-outcome configurations in ever more detail.

## Literature


Akkermans, H.A. (1992). Participative Modeling to Support Strategic Decision Making in Operations: A Case Study. *Proceedings of the 1992 International System Dynamics Conference*. Utrecht, The Netherlands, 21 – 31.

Akkermans, H.A. (1995). *Modeling with Managers. Participative Business Modeling for Effective Strategic Decision-Making*. Doctoral Dissertation. Eindhoven Technical University, Eindhoven.

Akkermans, H.A. (1995). Quantifying the Soft Issues: A Case Study in the Banking Industry. *Proceedings of the 1995 International System Dynamics Conference*. Tokyo, Japan, 313 – 322.

Akkermans, H.A., and Bosker, J. (1994). Design Guidelines for Participative Business Modeling Projects: Lessons from an Unsuccessful Case Study. *Proceedings of the 1994 International System Dynamics Conference. Problem Solving Methodologies*. Stirling, Scotland, 15 – 24.

Akkermans, H.A., and Vennix J.A.M. (1997). Clients' Opinions on Group Model Building: An Exploratory Approach. *System Dynamics Review*, 13 (1), 3 – 31.

Alanne, P.P., and Jambekar, A.B. (1996). Putting Systems Thinking to Use: A Case Study. *Proceedings of the 1996 International System Dynamics Conference*. Cambridge, Mass, 9 – 12.

Andersen, D.F., and Richardson, G.P. (1997). Scripts for Group Model Building. *System Dynamics Review*, 13 (2), 107 – 130.

Andersen, D.F., Richardson, G.P., and Vennix J.A.M. (1997). Group Model Building: Adding More Science to the Craft. *System Dynamics Review*, 13 (2), 187 – 203.

Bentham, J.B., and Visscher, A.G. de. (1994). Systems Thinking and Its Influence on Operational Culture. *Proceedings of the 1994 International System Dynamics Conference. Organisational Learning*. Stirling, Scotland, 1 – 12.

Berkvens, J.G.H.M., and Neomagus, L.J.H. (1997). *Logistiek Perspectief. Integrale Visie-Ontwikkeling over Logistieke Bereikbaarheid [The Logistic Perspective. Developing an Integral Vision on Logistic Accessibility]*. Doctoral Thesis. Catholic University Nijmegen, Nijmegen Business School, Nijmegen.

Bronkhorst, E.M., Wiersma, T., and Truin, G.J. (1991). Using Complex System Dynamics Models: An Example Concerning the Dutch Dental Health Care System. *Proceedings of the 1991 International System Dynamics Conference*. Bangkok, Thailand, 155 – 163.

Bull, M., Ford, A., Naill, G. (1985). The Importance of Feedback in the Pacific Northwest Electric Conservation Planning Model. *Proceedings of the 1985 International System Dynamics Conference*. Keystone, Colorado, 85 – 101.

Byrne, M., and Davis, A. (1991). Stella Modeling Process for a Manpower Strategy. *Proceedings of the 1991 International System Dynamics Conference*. Bangkok, Thailand, 73 – 81.

Cavaleri, S., and Sterman, J.D. (1997). Towards Evaluation of Systems Thinking Interventions: A Case Study. *System Dynamics Review*, 13 (2), 171 – 186.

Cook, Th.D., and Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis for Field Settings.* The University of Chicago Press, Chicago.

Cooper, K.G. (1980). Naval Ship Production: A Claim Settled and a Framework Build. *Interfaces*, 10 (6), 20 – 36.

Coyle, R.G. (1989). System Dynamics and Defence Analysis. *Proceedings of the 1989 International System Dynamics Conference*. Stuttgart, Germany, 599 – 607.

Davis, A. (1994). The Application of System Dynamics to Re-Engineering Career Plans. *Proceedings of the 1994 International System Dynamics Conference. Business Decision Making*. Stirling, Scotland, 31 – 39.

Deakins, E., and Winch, G. (1994). Helping Not-For-Profit Enterprises Become More "Business-Like": A Learning Organisation Approach. *Proceedings of the 1994 International System Dynamics Conference. Microworlds*. Stirling, Scotland, 22 – 30.

Delbecq, A.L., Ven, A.H. van de, and Gustafson, D.H. (1975). *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*, Scott, Foresman and Co, Glenview (Ill.).

Doyle, J. (1997). The Cognitive Psychology of Systems Thinking. *System Dynamics Review*, 13 (3), 253 – 266.

Doyle, J., Radzicki, M., and Scott Trees, W. (1998). Measuring Change in Mental Models of Dynamic Systems: An Exploratory Study, *Proceedings of the 1998 International System Dynamics Conference*. Quebeq, Canada, 30 – 31.

Draper, F., and Swanson, M. (1990). Learner-Directed Systems Education: A Successful Example. *System Dynamics Review*, 6 (2), 209 – 213.

Duman, E., Balibek, E., Firat, A., and Barlas, Y. (1997). A Dynamic Feedback Model for Strategic Management of an Insurance Company. *Proceedings of the 1997 International System Dynamics Conference*. Istanbul, Turkey, 685 – 688.

Els, S.A. (1997). Better Results through Better Communication Techniques in Consulting. *Proceedings of the 1997 International System Dynamics Conference*. Istanbul, Turkey, 571 – 578.

Eriksen, H.E., and Nielsen, E.H. (1985). A Report on the Utilization of the MOSES Computer as a Tool in Development and Dessimination of System Dynamics Models. *Proceedings of the 1985 International System Dynamics Conference*. Keystone, Colorado, 237 – 248.

Fey, W.R. (1978). An Industrial Dynamics Case Study. In Roberts, E.B. (ed.) *Managerial Applications of System Dynamics*. Productivity Press, Cambridge, Massachusetts [etc.], 117 – 138.

Fey, W., and Trimble, J. (1992). The Evaluation and Development of Knowledge Acquisition in System Dynamics Studies, *Proceedings of the 1992 International System Dynamics Conference*, Utrecht, The Netherlands, 173 – 182.

Ford, D.N., and Sterman, J.D. (1998). Expert Knowledge Eilicitation to Improve Formal and Mental Models. *System Dynamics Review*, 14 (4), 309 – 340.

Genta, P.J., Kreutzer, D.P., Anderson, G., Hinote, R., Hood, W.C., and McMillan, K. (1994). How to Use System Dynamics to Create Your Own Future: A Case Study of a Worldwide Oil and Gas Exploration Group. *Proceedings of the 1994 International System Dynamics Conference. Business Decision Making*. Stirling, Scotland, 53 – 61.
Genta, P.J., and Sokol, N. (1993). Applying a Systems Thinking Approach to Business Process Re-Engineering: A Case Study of a Canadian Oil and Gas Producer. *Proceedings of the 1993 International System Dynamics Conference*. Cancun, Mexico, 142 – 150.
Geus, A.P. de (1988). Planning as Learning. *Harvard Business Review*, March-April, 70 – 74.
Ginsberg, A., and Morecroft, J. (1995). Systems Thinking and the Case Method. *Proceedings of the 1995 International System Dynamics Conference*. Tokyo, Japan, 74 – 88.
Hall, R.I., and Menzies, W.B. (1983). A Corporate System Model of a Sports Club: Using Simulation as an Aid to Policy Making in a Crisis. *Management Science*, 29 (1), 52 – 64.
Henderson, S.M., and Wolstenholme, E.F. (1990). The Application of a Dynamic Methodology to Assess the Benefit of a Battlefield Information System. *Proceedings of the 1990 International System Dynamics Conference*. Chestnut Hill, Massachusetts, 480 – 494.
Hickson, D.C. (1978). A Banker's Appreciation of Systems Considerations. In Roberts, E.B. (ed.) *Managerial Applications of System Dynamics*. Productivity Press, Cambridge, Massachusetts [etc.], 477 – 483.
Hines, J.J., and Johnson, D.W. (1994). Launching System Dynamics. *Proceedings of the 1994 International System Dynamics Conference. Business Decision Making*. Stirling, Scotland, 83 – 93.
Hutjes, J.M., and van Buuren, J.A. (1992). *De Gevalsstudie. Strategie van Kwalitatief Onderzoek [The Case Study. Strategy of Qualitative Research]*. Boom / Open Universiteit, Meppel, The Netherlands.
Huz, S., Andersen, D.F., Richardson, G.P., and Boothroyd, R. (1995). Measuring Mental Model Alignment from a Group Model-Building Intervention, *Proceedings of the 1995 International System Dynamics Conference*. Tokyo, Japan, handout.
Huz, S., Andersen, D.F., Richardson, G.P., and Boothroyd, R. (1997). A Framework for Evaluating Systems Thinking Interventions: An Experimental Approach to Mental Health System Change. *System Dynamics Review*, 13 (2), 149 – 169.
Huz, S., Andersen, D.F., Richardson, G.P, and Boothroyd, R. (1997). Evaluating Group Model Building in Mental Health and Vocational Rehabilitation Service Delivery. *Proceedings of the 1996 International System Dynamics Conference*. Cambridge, Massachusetts, 233 – 236.
Hsiao, N., and Richardson, G.P. (1999). In Search of Theories of Dynamic Decision Making: A Literature Review. Forthcoming in *Proceedings of the 1999 International System Dynamics Conference*. Wellington, New Zealand.
Jambeker, A.B. (1990). System Dynamics Mapping to Influence Mental Models: A Case Study. *Proceedings of the 1990 International System Dynamics Conference*. Chestnut Hill, Massachusetts, 511 – 521.
Lane, D.C. (1992). Modeling as Learning: A Consultancy Methodology for Enhancing Learning in Management Teams. *European Journal of Operations Research*, 59, 64 – 84.

Lane, D.C. (1993). From Discussion to Dialogue: How an Interactive Modeling Approach Was Used with Managers to Resolve Conflict and Generate Meaning. *Proceedings of the 1993 International System Dynamics Conference*. Cancun, Mexico, 231 – 234.

Lane, D.C. (1993). The Road Not Taken: Observing a Process of Issue Selection and Model Conceptualization. *System Dynamics Review*, 9 (3), 239 – 264 (also as working paper, 1995).

Larsen, E.R., Morecroft, J.D.W., and Murphy, J. (1991). Helping Management Teams to Model: A Project in the Consumer Electronics Industry. *Proceedings of the 1991 International System Dynamics Conference*. Bangkok, Thailand, 281 – 290.

Lyneis, J.M. (1999). System Dynamics for Business Strategy: A Phased Approach. *System Dynamics Review*, 15 (1), 37 – 70.

Machuca, J.A.D. (1992). Are We Losing One of the Best Features of System Dynamics? *System Dynamics Review*, 8 (2), 175 – 177.

Mahmoud, M., and Genta, P. (1993). Microworld of an Open University: A Strategic Management Learning Laboratory. *Proceedings of the 1993 International System Dynamics Conference*. Cancun, Mexico, 318 – 325.

Mayberry, M., Hoxsey, K., McCracken, K., and Rendell, C. (1996). Using Systems Thinking and Dynamic Simulations to Reengineer Manufacturing Processes at Silicon Graphics. *Proceedings of the 1996 International System Dynamics Conference*. Cambridge, Massachusetts, 365 – 368.

Morecroft, J.D.W. (1988). System Dynamics and Microworlds for Policy Makers. *European Journal of Operations Research*, 35, 301 – 320.

Morecroft, J.D.W. (1992). Executive Knowledge, Models and Learning. *European Journal of Operations Research*, 59, 9-27.

Morecroft, J.D.W., Lane, D.C., and Viita, P.S. (1989). Modeling a Biotechnology Startup Firm, *Proceedings of the 1989 International System Dynamics Conference*. Stuttgart, Germany, 157 – 176.

Morecroft, J.D.W., Lane, D.C., and Viita, P.S. (1991). Modeling Growth Strategy in a Biotechnology Startup Firm. System Dynamics Review, 7 (2), 93 – 116.

Morecroft, J.D.W., and Heijden, K.A.J.M. van der (1992). Modeling the Oil Producers: Capturing Oil Industry Knowledge in a Behavioural Simulation Model. *European Journal of Operations Research*, 59, 102 – 122 (short version presented at ISDC 1990).

Pawson, R., and Tilley, N. (1997). *Realistic evaluation*. Sage, London.

Randers, J. (1980). *Elements of the System Dynamics Method.* MIT Press, Cambridge.

Raynolds, P.A., and Raynolds, G.H. (1992). Jog Your Right Brain (JOG): A Case Study in Knowledge Elicitation and Evaluation. *Proceedings of the 1992 International System Dynamics Conference.* Utrecht, The Netherlands, 553 – 562.

Richardson, G.P., and Andersen, D.F. (1995). Teamwork in Group Model Building. *System Dynamics Review*, 11 (2), 113 – 137 (also in Proceedings ISDC 1992: 595 – 604).

Richardson, G.P., and Senge, P.M. (1989). Corporate and Statewide Perspectives on the Liability Insurance Crisis. *Proceedings of the 1989 International System Dynamics Conference*. Stuttgart, Germany, 442 – 457.

Richardson, G.P., and Pugh, A.L. (1981). *Introduction to System Dynamics Modeling with DYNAMO*. MIT Press, Cambridge, Massachusetts.

Richmond, B. (1987). *The Strategic Forum: From Vision to Strategy to Operating Policies and Back Again*. High Performance Systems, Lyme (NH), The Netherlands.

Richmond, B. (1997). The Strategic Forum: Aligning Objectives, Strategy and Process. *System Dynamics Review*, 13(2), 131 – 148.
Rios, J.P., and Schwaninger, M. (1996). Integrative Systems Modeling: Leveraging Complementarities of Qualitative and Quantitative Methodologies. *Proceedings of the 1996 International System Dynamics Conference*. Cambridge, Massachusetts, 431 – 434.
Roberts, E.B., Abrams, D.I., and Weil, H.B. (1978). Policy Formulation in a Vertically-Integrated Firm. In Roberts, E.B. (ed.), *Managerial Applications of System Dynamics*. Productivity Press, Cambridge, Massachusetts [etc.], 427 – 446 (first published in Management Science, 14 (12), 1968).
Roberts, E.B. (1978a). Some Insights into Implementation. In Roberts, E.B. (ed.), *Managerial Applications of System Dynamics*. Productivity Press, Cambridge, Massachusetts [etc.], 155 – 161.
Roberts, E.B. (1978b). Strategies for Effective Implementation of Complex Corpor ate Models. In Roberts, E.B. (ed.) *Managerial Applications of System Dynamics*. Productivity Press, Cambridge, Massachusetts [etc.], 77 – 85.
Rohrbaugh, J.W., Andersen, D.F., Richardson, G.P., and Zagonel-Santos, A. (1997). *Group Model Building for Social Services in Cortland County*. Working paper / project report SUNY Albany (3 volumes) .
Roman Sliwa, K. (1993). Systems Thinking and Organizational Growth. Personnel Pressure and Organizational Equilibrium Scissors: A Case of the Company "Beta". *Proceedings of the 1993 International System Dynamics Conference*. Cancun, Mexico, 398 – 407.
Roos, E. (1996). System Dynamics Modeling: A Case Study from the Software Industry. *Proceedings of the 1996 International System Dynamics Conference*. Cambridge, Massachusetts, 443 – 446.
Roos, E. (1997). Group Model Building with a Client Using System Dynamics Modeling. *Proceedings of the 1997 International System Dynamics Conference*. Istanbul, Turkey, 613 – 616.
Rosenhead, J. (ed.) (1989). *Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict*. Wiley, Chichester.
Rouwette, E.A..J.A., Vennix, J.A.M., and Thijssen, C.M. (1997). Group Model Building: A Decision Room Approach. *Proceedings of the 1997 International System Dynamics Conference*. Istanbul, Turkey, 619 – 622.
Rufat-Latre, J. (1994). Strategy and Systems Thinking through Dynamic Storytelling. *Proceedings of the 1994 International System Dynamics Conference. Organisational Learning*. Stirling, Scotland, 45 – 54.
Rufat-Latre, J., Jamieson, M., and Mora, M. (1993). Transferring Systems Thinking and Circumscribing Problems: A Case Study. *Proceedings of the 1993 International System Dynamics Conference*. Cancun, Mexico, 427 – 434.
Sancar, F.H. (1987). Implementation of a Modeling Approach to Community Development Planning. *Proceedings of the 1987 International System Dynamics Conference*. Shanghai, China, 446 – 461.
Scheper, W.J. (1991). *Group Decision Support Systems. An Inquiry into Theoretical and Philosophical Issues*. Unpublished Doctoral Dissertation, Univeristy of Utrecht, Utrecht, The Netherlands.
Schlager, K.J. (1978). How Managers Use Industrial Dynamics. In Roberts, E.B. (ed.) *Managerial Applications of System Dynamics*. Productivity Press, Cambridge,

Massachusetts [etc.], 145 – 153 (first published in Instrumentation Technology, 1963).
Senge, P. (1987). Catalyzing Systems Thinking within Organizations. *Proceedings of the 1987 International System Dynamics Conference*. Shanghai, China, 855 – 887.
Swanborn, P.G. (1987). *Methoden van Sociaal-Wetenschappelijk Onderwijs* [Research Methods in the Social Sciences]. Boom, Meppel, The Netherlands.
Veit, K.P. (1978). System Dynamics in Corporate Long-Range Strategic Planning. In Roberts, E.B. (ed.), *Managerial Applications of System Dynamics*. Productivity Press, Cambridge, Massachusetts [etc.], 533 – 544 (first publication International Congress of Actuaries, 1976).
Vennix, J.A.M. (1990). *Mental Models and Computer Models. Design and Evaluation of a Computer-Based Learning Environment for Policy Making.* Unpublished Doctoral Dissertation. University of Nijmegen, Nijmegen, The Netherlands.
Vennix, J.A.M. (1995). Building Consensus in Strategic Decision Making: Insights from the Process of Group Model-Building. *Group Decision and Negotiation*, 4 (4), 335 – 355.
Vennix, J.A.M. (1996). Group Model Building: Facilitating Team Learning Using System Dynamics. John Wiley & Sons, Chichester.
Vennix, J.A.M., Akkermans, H.A., and Rouwette, E.A.J.A. (1996). Group Model Building to Facilitate Organizational Change: An Exploratory Study. *System Dynamics Review*, 12 (1), 39 – 58.
Vennix, J.A.M., Andersen, D.F., Richardson, G.P., and Rohrbaugh, J. (1992). Model-Building for Group Decision Support: Issues and Alternatives in Knowledge Elicitation. In Morecroft, J.D.W., and Sterman, J.D. (eds.) *Modelling for Learning. Special Issue of the European Journal of Operational Research*, 59 (1), 28 – 41.
Vennix, J.A.M., Gubbels, J.W., Post, D., and Poppen, H.J. (1990). A Structured Approach to Knowledge Elicitation in Conceptual Model Building. *System Dynamics Review*, 6 (2), 194 – 208.
Vennix, J.A.M., Scheper, W., and Willems, R. (1993). Group Model Building: What Does the Client Think of It? *Proceedings of the 1993 International System Dynamics Conference*. Cancun, Mexico, 534 – 543.
Verburgh, L. (1994). *Participative Policy Modeling Applied to the Health Care Insurance Industry*. Doctoral Dissertation. University of Nijmegen, Nijmegen, The Netherlands.
Wallace, S.D., and Sancar, F.H. (1988). An Integrative Approach to Water Resource Management: An Application in Middleton, Wisconsin. *Proceedings of the 1988 International System Dynamics Conference*. La Jolla, California, 447 – 459.
Watts, K.M., and Wolstenholme, E.F. (1990). The Application of a Dynamic Methodology to Assess the Benefit of a Logistics Information System in Defence. *Proceedings of the 1990 International System Dynamics Conference*. Chestnut Hill, Masssachusetts, 1286 – 1298.
Watts, K.M., and Wolstenholme, E.F. (1990). The Application of a Dynamic Methodology to Assess the Benefit of a Logistics Information System in Defence. *Proceedings of the 1990 International System Dynamics Conference*. Chestnut Hill, Massachusetts, 1286 – 1298.
Weil, H.B., and Veit, K.P. (1989). Corporate Strategic Thinking: The Role of System Dynamics. *Proceedings of the 1989 International System Dynamics Conference*. Stuttgart, Germany, 66 – 75.

Weil, H.B., and Etherton, R.L. (1990). System Dynamics in Dispute Resolution. *Proceedings of the 1990 International System Dynamics Conference*. Chestnut Hill, Massachusetts, 1311 – 1324.
White, L., Ackroyd, T., and Blakeborough, M. (1994). Learning about Modeling for Learning. *Proceedings of the 1994 Inernational System Dynamics Conference. Education.* Stirling, Scotland, 100 – 109.
Wile, K., and Smilonich, D. (1996). Using Dynamic Simulation for Resource Management Policy Design at the Minnesota Department of Transportation. *Proceedings of the 1996 International System Dynamics Conference*. Cambridge, Massachusetts, 569 – 512.
Winch, G.W. (1993). Consensus Building in the Planning Process: Benefits from a 'Hard' Modeling Approach. *System Dynamics Review*, 9 (3), 287 – 300.
Wolstenholme, E. (1992). The Definition and Application of a Stepwise Approach to Model Conceptualisation and Analysis. *European Journal of Operations Research*, 59, 123 – 136.
Wright, R.D. (1978). Growth Strategies for a Trucking Firm. In Roberts, E.B. (ed.) *Managerial Applications of System Dynamics*. Productivity Press, Cambridge, Massachusetts [etc.], 245 – 260 (first published in Sloan Management Review, 12 (1), Fall 1971, 71 – 86.
Yin, R.K. (1984). *Case Study Research: Design and Methods*. Sage, Londen.
Zazara, R., and Fisher, D. (1996). Experiences in Developing Single-Discipline and Cross-Curricular Models for Classroom Use. *Proceedings of the 1996 International System Dynamics Conference*. Cambridge, Massachusetts, 621 – 624.