

An Experimental Investigation Comparing the Effects of Case Study, Management Flight Simulator and Facilitation of these Methods on Mental Model Development in a Group Setting

Michelle Shields

Department of Psychology, North Carolina State University

3605 Mount Vernon Way, Plano, Texas, 75025 U.S.A.

Phone: 972-208-3150 Fax: 972-491-1290

michelle.shields@mindspring.com

ABSTRACT

The practice of building and exploring Systems dynamics models with groups of non-experts is still relatively new to most organizations. One primary concern in this practice is how to structure group modeling sessions so that participants' mental models of system functioning may be most effectively elicited and made more robust given a limited time frame in which to conduct group-based activities.

This paper is drawn from dissertation research conducted at a southwestern U.S. airline. The study was designed to look at the impact of model Conceptualization (via Case Study), simulation (via Management Flight Simulator) and facilitation of these processes on the elaboration and revision of individually held mental models and group dynamics. Included in the discussion is a summary of the background literature on which the study is based, the underlying theoretical model used for establishing the experimental framework, an overview of the research methodology and results.

Keywords: Case, Model Conceptualization, Simulation, Management Flight Simulator, Mental Models, Group Model Building, Facilitation, and Experimentation

THEORETICAL BACKGROUND

Development of The Mental Model Construct

Perhaps because of the “broad church” backgrounds of those working in the area of mental models (Wilson & Rutherford, 1989), few formal, widely accepted definitions of the mental model construct exist (Rouse & Morris, 1986). More often, we find that mental models are defined according to the field and context in which they are employed. While a relatively new subject of interest in the field of management (e.g. Senge, 1990b), the field of psychology has had a long history of the use of the term mental model. As early as 1943, Kenneth Craik proposed that, “we construct internal models of the environment around us that form the basis from which we reason and predict the outcome of events” (Rogers, 1992, p. 2). Craik’s view, as well as the opinions of nearly all others since, is that people create symbolic representations in their minds that “mirror” their perceptions of external events.

Forty years after Craik’s writing, the term mental models was elaborated on by cognitive psychologist Johnson-Laird (1983) who suggested that not only do people “understand the world

by constructing working models of it in their mind” (p. 10) but that they use these cognitive representations as “inference engines” which function recursively to enable the understanding of discourse and allow reasoning through the manipulation of symbols which are derived from the structure and meaning of speech (Luria, 1973). In this way we see the term mental models evolving from its position as a cognitive *product* (a representation) to a more active role in cognitive *process* (used for understanding the world).

Embracing this functional perspective, Rouse and Morris (1986) refer to mental models as specialized cognitive structures that enable a person to *describe, predict, and explain* the behavior of a given system. In their discussions, the term ‘system’ was used to describe engineering systems such as a schematic of an electronic component with a ‘concrete’ system of inputs/outputs. Making the leap from engineering to business psychology (as proposed here) requires that we apply this same thinking to an organizational system so that in talking about an organizational system, a mental model may be used for such tasks as *describing* the organizational system’s purpose and form, for *predicting* how policy changes may affect the system, and for *explaining* why the system functions as it does (Cannon-Bowers et al., 1993; Wickens, 1992).

To summarize, by attempting to describe, predict and explain how a concrete system of interacting physical components functions (Rasmussen, cited in Wilson & Rutherford, 1989) or as associated concepts in task descriptions (Schvaneveldt, et al., 1985), people manifest internal models as imagery (Johnson-Laird, 1983) reflecting both the spatial layout of the physical system and/or the more abstract human components and processes. What Rouse and Morris brought to the development of the mental model construct is the view that mental models are more than building blocks used for inference, they gave us an application-based perspective whereby mental models are also used for describing system functioning in general and for explaining *why* the system functions as it does. It is in the arena of understanding people’s view about “why” that we find the most fertile ground in research for organizational psychologists interested in mental models.

Mental Model Measurement and Reliability

In keeping with the idea that mental models are made up of different types of specialized knowledge (Rouse & Morris, 1986; and Converse and Kahler, 1992), a variety of methods may be required to elicit and capture the content of mental models (Gordon et al., 1993; Huff, 1990). Even with the use of specialized methods of mental model measurement however, Forrester (1994) and Senge (1990a) note that people have varying levels of difficulty articulating different aspects of a perceived system accurately and that different types of mental data may be articulated with varying degrees of accuracy. They identify three categories of data that decision makers possess which appear to correspond closely with the functional definition of mental models proposed by Rouse and Morris whereby mental models function to help a person describe, predict and explain their perceptions of system behavior. First, there is data about system structure and policies which are assumptions used to *describe* how variables interact with one another. According to Senge, assumptions about policies and structures can be reported with a fairly high level of consistency. However, mental models used to *explain* system behavior (changes that have happened or are happening), may be misinformed or erroneous. Lastly, mental data used to *predict* future system behavior represent detailed information encompassing *both* system description and explanation and represent intuitive solutions people give when asked to predict what will happen when structure and policy interact. As supported by Sterman’s studies

(1989) on the misperception of feedback, Senge and Forrester contend that these assumptions are *least* reliable as people consistently misjudge the dynamic behavior of how the pieces in a system will interact over time or how behavior would be altered by new policies.

Using System Dynamics in a Group Setting to Aid in Mental Model Development

There has been a growing emphasis on developing tools and processes that help decision-makers learn from System Dynamics models through the articulation and examination of their own mental models in a group context (e.g. Andersen & Richardson, 1997; Lane, 1994; Richardson & Andersen, 1995; Richmond, 1997; Vennix, et al., 1996). Building System Dynamics models with groups of people differs in its aims from previous efforts to derive expert-based models. Unlike modeling approaches where a goal is to replicate an actual system as closely as possible, the primary goal in modeling as a means to learning at the group level is not to derive a 'correct' model of the system but rather, to focus on the *process* of model building by engaging the group in a way that contributes to their understanding of the complex issues and which may lead to a new course of action to which the group feels committed (de Geus, 1988; Vennix, 1996; Vennix, et al., 1996).

Senge (1990a) and Senge and Sterman (1994) describe a recursive process of mental model development involving three stages: mapping mental models, challenging mental models to reveal inconsistencies, and improving mental models. Inherent in mapping mental models is the premise that mental models can not evolve unless they are first made explicit (de Geus, 1988; Forrester, 1975). By having group members talk about and answer questions about the variables in a system, they may draw on their own specific experiences and in the process of telling their personal stories make their mental models known to others (Bakken, et al., 1994; Eden, 1989; Narayanan & Fahey, 1990).

An articulated mental model may, as described earlier, possess varying degrees of accuracy against some objective criteria so that, in the second stage of mental model development, challenging mental models, an attempt is made to test an individual's existing mental models for validity by seeking to uncover internal contradictions, inconsistencies, or incompleteness (Senge, 1990a). The assumption being that if a person encounters a mismatch between what he expects will happen in a simulated experiment and what actually occurs, a third stage of mental model development may take place whereby an individual's mental models of a given situation and potentially, their subsequent real-life activities, may be revised in such a way as to bring these new expectations and outcomes into line (Argyris and Schön, 1978; Williams, et al., 1983). The process of revising a mental model is not usually immediate however (Senge, 1990a) despite the often hoped for, "ah-ha" experience of suddenly seeing the error in one's thinking. Rather, new conceptual perspectives may be assimilated gradually (Levitt and March, 1988) or not at all if existing models, though perhaps simpler or even erroneous, seem to function satisfactorily (Woods et al., 1994).

The Role of Social Interaction in Mental Model Development

Applying the Systems dynamics modeling framework in a group setting may enhance the group interaction process which may, in turn, affect mental model development (Lane, 1994; Morecroft and Sterman, 1994; Vennix, 1996). One reason for this is that mental models may simply be enriched the longer a person *thinks* about a topic (Morecroft, 1994; Woods et al., 1994). In other words, when group interaction occurs, people have an opportunity to recall

otherwise latent facts and concepts. The result may be a mental model that includes not only a network of 'familiar', often used concepts, but a vast matrix of *potential* connections brought to mind by the flow of conversation (Forrester, 1975). Anderson et al.'s (1992) videotapes of an 'interactive protocol' (live exchange between people working on a task) support this idea. They observed that when individuals worked with a peer, there were improvements in their mental models used for prediction, particularly if they contrasted their pre-test predictions with those of another person and entered into discussions as to possible explanations of the phenomena. Group interaction may not always improve problem solving performance however, as high status persons may dominate discussions so that participation among group members are unequal (Bakken, et al., 1994; Eden, et al., 1983; Hodgson, 1994; Vennix and Gubbels, 1994). Ensuring equal participation and avoiding deference to the most senior person in the group may necessitate a neutral third party that can inquire into the meaning of statements and ensure that all group members' contributions are heard.

The Role of Facilitator in Fostering Group Interaction

A review of the primary literature on Group Model Building (Akkermans & Vennix, 1997; Morecroft, 1994; Richardson & Andersen, 1995; Senge, et al., 1994; Vennix et al., 1994) suggests that facilitation of the process is vital for three reasons. One, the facilitator may affect the level of debate that occurs in the group. Two, she may mediate the power relationships that emerge in a group interaction and ensure that the group doesn't narrow their focus to a few approaches to the problem too soon. Third, she may have a positive effect on communication that transfers to other forms of group effectiveness.

First, whether or not alignment of mental models between two or more individuals occurs may be a function of how effective the facilitator is in fostering group interaction, as "the effectiveness of the learning cycle depends on...the skill which knowledge is elicited and...options and consequences are debated" (Morecroft, p. 11). A facilitator may encourage people to scrutinize and justify their reasoning regarding the viability of an idea while helping to side-step argumentation. When a Systems dynamics model is used to stimulate debate, the model may be seen as a unique point of view, which serves, as a focus rather than any one individual group member's point of view (Eden, 1989; Morecroft, 1984).

Secondly, a facilitator may mediate the power relationships that arise in a group. If a model is embedded in a computer-based simulation, the facilitator can establish a context in which the simulation is viewed as a tool for learning that goes beyond a position as an infallible 'black box' to occupy a more modest position as an inanimate *generator of opinions* that decision makers can and should challenge (Morecroft, 1994). Further, whether or not they intend to do so, the facilitator can effect the power relationships present in the team through the selection of what they incorporate in the model and how (Doyle, et al., 1996; Eden, 1994; Eden, et al., 1983).

Finally, effective facilitation may have an impact on the level and type of communication that occurs during the model building process. This in turn may have an effect on: the degree of group consensus, ownership of the model, and commitment to the recommendations that result (Akkermans & Vennix, 1997). First, Akkermans and Vennix found that communication during the process was positively related to forming consensus around the nature of an issue. In five of the six cases studied, good communication coincided with fair to high levels of resulting group consensus. Further, they observed that when people communicated openly and effectively, the group developed a greater feeling of ownership over the resulting model. This, they assert, may

contribute to a higher level of commitment to recommendations for policy changes. Reciprocally, this higher level of commitment may function as an indicator of greater levels of group dynamics generated through the process (Graham, et al., 1994).

Summary and Implications of the Theoretical Literature Review

An underlying assumption in the field of System Dynamics is that ‘Systems Thinking’ can be taught, that is, that people can be made aware of connections and feedback between structure and behavior. If so, then mental models about how the parts of a system interact may be revised so that they become more coherent (Lane, 1994). Unfortunately, possessing a mental model that is more coherent does not necessarily mean that it will be sufficient to handle the increased cognitive load that results when trying to predict what will happen when several events and processes interact. Studies have shown that the human brain is not capable of simulating interactions beyond a minimal number of variables. Overcoming this limitation requires mechanisms for freeing up cognitive resources of attention by providing a framework for organizing the data so that it can be stored as ‘chunks’ of information, which may be easier to process.

From a research point of view, we might assert that the more a method *challenges* people’s thinking, the greater the chances that their mental models will be altered. Yet challenging someone’s thinking is not to be undertaken lightly as it may generate a host of undesirable defensive behaviors. To elicit and confront inconsistencies in thinking requires a safe forum in which group members perceive risk taking as acceptable and an objective, neutral third party to encourage confrontation of existing beliefs, insure that less popular opinions are heard, and diffuse potentially negative interpersonal dynamics.

EMPIRICAL BACKGROUND

Ideally members of a decision-making team utilizing System Dynamics in policy making should be involved in all major aspects of the process. However, quite often there is insufficient time or opportunity for team members to meet on an on-going basis to undergo all aspects of model development and testing. Working within a constricted time-frame has led to an increased interest in finding ways to benefit from System Dynamics group modeling by utilizing sub-components from the overall model building process (e.g. Eden, 1994) and early experiments have highlighted the potential for enhancing mental models through the independent use of either Conceptual Model Building (e.g. Coyle & Alexander, 1997; Hodgson, 1994; Rosenhead, 1989; Wolstenholme & Coyle, 1983) or Simulation Experiments (e.g. Cavaleri & Thompson, 1996; Sterman and Senge, 1994).

Empirical Support for Abbreviating the Model Building Process

What follows is a review of studies that have used either Conceptual Modeling or Simulation as stand-alone interventions to teach decision-makers about complex system behavior. The research articles described here are distinguished along three lines: 1) those studies that report on efforts utilizing only Conceptualization or the creation of a “Cognitive Map” (Eden, 1989) that depicts the system under study by representing the meaning of a concept through its relationships to other concepts; 2) intervention studies that involve only simulation via a “Management Flight Simulator” (Morecroft & Sterman, 1994); and 3) those studies that compare the Case Method (a Conceptual Modeling task) to Gaming (a Simulation task) in a business education setting.

Evaluating Conceptual Modeling Used for Group Model Building

In the last few years, preliminary studies that support the use of qualitative or Conceptual Modeling for bringing about organizational learning have begun to emerge in the System Dynamics literature (e.g. Huz, et al., 1996; Vennix, et al., 1996; Wolstenholme, 1994). In a study of the effectiveness of Group Model Building techniques, Vennix, et al. (1996), reviewed cases involving Conceptual Modeling (the creation of a System Dynamics causal map) in order to evaluate whether it could induce, in a time efficient manner, the kind of change in management attitude and behavior considered necessary for organizational success. Their findings were in keeping with those of Huz et al. (1996) who observed that following Conceptual Model Building, participants were in greater alignment with regard to the goals of the organizational system but demonstrated no increase in alignment regarding *strategies* for change. That is to say, they developed more agreement on what the problem *was* but no further agreement on what to *do* about it. Even though participants in the Vennix et al. study said they gained considerable insight into the problem by revealing relationships and feedback processes between problem elements, they did not feel that their initial *opinions* had changed much. This suggests that no change in mental model occurred, or that if it did, that participants were able to distinguish this change as analysis of the individuals' workbooks indicated that the number of variables identified increased, concepts became more detailed, and new relationships were added following the intervention.

Evaluating Simulation Experiments Used for Group Model Building

Although using Management Flight Simulators in management development has generated controversy (Bakken, et al., 1994), recent evaluation studies (e.g. Akkermans & Vennix, 1997; Bakken, et al., Doyle, et al., 1996) support their use for enhancing mental models. In a study by Doyle et al. in which half the subjects were allowed to play a simulation game designed to coincide with data demonstrating a particular organizational dynamic, some participants showed different content in mental models, an outcome they attributed to the use of the Simulation. The researchers noted that member's new mental models did not *replace* their original views but rather, new concepts were integrated into old as demonstrated by the addition of variables to the model. In essence, it may be said that the computer provided an additional "voice" in the ongoing conversation about the system that stimulated a greater array of 'possible' cognitive associations.

However, if we take a hard line in defining organizational learning as a process that calls for not only changes in *thinking* but changes in organizational *behavior* as well (e.g. Argyris and Schön, 1996), then comparing the methods from this perspective may provide a stricter, but more refined basis for evaluation. Of the studies reviewed, in some cases, Simulation is reported to have led to learning at the organizational or policy level. In experiments in which only a Management Flight Simulator was used, Bakken, et al. (1994) report that when there was a difference in the mental models of two participants, the Management Flight Simulator provided a framework for discussion which ultimately led to a reconciliation of the two individual opposing views and instigated changes in one participant's organizational incentive policy. In another Case Study, Senge and Sterman (1994) cite Bergin and Rusko (1990) who quote a participant as saying, "Before the lab, I would have said the lack of quality was the only important factor. After the lab, it was obvious to me that productivity was also a key issue [s]o I restructured some units to enhance their ability to settle claims." (p. 35). In these cases, cognitive changes were followed

by changes in behavior on an organizational level such that organizational learning may be said to have occurred.

Finally, Cavaleri and Thompson (1996) suggests that there may be specific factors such as the backgrounds of the users that influences the extent to which benefits may be derived from simulation use. In a questionnaire administered following the use of a computer simulation by four groups of business students and managers, Cavaleri and Thompson observed that managers, more than students, reported the microworld to be useful and effective for deepening their understanding of management practice.

Widening the Scope: Experimental Studies Comparing the Case Method to Gaming

Business schools have long utilized the Case Study method and Simulation or Gaming methods as experiential approaches to teaching business strategy (Wolfe, 1976). While Management Simulations differ from Games in terms of the degree to which individuals give input and make decisions, many experts view Simulations and Games as synonymous (e.g. Lane, 1995). Likewise, the Case Study method, which is characterized by an analysis of those variables considered most critical to the problem, is similar to the Conceptual Modeling component of the System Dynamics methodology in that they both encourage decision makers to think strategically, view the business as a whole, and adopt the perspective of the general manager (Graham, et al., 1994).

Research by Moore (1967), Strother (1966, cited in Wolfe, 1975b), Wolfe (1973; 1975a; 1975b) and Wolfe & Guth (1975) compared the relative contributions of the Case and Gaming methods to students' understanding of business issues. Overall, results of these studies have been contradictory. In one study, Moore (1967) found that games utilized in teaching business policy were *not* superior to traditional methods in teaching Production Management, while Wolfe (1973) and Wolfe and Guth (1975) found that the use of games in teaching business policy *were* superior to those of case teaching - but only if teacher guidance and structure (i.e. facilitation of the process) were provided.

Wolfe's (1973; 1975a) research supports the importance of effective communication in student performance in simulated business environments. One study (Wolfe, 1975a) looked at effective performance of business students who utilized a simulated policy and decision-making environment. Among the behaviors associated with successful group performance were: a) formulating a long run strategy or plan; b) talking with other individuals during play; c) quantifying statements and rationalizing techniques; d) taking ample time for discussion among team members; d) taking an experimental and questioning attitude; and e) demonstrating flexibility in the face of changing conditions.

In a second study, Wolfe (1975b) pointed out the importance of facilitation in the learning process when he compared a traditional teaching approach (first structuring and then leading the learning process) to an experiential teaching approach in which the instructor role was largely passive once the initial learning structure was established. In a comparison of the amount and type of knowledge acquired by each group using a six-question test before and after the play, Wolfe found no gain in test scores in the learning environment in which there was no facilitation of the learning process; while in the facilitated learning process using the traditional approach, a gain in overall knowledge and principle mastery was observed.

Finally, Strother et al. (cited in Wolfe, 1975b) observed that students who utilized Gaming seemed to inconsistently apply decision-making techniques ad hoc. He asserts that students in a

gaming situation are often aware of issues or problems during play but fail to apply formal and rational analyses needed to solve them. He further notes that participants become so involved in play that they do not take time to objectively understand what they are doing. Many of these problems, says Wolfe, could be handled through facilitation of the process.

Summary and Implications of the Empirical Review

While studies have been few in number, outcomes of the evaluations of Conceptual Model Building and the use of Management Flight Simulators for enhancing mental models provide guiding principles in formulating hypotheses and designing a research methodology for the study and described in this paper. First, the studies reported on above support the theoretical assertion that both Conceptualization and Simulation can be effective in making mental models more explicit and both have been shown to help participants develop a deeper understanding of their organizations. However, based on the studies reviewed here in which modeling sessions resulted in specific policy changes in the organization, it might appear that Simulation holds greater potential for meeting a more stringent evaluation criteria. Finally, with either approach, studies involving the comparison of traditional Case Study method to experimental Gaming techniques support the anecdotal evidence that some level of facilitation of the group's process may be critical to learning.

A RESEARCH DESIGN FOR EVALUATING TWO ABBREVIATED GROUP MODEL BUILDING METHODOLOGIES

Operationalizing the Terminology Used in the Study

Accepting mental model development as a meaningful indicator of the effectiveness of a group modeling technique requires that the construct be operationalized in a way that general theories about its development can be assessed. As noted earlier, mental models may take many forms and represent different kinds of knowledge structures. For purposes of this study, the term *mental model* relates specifically to *strategic* mental models (a combination of declarative and procedural knowledge developed over time and adapted to specific contexts). The strategic mental models of interest here are those dealing with a *complex organizational system*, that is, a bounded system of both structural and social components that reflects the flow of information, products and people within a business context. So that a mental model of a complex system here refers to the individual's articulation of what these components are, how they are interconnected, and what happens to them over time. Next, *facilitation* is operationalized as a process whereby individuals are asked to: *describe* their assumptions about interactions in the model making them known to others in a group; *predict* what will happen when a chosen strategy is implemented; and *explain* results of outcomes following feedback on performance. Finally, *group dynamics* refers to the relative levels of group debate, reasoning, and strategy articulation present during a group decision-making activity.

A Theoretical Model of Mental Model Development from Group Model Building

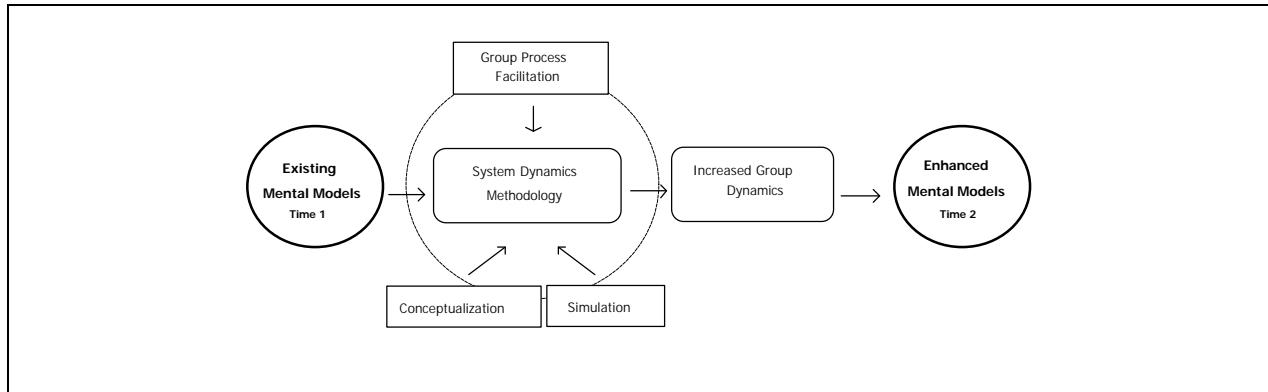


Figure 1

In the theoretical model in Figure 1 it is proposed that through the application of the System Dynamics Group Model Building framework comprised of Conceptualization, Simulation and Group Facilitation techniques that individuals will: make their existing (Time1) mental models of system structure and behavior more explicit; have opportunities to test their assumptions; and revise their thinking if and when it is shown that their mental models are in some way erroneous. The application of this framework is believed to be effective because it increases the relative level of group dynamics. This group interaction serves as the catalyst for safely challenging existing mental models so that they are measurably enhanced at Time2. An enhanced mental model is one which possess any of the following: a greater number of variables identified by the group member as important to controlling the system; an increase in the number of connections between system components articulated; a greater level of detail in strategies proposed; and/or inclusion of characteristics that may be considered properties of the “whole system”.

Hypotheses

Based the above theoretical model, which is based on the preceding theoretical and empirical studies review, the following outcomes are asserted:

- Ho₁: The method used will have a significant effect
- Ho₂: Facilitation of the group process will have a significant effect
- Ho₃: Time will have a significant effect
- Ho₄: There will be significant interaction between Method x Time effect
- Ho₅: There will be a significant interaction between Facilitation x Time effect
- Ho₆: There will be a significant interaction between Method x Facilitation x Time

RESEARCH METHODOLOGY

The main interest in this study is how an individual’s mental model of a complex organizational system changes as a result of participating in one of two methods for learning about complex systems one, utilizing the Case Study method and the other, a Management Flight Simulator. The remainder of this paper reports on a study that incorporates repeated measures

and multiple methods to assess the enhancement of mental models. In this study which was conducted at a major airline located in the United States, there were two experiments in the overall research design, one to assess mental model development and one to measure the level of group dynamics present during the Group Model Building activity. Due to the space limitations for this paper only the first experiment used to assess mental model development will be reported here.

Assessing Mental Model Development

The aim of Experiment I was to test hypotheses related to two modeling strategies and two levels of facilitation on mental model development. In the 2x2x2 repeated measure design shown in Figure 2, two levels of modeling strategies - *Conceptualization and Simulation*, and two levels of facilitation - *Scripted Facilitation and Non-Facilitation* made up the four experimental groups. Under this repeated measures design, each experimental group served as its own control group.

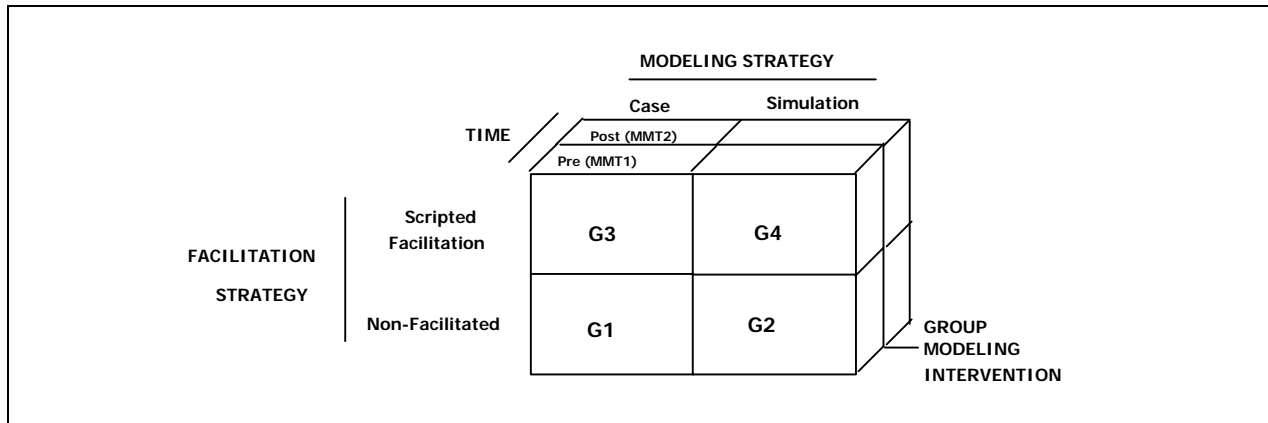


Figure 2

Dependent Variables

Three measurement methods were used to assess the extent of mental model development: 1) an Open-Ended Question; 2) a relatedness Ratings Task with two parts and; 3) a diagramming task with two scored components making a total of five dependent variables. Each of these measurement approaches was selected as a way to tap into unique but overlapping aspects of participants' mental models. The Open-Ended Question was used to gather strategic knowledge about a general strategy for controlling the system. The two Ratings Tasks were used to assess participants' perceptions about the impact that each control variables had on the system overall (Task I) and their views on how the variables effect specific populations of people within the system (Task II). Lastly, the Diagramming task was designed to capture individuals' mental models about system complexity (Measure 1) and System Dynamics (Measure 2).

Scoring of the Dependent Variables

- **Open-Ended Complexity:** An open-ended question about the best strategy for achieving the desired system change. Scoring: Objective Weighted Average Score assessing the presence of multiple strategies, variations in levels of strategies applied, changes in strategy over time,

effects of strategies on specific populations of people and the presence of potential feedbacks, flows and delays within the system described in the text.

- **Rating Task I:** Rating the six control variables in the Case/Simulation according to their impact on the system. Scoring: Comparison to Expert Criterion Score to derive a differential delta between the Subject Expert's and the Participant's ratings. The lower the score, the closer the responses were to the modeler's.
- **Rating Task II:** Rating the six control variables in the case/simulation according to the impact each has on the three specific populations of people in the hypothetical organizational system. Scoring: Comparison to Expert Criterion Score to derive a differential delta.
- **Diagramming:** Placing pre-labeled elements of the system (control variables and affected populations of people) in relation to one another on a blank page and drawing connecting line and arrows as well as relevant System Dynamics language such as delays and indicators of direction of change/flow in the system. Scoring:
 - (**Measure 1: Complexity**): A weighted cumulative score of the number of variables and connections used in the diagram.
 - (**Measure 2: System Dynamics**): A cumulative score of the use of same/opposite directional flow indicators, flows between populations of people in the system, feedback between variables and delays of effect over time.

Sample Description

Management and professional employees at the airline were invited to attend a half-day Change Management seminar designed to introduce System Thinking concepts. Participants who responded self-selected into one of the four training sessions offered (dates and times were published but specific training procedures to be used i.e. Case or Simulation, were not). The sample made up of 58 participants assigned to four experimental conditions were distributed as follows: **Group 1: Non-Facilitated Case** n=9; **Group 2: Non-Facilitated Simulation** n=14; **Group 3: Facilitated Case** n=18; **Group 4: Facilitated Simulation** n=17. *Total Non-Facilitation* n=23; *Total Facilitation* n=35; *Total Case* n=27; *Total Simulation* n=31.

Of the 58 participants, 88% had no prior exposure to System Dynamics; none had prior experience using a Management Flight Simulator; 88% of the 58 had no prior exposure to Systems Thinking and 98% of the sample had no prior exposure to concepts related to "The Tipping Point" (Shapiro, 1998), the Management Flight Simulator used in the study. Further, of the sample, 79% had no prior Engineering experience and 63% had no prior experience working in the area of Organizational Effectiveness. The types of jobs held by participants included: Finance and Accounting (12%), Customer Service (9%), Training (7%), Management/Supervision (22%), Human Resource Development (16%), Engineering (14%), Project Management (7%), Analyst (10%), and Other (3%). Nearly half of the sample, (48%), were college graduates. Of the remainder, 4% were high school graduates, 31% had some college and 17% had possessed a Master's degree or beyond.

Experimental Protocol and Review of the Dependent Variables

Each experimental session began with a brief overview on Organizational Models of Change and an introduction to Systems Thinking which included an orientation in how to label a causal-loop diagram using arrows, same/opposite or +/- labels, system feedbacks and delays. Depending on whether a participant was in the Case Study or Simulation group, they either read a case

written for the workshop or were given an orientation to the simulation control panel of a newly developed Management Flight Simulator called The Tipping Point (Shapiro, 1998). The Simulation activity orientation, as with the Case Study orientation, included an explanation of each of the variables participants could use to control the system and an overview of the objectives in the hypothetical organization. Following a pre-set time provided so that they could become familiar with Case/Simulation, participants completed the pre-test, which included the five Dependent Variable tasks described above.

Group Activity Protocol

After the “pre-test”, the Group Model Building activity was described and participants were randomly assigned to small working groups made up of 4-5 members each. All groups were given the same assignment: to come up with a group strategy for implementing a change in a fictitious organizational system utilizing the six variables described in the Case or the Simulation. Those participants in the Case conditions worked in small groups to conduct an analysis of the case in order to establish a strategy to control the hypothetical organizational system. Their strategy was to include the necessary levels of each variable they would use to achieve their objectives and a prediction of what they thought would happen over time if they employed their strategy.

Depending on which Case condition they were in (Facilitated or Non-Facilitated), variations in assistance with group processes for completing these objectives were provided. If they were in the Non-Facilitated Case condition the group was instructed to use any group process techniques they chose to develop an overall strategy. If they were in the Facilitated Case condition they were assigned a trained facilitator (one of eight volunteers from a local Organizational Development professional association). The facilitator was instructed to guide the group through the case analysis with the aid of a pre-established script which included a description of his or her role and prompting questions to use such as, “what are the key relationships in the system?” (describe); “what makes these important?” (explain) and, “what will happen if you implement your proposed strategy”? (predict). With the exception of being given assistance in how to operate the Management Flight Simulator, the Non-Facilitated Simulation groups were given no further guidance. Those in the Facilitated Simulation groups were led through a scripted process that included the same questions utilized in the Facilitated Case groups. Following the fifty-minute group activity, participants repeated the Open-Ended Question, Ratings, and Diagramming tasks completed earlier on the pre-test.

RESULTS

In order to differentiate between pre-post differences in an intervention with multiple variables, the generalized multivariate analysis of variance (MANOVA) technique recommended by Cook & Campbell (1979) was used. In this approach, no one group in the study was untreated and pre-test scores served as the point of comparison for post-test scores. The purpose of this analysis is not to prove directly that there are *differences* among groups but rather that they are *not the same* (Kosecoff & Fink, 1982). The unit of analysis for this part of the study was individual and based on the use of averages and standard deviations on pre- and post-assessment scores. Further, the aim of the MANOVA is to address whether the level of one independent

variable alters the influence of another and is particularly important where there are multiple dependent variables which are correlated both theoretically and statistically (Weinfurt, 1998).

Within Group Means

Group Means and standard deviations for the pre- and post-test scores on the five dependent variables are compared in Table I below.

Table I

Overall Means and Standard Deviations for Pre- and Post-test Scores for a Systems Thinking Group Model Building Activity Using Five Dependent Variables

Variable	Pretest			Post-test		
	<u>M</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>n</u>
Open-Ended Complexity question - Complexity	.69	.42	58	.85	.42	58
Ratings Task I*	3.17	1.48	58	2.68	1.69	57
Ratings Task II*	15.0	3.01	57	16.17	3.77	54
Diagramming Task - Complexity	17.40	8.42	58	17.91	7.64	56
Diagramming Task - System Dynamics	3.24	2.16	58	3.34	2.14	56

* Lower Scores indicate responses that are closer to the expert modeler's ratings i.e. "less is better"

Overall, scores improved from T1 to T2 on four of the five dependent variables. Across all groups, the largest change from pre-to-post test was observed in the Open-Ended Complexity question (Mean Time1 = .69, MT2 = .85, %Change = +23%) followed by the Ratings Task I (MT1 = 3.17, MT2 = 2.68, %Change = +16%). Interestingly, the second Ratings task in which individuals were asked to indicate the level of influence that each of the six variables would have on the three populations of people in the system, yielded performance decrements in the post-test across all participants (MT1 = 15.0, MT2 = 16.17, %Change = -7%). Finally, across all participants, increased scores were observed evenly on the two components of the Diagramming Task (System Complexity: MT1 = 17.40, MT2 = 17.91, %Change = 3%; System Dynamics: MT1 = 3.24, MT2 = 3.34, %Change = 3%).

Between Group Means

Group Means and Standard Deviations for the pre- and post-test scores for the two levels of Facilitation and the two Methods used in the study are presented below in Table II and the Differences between these results are then compared in Table III.

Table II

Means and Standard Deviations for Two Levels of Facilitation Used Across Two Methods of Systems Thinking Group Model Building Activities and Five Dependent Variables Used to Measure Mental Model Development

Variable	<u>FACILITATION</u>							
	Level 1 Non-Facilitated				Level 2 Facilitated			
	Pretest		Post-Test		Pre-test		Post-Test	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Open-Ended Complex.	.55	.34.	.70	.30	.79	.45	.94	.47
Ratings Task I*	4.22	1.0	3.22	1.68	2.49	1.34	2.32	1.63
Ratings Task II*	15.39	2.50	17.10	3.33	14.74	3.32	15.58	3.95
Diagram. - Complex.	19.13	7.79	18.89	8.39	16.26	8.73	17.23	7.13
Diagram. - Sys. Dyn.	3.48	1.90	4.04	2.18	3.09	2.33	2.85	2.00

Variable	<u>METHODS</u>							
	Level 1 Case				Level 2 Simulation			
	Pretest		Post-Test		Pre-test		Post-Test	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Open-Ended Complex.	.76	.42	.89	.42	.64	.43	.82	.43
Ratings Task I*	3.48	1.55	3.00	1.74	2.90	1.38	2.42	1.63
Ratings Task II*	14.74	3.75	15.75	3.66	15.23	2.19	16.50	3.88
Diagram. - Complex.	18.19	9.27	17.80	8.63	16.71	7.69	18.00	6.90
Diagram. - Sys. Dyn.	3.15	2.23	3.24	2.20	3.32	2.14	3.42	2.13

* Lower Scores indicate responses that are closer to the expert modeler's ratings i.e. "less is better"

Table III

Comparisons of the Mean Scores for Facilitation

<u>Dependent Variable</u>	<u>FACILITATION</u>				
	Level 1 Non-Facilitated		Level 2 Facilitated		<u>≥ Δ</u>
	<u>T2 - T1</u>	<u>T2 - T1</u>	<u>T2 - T1</u>	<u>T2 - T1</u>	
<u>M Diff</u>	<u>% Δ</u>	<u>M Diff</u>	<u>% Δ</u>		
Open-Ended Complexity	.15	27%	.15	19%	Non-Fac
Ratings Task I	1.0	24%	.17	7%	Non-Fac
Ratings Task II	1.7	-11%	.84	-6%	Non-Fac
Diagramming Task – Complex.	.24	-1%	.97	6%	Facilitated
Diagramming Task – Syst. Dyn.	.56	16%	.24	-8%	Non-Fac

<u>Dependent Variable</u>	<u>METHOD</u>				
	Level 1 Case		Level 2 Simulation		<u>≥ Δ</u>
	<u>T2 - T1</u>	<u>T2 - T1</u>	<u>T2 - T1</u>	<u>T2 - T1</u>	
<u>M Diff</u>	<u>% Δ</u>	<u>M Diff</u>	<u>% Δ</u>		
Open-Ended Complexity	.13	17%	.18	28%	Simulation
Ratings Task I	.48	14%	.48	17%	Simulation
Ratings Task II	1.01	-7%	1.27	-8%	Simulation
Diagramming Task – Complex.	.39	-2%	1.29	8%	Simulation
Diagramming Task – Syst. Dyn.	.09	3%	.10	3%	Case/Simul.

Overall, the Non-Facilitated and Simulation conditions appear to have more favorable results in terms of changes in performance from Time1 to Time2. In looking at the differences between the Non-Facilitated and Facilitated conditions, the Non-Facilitated conditions appear to have the greatest impact on the Open-Ended Complexity question ($\underline{MT2-MT1}=.15$; $\% \Delta=27\%$) followed by the first Ratings task ($\underline{MT2-MT1}=1.0$; $\% \Delta=24\%$). As with the Facilitation comparison, in looking at the differences between the Methods, the Simulation method appears to have greatest impact on the Open-Ended Complexity question ($\underline{MT2-MT1}=.18$; $\% \Delta=28\%$) followed by the first Ratings task ($\underline{MT2-MT1}=.48$; $\% \Delta=17\%$).

Across Group Means

Means and Standard Deviations for each of the dependent variables are given in Table IV.

Table IV

Means and Standard Deviations for Four Conditions of Systems Thinking Training as Measured by Five Dependent Variables

Variable	Group 1 <u>Non-Facilitated Case</u>						Group 2 <u>Non-Facilitated Simulation</u>					
	Pretest			Posttest			Pretest			Posttest		
	<u>M</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>n</u>
Open-Ended Q. - Complexity	.63	.42	9	.81	.34	9	.50	.29	14	.64	.27	14
Ratings Task I* - Macro Influe.	4.78	.44	9	4.56	1.42	9	3.86	1.10	14	2.36	1.22	14
Ratings Task II* - Micro Influe.	15.67	3.0	9	16.13	2.48	8	15.21	2.23	14	17.69	3.73	13
Diagramming - Complexity	20.11	7.66	9	<u>20.50</u>	9.51	9	18.50	8.10	14	17.86	7.78	14
Diagramming - System Dyn.	3.11	1.96	9	4.11	2.42	9	3.71	1.90	14	4.00	2.11	14

Variable	Group 3 <u>Facilitated Case</u>						Group 4 <u>Facilitated Simulation</u>					
	Pretest			Posttest			Pretest			Posttest		
	<u>M</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>n</u>
Open-Ended Q. - Complexity	.82	.41	18	.92	.46	18	.76	.49	17	.97	.49	17
Ratings Task I* - Macro Influe.	2.83	1.51	18	<u>2.18</u>	1.29	17	2.12	1.05	17	2.47	1.94	17
Ratings Task II* - Micro Influe.	14.28	4.07	18	<u>15.56</u>	4.20	16	15.25	2.24	16	15.59	3.84	16
Diagramming - Complexity	17.22	10.05	18	16.28	8.00	16	15.24	7.24	17	18.12	6.32	17
Diagramming - System Dyn.	3.17	2.41	18	2.75	1.98	16	3.00	2.32	17	2.94	2.08	17

* Lower scores indicate ratings that are closer to criteria i.e. lower is better

Bold = Best Performance on Pretest

Underline = Best Performance on Post-test

Bold/Underline = Greatest %Change T2-T1 for Variable

Overall, for all four groups, performance improved from Time1 to Time2 on the Open-Ended Complexity question in which individuals were asked to describe the strategy they would employ to implement the change using the six variables provided. The participants' answers were rated in a weighted fashion according to whether they gave: multiple strategies; variations in the levels of use of the strategies; suggested changes in strategy over time; and whether they indicated effects of the strategy on a specific population of people within the system. Secondly, overall, performance fell from T1 to T2 on the Ratings II task. Comparing the groups on pre-test scores, no one group had the best performance on pre-test across all dependent variables and the best performance was evenly distributed across all four groups with the exceptions of the Facilitated Case which performed best on two of the five dependent variables at T1. Similarly, the best performance at T2 was fairly evenly distributed across groups for each of the five dependent variables. However, the Non-Facilitated Case condition performed best on both aspect of the Diagramming task,; the Facilitated Case group performed best on the two Ratings tasks; and the Facilitated Simulation group performed best on the Open-Ended Complexity question. Of interest here, is that in looking at the percentage change in performance from Time1 to Time2 ($(MT2 - MT1)/T1$) the greatest gains for the Open-Ended Complexity question were observed in the Non-Facilitated Case condition ($\% \Delta = 29\%$); the greatest percentage change in the first Ratings task was in the Non-Facilitated Simulation condition ($\% \Delta = 39\%$); the greatest gain in the Complexity of the Diagramming task was in the Facilitated Simulation condition ($\% \Delta = 19\%$); while for the System Dynamics component of the Diagramming task, the greatest gain was observed in the Non-Facilitated Case ($\% \Delta = 32\%$). While performance dropped for all conditions from T1 to T2 on the 2nd Ratings task, the smallest drop in performance was observed in the Facilitated Simulation condition ($\% \Delta = -2\%$). Overall, the most dramatic changes occurred in the Non-Facilitated Case condition on the System Dynamics component of the Diagramming task ($\% \Delta = 32\%$) and the Open-Ended Complexity question (29%).

Summary Results Based on Group Means

Based on the Mean Scores across groups presented in Table IV it may be said that no one group shows a clear performance gain from Time1 to Time2 across all dependent variables. Of the four variables that exhibited performance gains, greatest impact was found in the Non-Facilitated Case (2 of 5 Variables) with greatest percentage change on at least one variable in each of the Non-Facilitated and Facilitated Simulation groups. Interestingly, while the greatest scores on the post-test for two of the five dependent variables were observed in the Facilitated Case condition, this group did not show the greatest gains in performance over time. All in all, these distributed results suggest that no one task or dependent variable is a clear indicator of change in mental models as a result of the Systems Thinking group activity.

While gain scores provide a general feel for the results overall and the trends to be found within these results, they do not tell us much about the statistical robustness of the results and miss subtle distinctions within and across the Dependent Variables. When several dependent variables are used, while tapping into different aspects of mental models, they are often correlated with one another. Therefore, it is necessary to look at which variables are significant to results overall and whether when taken together, the dependent variables define one or more theoretical constructs such that the differences due to Facilitation or Method may only be observed when they are taken together as a whole system. To begin to understand the statistical robustness of the differences in Mean scores, further statistical analysis was conducted.

Factorial Analysis of Variance

In order to determine whether any of the Group Means and gains in performance over time observed on the five dependent variables for the various groups were statistically significant, an Analysis of Variance (ANOVA) was conducted. The purpose of the ANOVA is to determine whether the means of the dependent variables for each level of an independent variable are significantly different from each other. Results of the Factorial Analysis of Variance in which the effects of the Independent Variables on each of the five dependent variables are presented in Table V.

Table V

ANOVA of the 2 x 2 x 2 Repeated Measure Analysis

(Type III SS, N=116 [58x2], df=61, 54)								
<u>Source</u>	<u>Facilitation</u>		<u>Method</u>		<u>Time</u>			
	<u>F</u>	<u>Pr > F</u>	<u>F</u>	<u>Pr > F</u>	<u>F</u>	<u>Pr > F</u>		
Open-Ended Complexity	9.28	.0036***	1.23	.2727	4.63	.0358**		
Ratings Task I	46.52	.0001***	14.02	.0004***	6.00	.0176**		
Ratings Task II	1.44	.2355	1.23	.2733	3.51	.0670*		
Diagramming - Complexity	5.36	.0246**	1.47	.2311	.31	.5802		
Diagramming - Syst. Dynamics	6.45	.0141***	.15	.6961	.51	.4802		

<u>Source</u>	<u>Facilitation x Method</u>		<u>Facilitation x Time</u>		<u>Method x Time</u>		<u>Facilitation x Method x Time</u>	
	<u>F</u>	<u>Pr > F</u>	<u>F</u>	<u>Pr > F</u>	<u>F</u>	<u>Pr > F</u>	<u>F</u>	<u>Pr > F</u>
	Open-Ended Complex.	1.02	.3167	.00	.9944	.04	.8417	.24
Ratings Task I	10.24	.0023***	1.96	.1678	.01	.9103	7.52	.0083***
Ratings Task II	.23	.6323	.01	.9281	.24	.6240	3.73	.0593**
Diagramming - Complex.	.73	.3970	.46	.4993	.29	.5952	1.08	.3044
Diagramming - SD	.18	.6725	2.06	.1571	.11	.7410	.74	.3936

* = p < .10 ** = p < .05 *** = p < .01

Discussion

Main Effects: For Facilitation, all but the second ratings task was significant at the $p < .05$. For Methods, only Ratings Task I was significant ($p < .01$). Time was significant on: Open-Ended Complexity ($p < .05$), and Ratings Tasks I ($p < .05$) and Ratings Task II ($p < .10$).

Interaction Effects: Significant interactions were observed on FacilitationxMethod (FxM) for Ratings Task I ($p < .01$) and FacilitationxMethodxTime (FxMxT) for Ratings Task I ($p < .01$) and Ratings Task II ($p < .10$). Given the significance of three of the dependent variables on Time and the interactions of MxFxT we can not assume that the significance observed on Method or Facilitation can be interpreted on their own. Of most concern for further analysis therefore are Time and the interactions of FxM and FxMxT for Open-Ended Complexity, Ratings Task I and

¹ Please note: The results presented and discussed in this paper have not been reviewed by the author's graduate study committee and are therefore subject to potential revisions.

Ratings Task II. Given the nature of the tasks on each of the variables, it was highly likely that if a participant performed well on one of the tasks, they might perform equally well on another. Even though each dependent variable was unique in some aspect of the mental model to be captured, it was expected that there would be overlaps in content across dependent variables such that the five dependent variables could be correlated with each other. Given this assumption, a review of the correlations between variables was conducted.

Dependent Variable Correlations

As expected, significant correlations were observed between the pre- and post-tests for each of the five dependent variables with results ranging from correlations of .24-.50. All correlations between the pre-and post variable scores were significant at the $p < .01$ level except for the Open-Ended Complexity measure which was significant at the $p < .10$ level. Between dependent variables, but within the same test time element, significant correlations on the pre-test were observed between: Diagramming (Complexity) and Diagramming (System Dynamics) ($p < .01$) and on the post-test between: Open-Ended Complexity and Ratings Task I ($p < .05$); Open-Ended Complexity and Diagramming (Complexity) ($p < .05$); Ratings Task I and Diagramming (System Dynamics) ($p < .10$); and Diagramming (Complexity) and Diagramming (System Dynamics) ($p < .01$).

Multivariate Analysis of Variance

When an experimental study incorporates multiple dependent variables, which are highly correlated, the use of the Multivariate Analysis of Variance (MANOVA) is recommended. One assumption in the use of the MANOVA is that a measurement taken with one dependent variable may contain properties of response also found in another. Results of the 4 (Group) x 2 (Time) repeated measures MANOVA utilizing the three dependent variables found to be significant utilizing the Type III Analysis of Variance (ANOVA) presented earlier: Open-Ended Complexity, Ratings Task I, and Ratings Task II are presented in Table VI.

Table VI

**The 2 x 2 x 2 Doubly Multivariate Repeated Measures MANOVA
Using Significant Dependent Variables from ANOVA: Open-Ended Complexity, Ratings Task I and Ratings Task II**

Multivariate Effect	Λ	F	Df	<i>p</i>	η^2	Effect Size
Facilitation	.6378	9.84	3, 52	.0001***	.3622	Medium
Method	.8441	3.20	3, 52	.0307**	.1559	Small
Facilitation x Method	.8844	2.26	3, 52	.0918*	.1156	Small
Time	.8221	3.39	3,47	.0256**	.1779	Small
Facilitation x Time	.9657	.56	3, 47	.6468	.0343	Small
Method x Time	.9938	.10	3, 47	.9610	.0062	Small
Facilitation x Method x Time	.8429	2.92	3, 47	.0437**	.1571	Small
* = $p < .10$ ** = $p < .05$ *** = $p < .01$						

The MANOVA results shown in Table VI reveal a significant multivariate Main Effect for Facilitation ($p < .01$), Method ($p < .05$), and Time ($p < .05$). They also show a significant Interaction Effect for TimexMethod ($p < .10$) and TimexFacilitationxMethod ($p < .05$). However, since the Time x Facilitation x Method interaction is significant ($p = .04$) it may not be appropriate to interpret the components of this interaction independently of one another as they each may contain some portion of the of the significant TxMxF. Given this, the meaningful statistically significant interaction is TimexFacilitationxMethod. Based on these results, the following hypotheses are rejected:

- Ho₁: The method used will have a significant effect
- Ho₂: Facilitation of the group process will have a significant effect
- Ho₃: Time will have a significant effect
- Reject Ho₄: There will be significant interaction between Method x Time effect
- Reject Ho₅: There will be a significant interaction between Facilitation x Time effect
- Ho₆: There will be a significant interaction between Method x Facilitation x Time

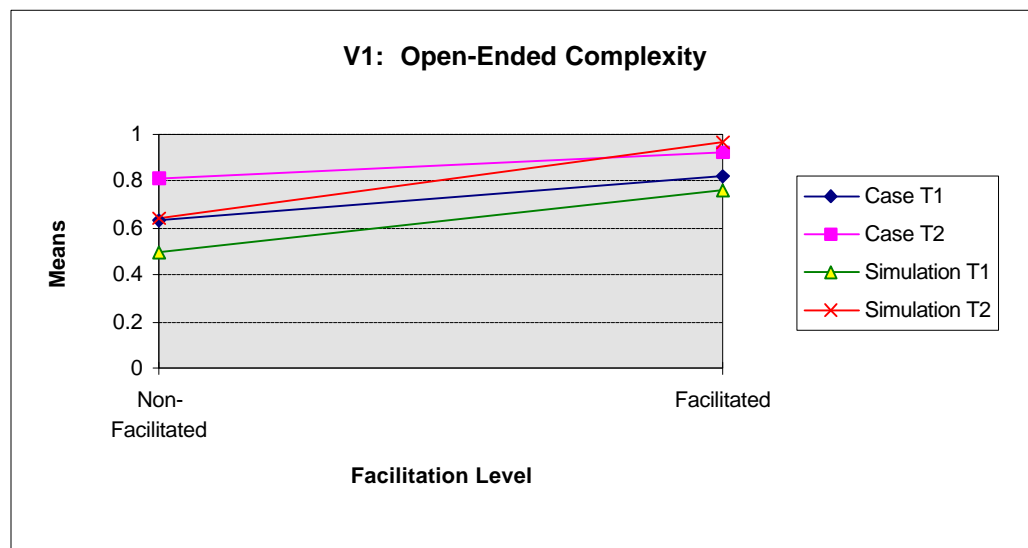
Discussion

In order to understand the dependent variable(s) that contribute to this three way (MxFxT) interaction it is helpful to look at the dependent variables in visual contrast to one another. The graphs below in Figure IIIa-IIIj show first the four conditions in relation to one another with Time1 and Time2 separate and then in the second graph, the “differences between the differences” or how, when time is taken into consideration the dependent variables compare under each of the two methodologies and levels of facilitation.

Figure III

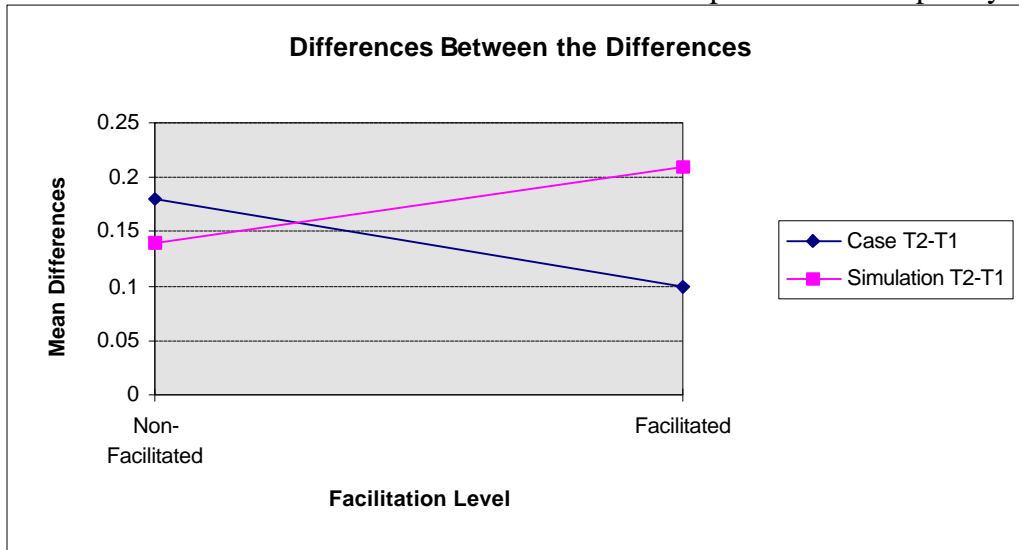
Interaction of Facilitation x Methods x Time

IIIa Comparing the Differences Between the Means: Open-Ended Complexity



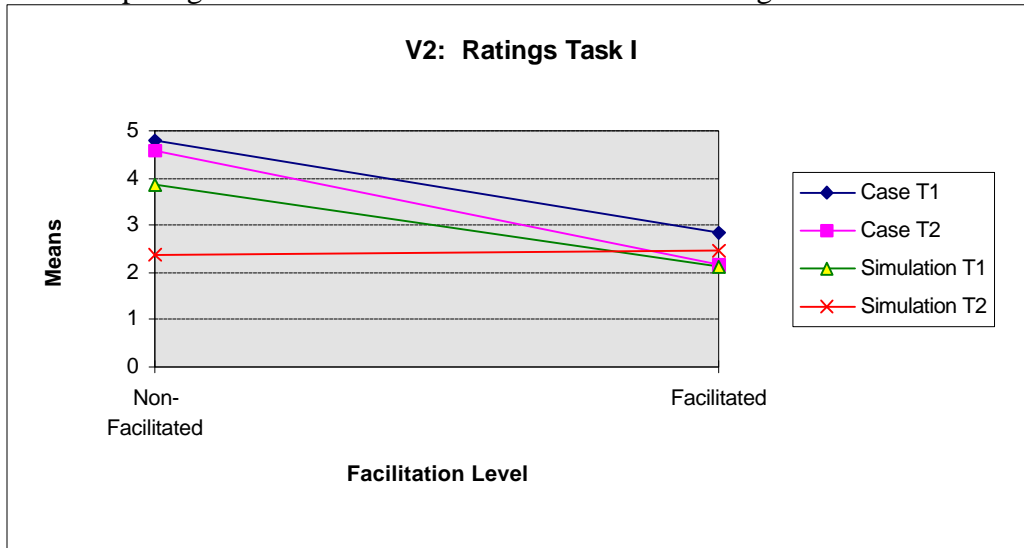
In general, it appears that facilitation of the process contributes to higher mean scores as all Facilitated conditions were shown to be higher than Non-Facilitated conditions. However, facilitation only occurs in the experimental process between T1 and T2 so it is necessary to consider the effect of Time more closely on this variable. The greatest difference between for Facilitation Level can be seen in the Facilitated Simulation condition that exhibited the greatest rise in slope between the difference of T2-T1.

IIIb Differences Between the Differences of T2-T1: Open-Ended Complexity



In looking at the differences between T2 and T1, it can be seen that while the Simulation method shows an improvement when Facilitated, the Case method does not, and performance over time across group members declines. Looking back to the MANOVA results, even though the MxF interaction has been removed as a source of significance because it may also contain the effects of Time, we can presume that the original significance level from MxF was probably partly derived from the Facilitated Simulation condition difference for Open-Ended Complexity variable as illustrated in IIIb here.

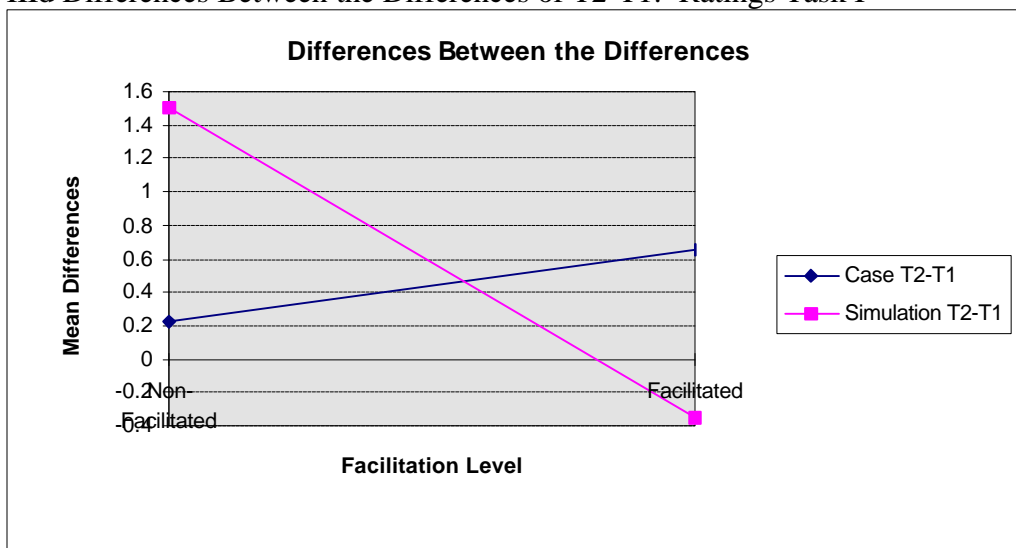
IIIc Comparing the Differences Between the Means: Ratings I



(Note: the closer the Mean is to (0) the better the performance on this task)

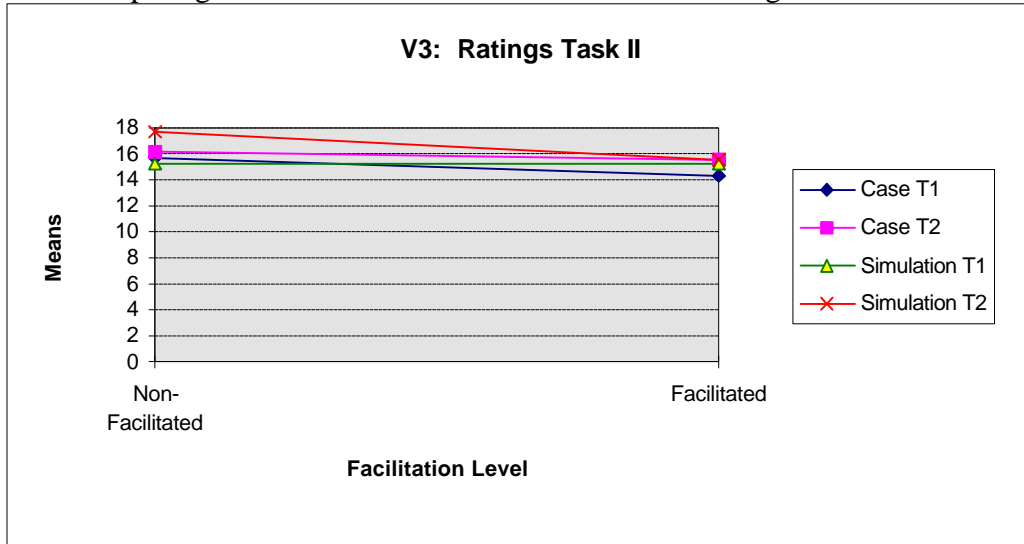
Looking at a comparison of the Facilitated vs. Non-Facilitated condition for Ratings Task I in IIIc, it appears that there is little difference between Facilitation and Non-facilitation in the Simulation condition as the slopes are quite flat between the two levels. This stands in stark contrast to the Open-Ended Complexity question where significant contrast between Facilitation and Non-Facilitation for the Simulation method was observed. Perhaps more notable is the trend demonstrated in Facilitation on all both conditions over time. If we rule out T1 results we can still see that there is a difference between Facilitation and Non-Facilitation groups in the Case condition on the Ratings I task with better performance observed in Facilitation. To understand better the nature of the contrast in the Case condition, consider the effect of time across the four conditions in Figure III d.

III d Differences Between the Differences of T2-T1: Ratings Task I



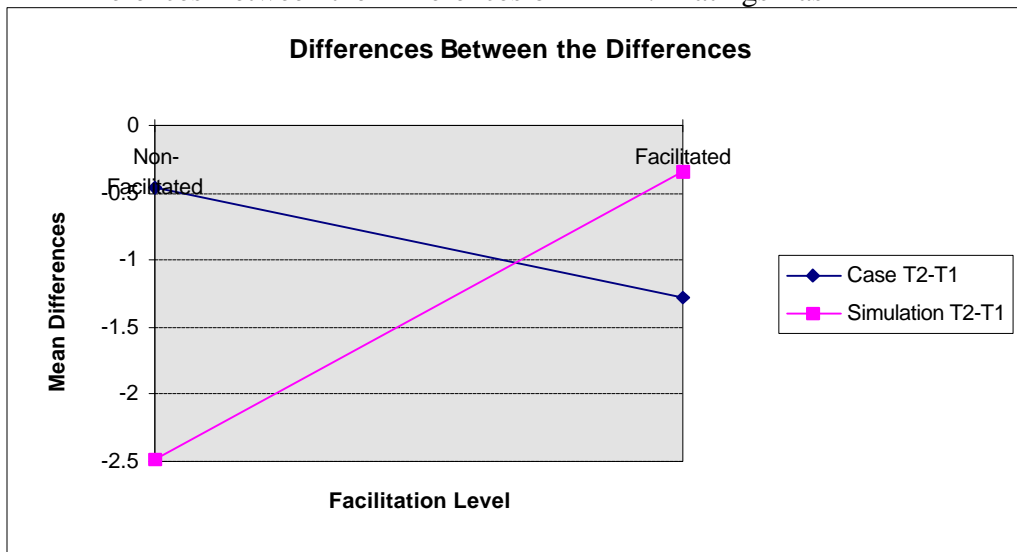
The Non-Facilitated Case group started out with the highest Mean score on the Ratings I task, but over time, the Case method showed greater improvement when Facilitated as can be seen in IIIc. On the other hand, the Facilitated Simulation showed a dramatic drop from T1 to T2.

IIIe Comparing the Differences Between the Means: Ratings II



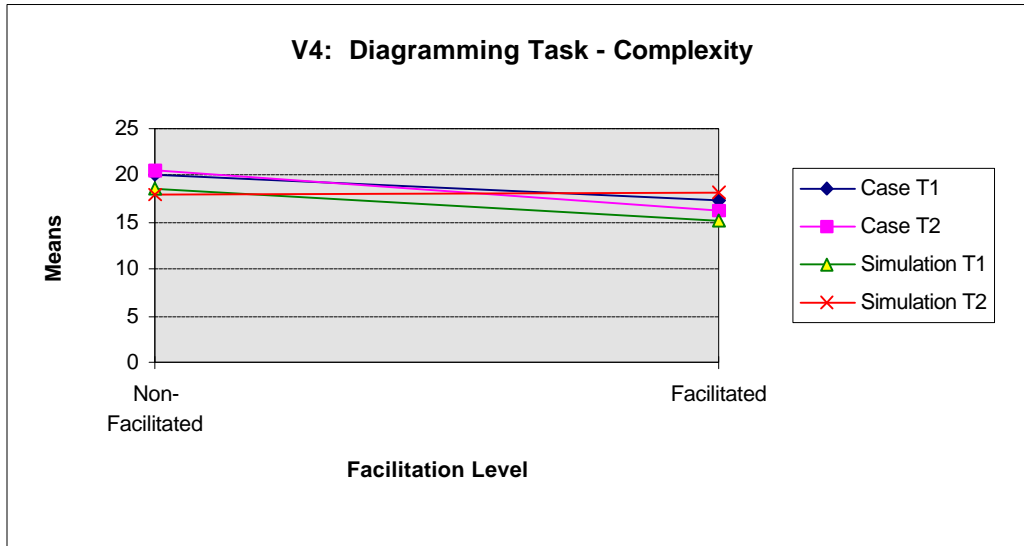
On Ratings II Facilitation appeared to be less influential and under Simulation may actually be somewhat of a detractor to performance as can be seen in IIIe.

IIIf Differences Between the Differences of T2-T1: Ratings Task II



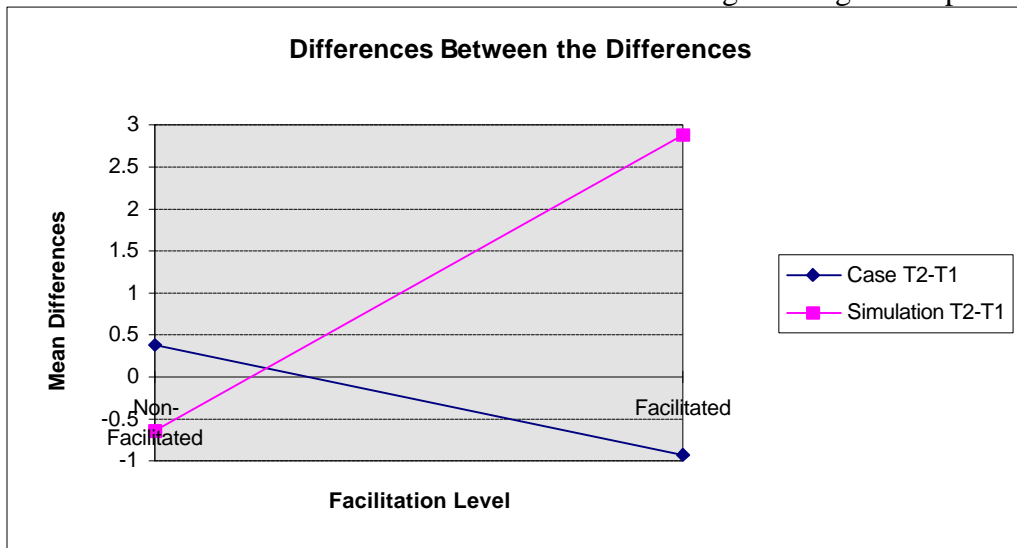
In looking at IIIf, there are striking contrasts between the Facilitation and Non-Facilitation in the Simulation condition as mean score differences from T1 to T2 for the Facilitated condition which were much higher under Simulation than the Non-Facilitation. It should be remembered however, that in talking about the Ratings II task, we are looking at performance loss across all groups in general as all exhibited “negative” or below (0) T2-T1 scores on this task.

IIIg Comparing the Differences Between the Means: Diagramming Task – Complexity



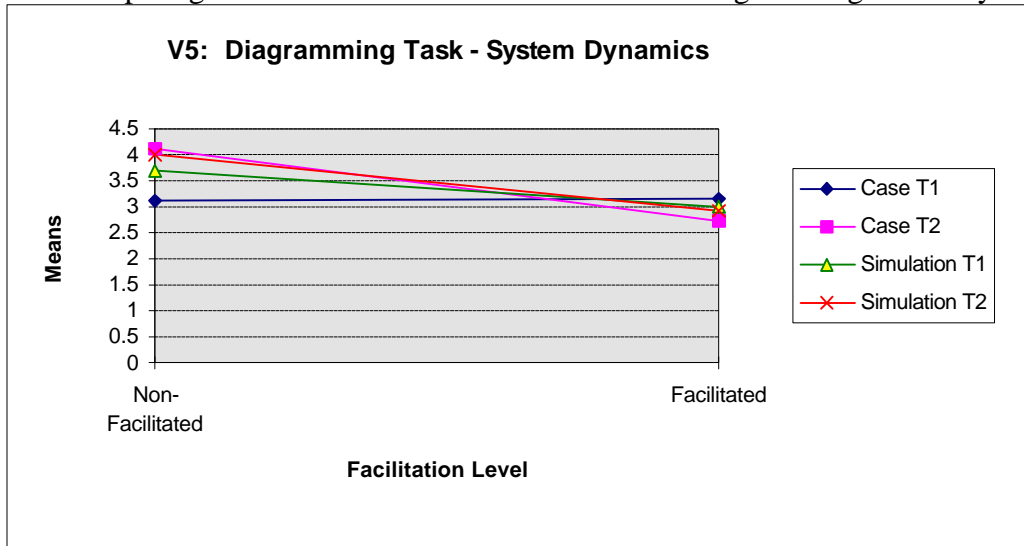
As with the second ratings task, there is not a big difference in Facilitation between Methods although a difference can be seen in Simulation at T2 for Facilitation as evidenced by a rise in slope. In all other instances, Non-Facilitation appears to be favorable on the Diagramming task for Complexity.

IIIh Differences Between the Differences of T2-T1: Diagramming – Complexity Component



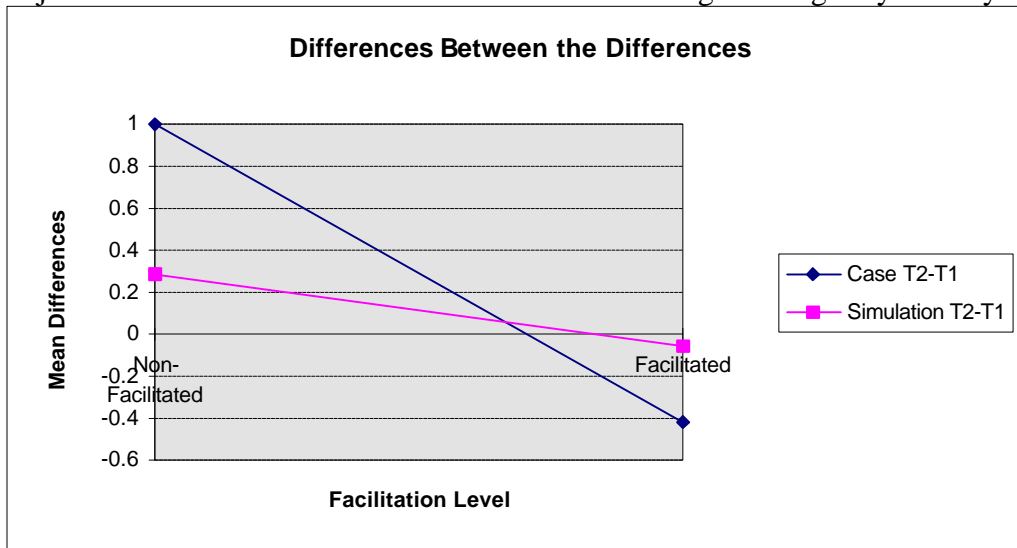
In looking at the differences over time however, a difference can be seen in Facilitated Simulation. While performance levels were below or near zero at T1 for both methods, at T2 Facilitation had a negative impact on the Case method and a large, positive impact on Simulation. These results are consistent with those observed on the Open-Ended Complexity variable.

IIIi Comparing the Differences Between the Means: Diagramming Task - System Dynamics



Finally, Figure IIIi shows a trend consistent with the previous three variables in that Facilitation has a Mean score across groups that is less than Non-Facilitation in all instances except the Case condition at T1 where Facilitation scores were slightly higher to begin with.

IIIj Differences Between the Differences of T2-T1:Diagramming – System Dynamics Component



As with the previous variables however, there is a difference when Time is taken into consideration as Non-Facilitation appears to have a much greater impact across methods on the System Dynamics aspects of the Diagramming task. While the difference between Facilitation and Non-facilitation is evident in the Simulation condition, a more dramatic difference can be seen in the Case condition whereby Non-facilitation shows a dramatically larger improvement in performance from T1 to T2. This stands in sharp contrast to the previous element of the Diagramming task - the Complexity component.

Discussion of Interaction of Facilitation and Method

In looking at the “Differences between the Differences” figures for all five dependent variables, facilitation has a positive impact under some methods on some variables and a negative impact on others as reflected by a drop in slope from right to left on these Figures. On three of the five task variables, facilitation had a positive effect when combined with the Simulation method (Open-Ended Complexity, Ratings II and Diagramming Complexity). Were there increases in slope for the Case method on these variables as well we might be inclined to say that Facilitation in general has a positive effect. But, when looking at the results presented here, in actuality Facilitation had a “negative” impact when combined with the Case method. When Facilitation is taken into consideration on these same three variables, the gains in performance were much less.

Further, Facilitation seems to be a hindrance for both methods in helping individuals develop a better understanding of the dynamics at work within the system as Facilitation resulted in much less performance improvement for both Case and Simulation as measured on the System Dynamics aspect of the Diagramming variable represented in Figure IIIj. However, if we look at the effect of Facilitation on the variables taken as a whole, we can not come to the same conclusion as shifts in performance between Case and Simulation occur under Facilitation on some variables. On one of the five dependent variables (Open-Ended Complexity) the Facilitated Simulation group performed better than the Case group and on three of the five variables (Ratings Task I and the two Diagramming assessments) the Case condition outperformed the Simulation groups on the post-test even though the Simulation groups started out with better pre-test scores.

Summary of Results

It would seem evident by these results that there is more to the relationship between the Methodology, Facilitation and mental models measurement than might first appear. Given this, as researchers and practitioners designing and evaluating Systems Thinking interventions, it would be in our best interest to ask, “which level of Facilitation under which Method has an impact in the “as expected” direction? Moreover, which mental model measurement methods should be used to assess results? While the first two dependent variables and measurement methods appear to be more robust as reflected in “as expected” direction in results, the Ratings I task seems to be more discerning or sensitive in detecting changes in participants’ mental models and in detecting differences across methods and facilitation levels. This then raises the question, “do some people really “know more” about what’s being asked in the Ratings task or is the Open-Ended Complexity question simply easier from the start?”

Clearly Facilitation of the Group Model Building process helps in some cases. Based on results found here it appears to have the greatest effect in helping individuals understand better the need for varying the types and levels of the control variables used to alter the system over time as suggested by better performance on the Open-Ended Question of strategy. Just how helpful Facilitation is in aiding a group on other Systems Thinking tasks is not as clear however. Certainly there may be other factors at work that may account for differences in effectiveness across variables found here.

Potential Confounding Factors

Of the methods used for assessing mental models only the Open-Ended Complexity question and the Ratings I task had a consistent “as expected” result in scores from T1 to T2. Perhaps these two methods are better suited to capturing changes in mental models for of the task the participants were asked to perform, these two were simpler while the second Ratings task and the Diagramming task were much more difficult and required more cognitive effort.

In looking at the results, it may be that the second Ratings task was very difficult to complete, as this was the only variable on which there was missing data. On this task individuals were asked to indicate the level of influence (using a five-point scale) that each of the six control variables would have on each of the three potential populations in the system. This required that participants think through and rate 18 different combinations. In contrast, on the first Ratings task which was much simpler (discerning the effect using a three point scale of the six variables on the system overall) performance was closer to the ratings given by the model expert. The differences in these two tasks suggest that the second ratings in which performance dropped on the post-test for all four experimental conditions could have confounded the overall results.

Another potential confounding factor affect results may be the presence of a fatigue factor in which performance on tasks completed at the end of a repeat measures design are given less attention and effort by participants in the study due to tiredness. If so, it should occur in this study on the Diagramming task that came last. But “opposite from expected” results on the Diagramming task were inconsistent across groups. Both methodologies and facilitation levels did better on some aspects of the Diagramming task but not on others. Hence, there is no conclusive evidence to presume a fatigue factor that could explain the drop in scores on the post-tests but it should not be ruled out as a potential confounding factor for some participants particularly those attending afternoon sessions of the training (Facilitated sessions).

Lastly, even though skilled facilitators were and a scripted facilitation process was employed, it can not be said with certainty that the Facilitators exhibited an equal level of “facilitation” within their groups. The audio recordings made of the group activities being analyzed in the second portion of this study may provide further insights into the equality of groups on the facilitation dimension.

How Results Fit with Earlier Studies

In the study by Vennix et al., (1996), participants reported having learned more about the complexity of relationships in their organizational system as a result of Conceptual Model Building (a Case Study corollary) but reported that their views on strategy had not changed. In this study, the findings under the Case methodology partially reflect this observation. Whereas the written strategies in this study became more detailed and included variations in strategy over time, only the Non-Facilitated groups reflected more complexity and understanding of System Dynamics in their diagrams. The Facilitated Case group surprisingly, did not reflect a deeper understanding for the complexity of the system in their post-test diagrams and reduced or eliminated some of the detail present in their pre-test assessments.

Secondly, whereas Wolfe (1975) found that Non-Facilitated Case groups showed no gain in test scores but a gain in Facilitated groups, in the application of the Case method used here, this was not the case as the Non-Facilitated groups performed better than the Non-Facilitated Case groups in general. Lastly, the observations made during the experimental process used in this study support the observations made in earlier Gaming studies (Wolfe, 1975b) in that participants

using Simulation often failed to apply formal, rational analyses to finding the best strategy and did not take the time to articulate the rationale for their overall strategy without being prompted by a facilitator to do so.

In comparing results found here with earlier evaluations of the Simulation method, in the Doyle et al (1996) study, it was reported that participants models were more complex (more variables added post) as a result of using Simulation. Results here reflect this finding but only in the Facilitated conditions. In the Non-Facilitated Simulation groups, less complexity was found in the diagrams after the group activity while more System Dynamics understanding was articulated.

Implications of the Findings for Research and Practice

When the differences in results associated with Facilitation found here are put side-by-side, Non-Facilitation of the Case method appears to be more effective in enhancing complexity of participants' mental models on two of five variables whereas, Facilitation resulted in the largest increases in performance on three of the five dependent variables. Given this, So in practice, we might expect that Simulation via a Management Flight Simulator might help people understand the complexity of their systems better.

Given this contrast, it could be said that clearly, facilitation of the process is important but to what extent and how effective it will be on different aspects of learning is not clear. Overall, it may be said that results this study share one thing in common with the finding of the studies that compared Case method to Gaming in that results here have been contradictory and oftentimes ambiguous suggesting that further research is needed.

Final Remarks

Even though this study is considered experimental in nature because of the controlled conditions that were put into place in the execution of the training workshops, the real-life conditions under which it was conducted and the general "newness" of the technology employed in the workshops lend themselves to classifying this type of study as exploratory – it was impossible to predict ahead of time what would happen when employing measurement techniques, training methodologies and content that were in themselves brand new and tried only in one previous trial study. By using a simulation that was recently developed it ensured that participants in the study had no prior exposure, however, it also opened up a whole host of concerns around how well participants would understand the simulation task and how to use the Management Flight Simulator. In addition, with the exception of the trial study, the quality and integrity of the case used was unknown. Finally, employing the Diagramming and Ratings tasks was new and with the exception of one trial study untested.

When new training and research methods are employed as they were here there are sure to be a few surprises in the implementation of the process that in and of themselves provide fuel for future studies. While not the focus of this research program, the following are a few of the things observed that should be considered in the Group Model Building practice:

- It was difficult to entice Simulation groups to document what they learned as they went, they usually developed a rationale for their strategy after the fact.
- An assigned facilitator may not be strong enough to guide the facilitation process.
- Non-facilitated groups may still have a strong facilitative member.
- The case process was more successful in eliciting a rationale from the group for their chosen strategy.

In terms of future evaluation Group Model Building evaluation studies:

- Scoring diagramming is very difficult – there is a lot of variation in how people go about the task.
- Determining which aspects of mental models one is most interested in is critical but equally important is finding measurement methods that best captures those aspects of thinking. At present they are not readily pre-defined in the literature and may require development.
- Don't assume that the results of one measure will agree with the results of another – be prepared for contradictory and unexplainable findings.

In conclusion, no research methodology is without its weaknesses but that does not provide reason enough to abandon an evaluation effort. Given the willingness of a host organization to participate in the study in order to give their employees an opportunity to learn about something new, has secondary benefits that go beyond the research agenda. The efforts described here not only looked at the robustness of the methods, they provided a forum for talking about organizational systems in a way that was different, provocative and a lot of fun. These secondary benefits should not be underestimated in their ability to take us a long way toward learning about and improving our efforts.

References

- Akkermans, H. A. & Vennix, J. A. M. (1997). Clients' opinions on group model-building: an exploratory study. System Dynamics Review, 13, 1, 3-31.
- Andersen, D. F., & Richardson, G. P. (1997). Scripts for Group Model Building. System Dynamics Review, 13, 107-129.
- Anderson, T., Tolmic, A., Howe, C., Mayes, T., & MacKenzie, Mhairi (1992). Mental models of motion. In Y. Rogers, A. Rutherford and P. A. Bibby (Eds.) Models in the mind: Theory, perspectives and application (pp. 57-71). San Diego: Academic Press.
- Argyris, C. (1990). Overcoming organizational defenses. Boston: Allyn and Bacon.
- Argyris, C. & Schön, D. (1978). Organizational learning: A theory of action perspective. Reading, MA.: Addison-Wesley.
- Argyris, c. & Schön, D. (1996). Organizational learning II: Theory, method, and practice. Reading, MA: Addison-Wesley.
- Bakken, B., Gould, J., & Kim, D. (1994). Experimentation in learning organizations: A Management Flight Simulator approach. In J. D. W. Morecroft & J. D. Sterman, (Eds.), Modeling for learning organizations (pp. 243-266). Portland, OR: Productivity Press.
- Cannon-Bowers, J. A., Salas, E., and Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Costellan, Jr. (Ed.), Individual and group decision making (pp. 221-246). Hillsdale, NJ: Erlbaum.
- Carley, K. & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. Social Forces, 70, 602-636.
- Cavaleri, S. A. & Thompson, J. P. (1996). Assessing the efficacy of microworlds for promoting Systems thinking. In G. P. Richardson & J. P. Sterman (Eds.), Proceedings of the 1996 International System Dynamics Society (pp. 189-199). Cambridge, MA.
- Converse, S. A. & Kahler, S. E. (1992). Knowledge acquisition and the measurement of shared mental models. (Contract No. DAAL03-86-D-0001). Orlando, FL: Naval Training Systems Center, Team Performance Laboratory.
- Coyle, R. G. & Alexander, M. D. W. (1997). Two approaches to qualitative modeling of a nation's drugs trade. System Dynamics Review, 13, 3, 205-222.
- Craik, K. J. W. (1943 reprint 1952). The nature of explanation. Cambridge: University Press.
- de Geus, A. P. (1988, March - April). Planning as learning. Harvard Business Review, pp. 70-74.

- Doyle, J. K., & Ford, D. N. (1998). Mental models concepts for Systems dynamics research. System Dynamics Review, 14, 3-29.
- Doyle, J. K., Radzicki, M. J., & Trees, W. S. (1996). Measuring the effect of Systems thinking interventions on mental models. In G. P. Richardson & J. P. Sterman (Eds.), Proceedings of the 1996 International System Dynamics Society (pp. 129-132). Cambridge, MA.
- Drath, W. H., & Palus, C. J. (1994). Making common sense: Leadership as meaning-making in a community of practice. Greensboro, N.C.: Center for Creative Leadership.
- Eden, C. (1989). Using cognitive mapping for strategic options development and analysis (SODA). In J. Rosenhead (Ed.), Rational analysis for a problematic world: Problem structuring methods for complexity, uncertainty and conflict (pp. 21-42). New York: John Wiley & Sons
- Eden, C. (1994). Cognitive mapping and problem structuring for Systems dynamics model building. System Dynamics Review, 10, 2-3, 257-276.
- Eden, C., Jones, S. & Sims, D. (1983). Messing about in problems: An informal structured approach to their identification and management. New York: Pergamon
- Forrester, J. W. (1975). Counterintuitive behavior of social systems. Collected papers of Jay W. Forrester (pp. 211-234). Cambridge, MA.
- Forrester, J. W. (1994). Policies, decisions, and information sources for modeling. In J. D. W. Morecroft & J. D. Sterman, (Eds.), Modeling for learning organizations (pp. 51-84). Portland, OR: Productivity Press.
- Gordon, S. E., Schmierer, and Gill, R. T. (1993). Conceptual graph analysis: Knowledge acquisition for instructional system design. Human Factors, 35, 459-481.
- Graham, A. K., Morecroft, J. D. W., Senge, P. M., & Sterman, J. D. (1994). Model-supported case studies for management education. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 219-241). Portland, OR: Productivity Press.
- Hodgson, A. M. (1994). Hexagons for Systems thinking. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 359-374). Portland, OR: Productivity Press.
- Huff, A. S. (Ed.) (1990). Mapping strategic thought. New York: John Wiley and Sons.
- Huz, S., Andersen, D. F., Richardson, G. P., & Boothroyd, R. (1996). Evaluating Group Model Building in mental health and vocational rehabilitation services. In G. P. Richardson & J. P. Sterman (Eds.), Proceedings of the 1996 International System Dynamics Society (pp. 233-236). Cambridge, MA.

- Isaacs, W. & Senge, P. M. (1994). Overcoming limits to learning in computer-based learning environments. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 267-287). Portland, OR: Productivity Press,.
- Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard University Press.
- Kelly, G. A. (1955). The psychology of personal constructs. Vols. 1 and 2. New York: Norton.
- Lane, D. C. (1994). Modeling as learning: A consultancy methodology for enhancing learning in management teams. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 85-117). Portland, OR: Productivity Press.
- Lane, D. C. (1995). On a resurgence of management simulations and games. Journal of the Operational Research Society, 46, 604-625.
- Levitt, B., & March, J. (1988). Organizational learning. Annual Review of Sociology, 14, 319-340.
- Luria, A. R. (1973). The working brain: An introduction to neuropsychology. Middlesex, England: Penguin Books, Ltd.
- Moore, L. F. (1967). Business games vs. cases as tools for learning. Training and Development Journal, 21, 13-23.
- Morecroft, J. D. W. (1984). Strategy support models. Strategic Management Journal, 5, 215-229.
- Morecroft, J. D. W. (1994). Executive knowledge, models, and learning. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 3-28). Portland, OR: Productivity Press.
- Morecroft, J. D. W. & Sterman, J. D. (Eds.) (1994). Modeling for learning organizations Portland, OR: Productivity Press.
- Narayanan, V. K., & Fahey, L. (1990). Evolution of revealed causal maps during decline: A Case Study of Admiral. In A. S. Huff (ed.), Mapping strategic thought (pp. 109-133) New York: John Wiley & Sons.
- Richardson, G. P. & Anderson, D. F. (1995). Teamwork in Group Model Building. System Dynamics Review, 11, (2), 113-137.
- Richmond, B. (1997). *The Strategic Forum: Aligning objectives, strategy and process*. System Dynamics Review, 13, (2), 131-148.

- Rogers, Y. (1992). Capturing mental models. In Y. Rogers, A. Rutherford and P. A. Bibby (Eds.) Models in the mind: Theory, perspectives and application (pp. 241-246). San Diego: Academic Press.
- Rosenhead, J. (1989). Introduction: old and new paradigms of analysis. In J. Rosenhead (Ed.), Rational analysis for a problematic world: Problem structuring methods for complexity, uncertainty and conflict (pp. 1-20). New York: John Wiley & Sons
- Rouse, W. B. & Morris, N. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100, 349-363.
- Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., & Cooke, N. M. (1985). Measuring the structure of expertise. International Journal of Man-Machine Studies, 23, 699-728.
- Senge, P. M. (1990a). Catalyzing Systems thinking within organizations. In F. Massaryk (Ed.), Advances in organization development (pp. 197-246). Norwood, NJ: Ablex..
- Senge, P. M. (1990b). The fifth discipline: The art and practice of the learning organization. New York: Doubleday Currency.
- Senge, P. M., Kleiner, A., Roberts, C., Ross, R. B., Smith, B. J . (1994). The fifth discipline fieldbook: Strategies and tools for building a learning organization. New York: Doubleday.
- Senge, P. M. & Sterman, J. D. (1994). Systems thinking and organizational learning: Acting locally and thinking globally in the organization of the future. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 195-216). Portland, OR: Productivity Press.
- Shapiro, A. (1998). The Tipping Point: A Simulation of Organizational Change. <http://www.the-tipping-point.com>.
- Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. Management Science, 35, 321-339.
- Vennix, J. A. M. (1996). Group Model Building. New York: John Wiley & Sons.
- Vennix, J. A. M., Akkermans, H. A., & 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. & Rohrbaugh, J. (1994). Model building for group decision support: Issues and alternatives in knowledge elicitation. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 29-50). Portland, OR: Productivity Press.

- Vennix, J. A. M. & Gubbels, J. W. (1994). Knowledge elicitation in Conceptual Model Building: A Case Study in modeling a regional Dutch health care system. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 121-145). Portland, OR: Productivity Press.
- Wickens, C. D. (1992). Engineering psychology and performance. (2nd Edition). New York: Harper Collins.
- Williams, M. D., Hollan, J. D., & Stevens, A. L. (1983). Human reasoning about a simple physical system. In D. Gentner & A.L. Stevens, (Eds.), Mental models (pp. 131-153). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wilson, J. R., & Rutherford, A. (1989). Mental models: Theory and application in human factors. Human Factors, 31, 617-634.
- Wolfe, J. (1973). The comparative learning effects of a management games vs. case work in the teaching of business policy. Proceedings of the 33rd Annual Meeting of the Academy of Management, (pp. 291-298).
- Wolfe, J. (1975a). Effective performance behaviors in a simulated policy and decision making environment. Management Science, 21, 872-882.
- Wolfe, J. (1975b). A comparative evaluation of the experiential approach as a business policy learning environment. Academy of Management Journal, 18, 442-452.
- Wolfe, J. (1976). The effects and effectiveness of simulations in business policy teaching applications. Academy of Management Review, April, 47-56.
- Wolfe, J., & Guth, G. R. (1975). The game approach versus gaming in the teaching of business policy: An experimental evaluation. The Journal of Business, 3, 349-364.
- Wolstenholme, E. F. (1994). A systematic approach to model creation. In J. D. W. Morecroft & J. D. Sterman (Eds.), Modeling for learning organizations (pp. 121-145). Portland, OR: Productivity Press.
- Wolstenholme, E. & Coyle, R. G. (1983). The development of Systems dynamics as a methodology for system description and qualitative analysis. Journal Operational Research Society, 34, 569-581.
- Woods, D. D., Johannesen, L. J., Cook, R. I., Sarter, N. B. (1994). Behind human error: Cognitive systems, computers, and hindsight. Wright-Patterson Air Force Base, Ohio: Crew Systems Ergonomics Information Analysis Center.