THE UNIVERSITY OF SYDNEY

When to Map and When to Model: The Effect of System Dynamics on Decision Making and Forecasting

RAJAT DHAWAN

A THESIS SUBMITTED TO THE UNIVERSITY OF SYDNEY IN
FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

BUSINESS INFORMATION SYSTEMS

JANUARY 2008

Faculty of Economics and Business

Declaration

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it contains no material previously written by another person nor material which to substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgement is made in the text.

[redacted]

Rajat Dhawan

ACID FREE
UNI 2

To

My parents, R.K. & Usha Dhawan,

who inculcated in me that education is of paramount importance

and

for their unconditional love, endless support and encouragement.

Acknowledgements

The "thesis journey" is a long and challenging one. It is not a journey that can be travelled alone and I need to say thank you to some very important people who helped me tread the path.

I owe an immeasurable debt to my primary supervisor, Professor Marcus O'Connor, who has intelligently guided me along the way. This work benefited immensely from Marcus' advice, constructive criticism, and encouragement. I express my deepest gratitude to him for believing in my potential and supporting me throughout the PhD.

Dr. Mark Borman played a much greater role than of a traditional associate supervisor. His contribution to the thesis is significant. I admire his acute sharpness through which he always made me see the flip side of things. His manner of thinking has had a profound impact on my intellect.

A special thanks to Monika Kent for helping out with coding of participants' responses. Several interesting discussions with her added a new dimension to my analysis.

Most of all I need to thank my brother Rohit. Rohit and I had numerous interesting discussions about the research in these three plus years. These helped me refine my ideas and challenge my thinking. If not for his company, this journey would have been quite lonely.

Finally, I wish to thank Divya, my wife, who has been a great source of encouragement for me and shared my never-ending PhD stories with patience. She spent too many Sundays alone as I worked on this project. Without her support and understanding it would have been much, much harder.

Abstract

Effective decision making in a world of growing complexity requires us to understand how the structure of complex systems creates their behaviour. System dynamics is a perspective that provides mapping and computer modelling tools to help decision makers take a holistic view of a system. Even though system dynamics has been widely applied, there is a lack of established experimental results regarding the individual and relative efficacy of qualitative and quantitative system dynamics. This research evaluated the effectiveness of system dynamics in decision making and forecasting.

Three rigorous experiments were conducted in a pre-test/ post-test setup. Around hundred participants were provided with qualitative and/or quantitative system dynamics training at appropriate interventions to complete tasks that were classified as simple and complex. The experiments measured performance, the logic applied and the level of confidence for each participant.

Overall, qualitative, quantitative and a combination of both were all useful tools in improving performance compared to a situation where none were employed. However, qualitative system dynamics is useful enough to solve simple tasks but complex tasks benefit from the use of quantitative computer modelling. Over a period of time, though it was easier for participants to retain previously learnt qualitative training, quantitative skills suffered a relapse.

There are two practical implications of this research. Firstly, these results provide guidance to practitioners in deciding when to apply qualitative tools and when to use quantitative modelling. Secondly, results of the longitudinal study provide trainers to assess the opportunity and cost of training employees on qualitative and quantitative tools.

Table of contents

List of tables

List of figures

Chapter 1
Introduction

1.1 Background to the research

A complex system is made up of multiple parts that when considered in interaction with each other exhibit a high level of complexity. Ant-hills, human economies and human beings are examples of complex systems.

System dynamics is a methodology to study complex feedback systems. Such systems are ubiquitous, especially in business organizations. The field was founded by J.W. Forrester at the Massachusetts Institute of Technology in the early 1960s. The focus of system dynamics is to study feedback loops and stocks and flows; taking a holistic view of the system under context. The System Dynamics Society describes the methodology as "a methodology for studying and managing complex feedback systems, such as one finds in business and other social systems."

The following subsection describes an application of the system dynamics process to a real-life problem. Although system dynamics has been applied to solve complex problems in various industries (Sterman 2000), its applications to project management have been most widespread (Rodrigues and Bowers 1996). These applications include but are not limited to software projects (Abdel-Hamid and Madnick 1991; Barlas and Bayraktutar 1992), ship building projects (Cooper 1980a; Sterman 2000) and research and development projects (Roberts 1964; Kelly 1970; Richardson and Pugh 1981; Keloharaju and Wolstenholme 1989). It is a well-known fact that projects consistently fall behind schedule, exceed the budget and are not able to produce deliverables of desired quality (Reichelt and Lyneis 1999). Traditional tools of project management and approaches to software engineering have often failed to produce desired results as they adopt a linear approach and are unable to deal with complexity (Lyneis et al. 2001; Rodrigues and Williams 1995). Such a problem is a typical candidate for system dynamics modelling.

An example of one such application is described in Lyneis et al. (2001). The study details of the application system dynamics to decisions that are taken up-front in designing the project and those decisions that have a long-term impact on the project. The authors of the study, as employees of PA Consulting group, undertook this project for their clients—Hughes Aircraft Company (now part of Raytheon Corporation). Hughes was building an aircraft for the US Air Force on behalf of the Kingdom of Saudi Arabia (the project was called the Peace Shield Air Defense System). Hughes had 54 months to complete the more than $1 billion dollar contract project that involved both software and hardware components. This time period was perceived unrealistic by many stakeholders. System dynamics models (both causal loop and stock/flow) were developed to highlight the relationship between key factors and to identify feedback loops. A stock/flow structure showing feedback loops is shown below (Figure 1.1). This model was applied to the Peace Shield program.

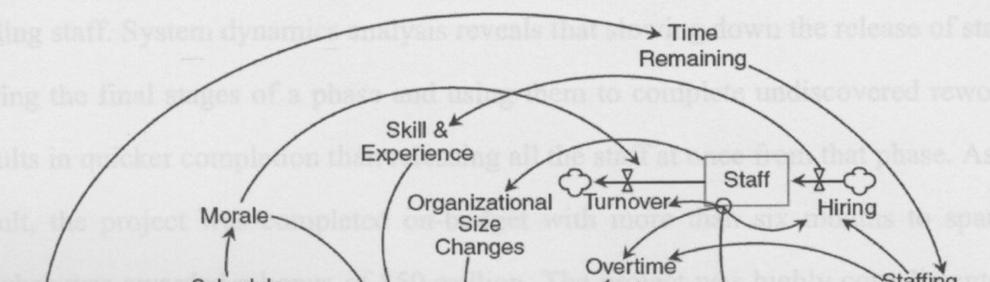


Figure 1.1: A typical system dynamics model showing causal relationships and stocks and flows (Lyneis et al. 2001)

To begin with, system dynamics was used to support the bidding process and risk assessment of the project. This system dynamics model was built on the basis of interviews and discussions with Hughes' managers and on a similar model that was previously constructed. Using the model, 'what-if' scenarios were conducted to analyse the impact of factors such as re-use of previously written code, availability of staff and time delays. This process helped in identifying the factor that was most 'sensitive' to the model output. Mitigation plans were then drawn based on this assessment. The model also helped to analyse likely bids from competitors. The knowledge gained from the model was an important input to the final bid placed by Hughes. PA consulting also supported the project during the execution phase. As the project progressed, the initial model was continually updated. The system dynamics model helped the client team in assessing alternative software development approaches as well. They found that a 'teaming' approach reduced time for completion when compared to the traditional waterfall approach. Another unconventional approach was also used for the release of software engineering and coding staff. System dynamics analysis reveals that slowing down the release of staff during the final stages of a phase and using them to complete undiscovered rework results in quicker completion than releasing all the staff at once from that phase. As a result, the project was completed on-budget with more than six months to spare. Hughes was awarded a bonus of $50 million. The project was highly complimented and rated as 'highly successful' by the U.S. Air Force. Similar projects that had been undertaken in the past were far from being as successful as the Peace Shield Program.

Use of system dynamics tools

In its initial years, system dynamics was perceived and used as a purely quantitative technique. It used stocks and flows and mathematical equations to create simulation models that could be run on digital computers (Coyle 2000). Around the 1980s, system dynamics borrowed qualitative tools from allied system methodologies such as soft systems methodology. Qualitative system dynamics provides a foundation for constructing and understanding formal simulation models. It has been well established that causal loop diagrams, one of the tools of qualitative system dynamics, may facilitate the way in which we can externalize mental models by introducing circular causality (Richmond 1997). They provide a way to simulate mental maps, their modes of behaviour, and our assessment of a changed system (Wolstenholme 1999). These pen and paper tools such as causal loop diagrams have become part of mainstream system dynamics process. In turn, causal loop diagrams benefit from the output of formal modelling that provide for enhanced mental models. Hence, the use of qualitative and quantitative methods for systems modelling may be perceived as iterative in nature, one benefiting from the other and vice-versa. The approach of combining the two methodologies is most common in the field today with most authors incorporating both to describe the complete system dynamics process (e.g. Wolstenholme 1983, Wolstenholme 1990; Maani and Cavana 2000). The majority of system dynamics applications consist of both, qualitative and quantitative analysis.

Though the use of a combination of qualitative system dynamics and quantitative system dynamics is popular for inferring the behaviour of a system through its feedback structure, the last twenty-five years have also seen enthusiastic support for the use of qualitative system dynamics alone, which some authors believe may be sufficient for solving problems of dynamic complexity (Wolstenholme 1999). Wolstenholme and Coyle (1983) argue that there could be value in simply using rigorous qualitative approaches to facilitate system description. Some of the significant studies using qualitative system dynamics alone include those by Wolstenholme (1983) and Cavana et al. (2004). On the other hand, others have doubted its efficacy and its adequacy in dealing with such problems (e.g. Homer and

Oliva 2001). Richardson (1996) acknowledges an increase in the use of qualitative system dynamics tools alone, and expresses concern over the usage of these tools by those who lack the knowledge of quantitative modelling. It has also been argued that causal loop diagrams may not make any distinction between information links and rate-to-level links, and the standard characterizations of positive and negative polarities may be wrongly interpreted (Richardson 1986/76).

The isolated usage of qualitative tools has sparked a debate between the relative effectiveness of qualitative system dynamics and quantitative system dynamics (e.g. Coyle 2000; Coyle 2001; Homer and Oliva 2001; Luna-Ryes and Andersen 2003; Howick et al. 2006). Consequently, system dynamics experts have asserted the need of an investigation to examine the value added by quantified modelling over qualitative analysis (Coyle 2000) and to the benefit in knowing when to use qualitative tools and when to use quantitative tools (Richardson 1999).

In general, experimental evaluation of the effectiveness of system dynamics has received little attention in comparison to the amount of consulting work involving system dynamics (Pala and Vennix 2005; Maani and Maharaj 2004; Cavaleri and Sterman 1997; Doyle 1997). Although some studies experimentally explore the links between system dynamics and task performance to some extent, the relative effectiveness of qualitative and quantitative stages of system dynamics still remains unclear and has not yet been subject to rigorous experimental testing. Furthermore, the long-term efficacy of system dynamics tools represents an unexplored area of research. Among the deterrents to such research is lack of commitment by clients, the associated costs and paucity of adequate methodologies to carry out such studies (Huz et al. 1997).

1.2 Research problem and research questions

System dynamics is a well-established methodology that has been applied to a range of problem areas including those in business and corporate policy to social and public policy (Scholl 1995 contains a comprehensive review). System dynamics is interdisciplinary, with its roots in general systems theory and is practiced all over the world. The System Dynamics Society boasts of over 1000 members in over 70 countries (System Dynamics Society 2007; accessed 1 January 2006). The methodology is taught in all regions of the world and is part of curricula of major business schools like London Business School and MIT Sloan School of Management. Qualitative and quantitative phases of system dynamics are applied alone or in combination with each other to understand complex problems. Successful applications of these variants can be found in leading management journals (such as Management Science, Organizational Behavior and Human Decision Processes).

Although there are ample anecdotal claims of the effectiveness of the system dynamics methodology, there are few experimental studies that test the links between system dynamics interventions and performance in dynamically complex tasks (Pala and Vennix 2005; Maani and Maharaj 2004; Cavaleri and Sterman 1997; Doyle 1997). For instance Sweeney and Sterman (2000) remark:

> "*...there is little evidence, or even systematic research, to support educators' and consultants' faith in its [system dynamics'] efficacy*" (p249).

Specifically, the relative efficacy of quantitative and qualitative phases has not yet been explored and is a subject of a long-standing debate that remains unanswered. Questions such as *"the field must address the relationships between qualitative mapping and quantitative modelling — in short, when to map and when to model"*

(Richardson 1999) and "*how can we measure the value added by the extra effort of simulation or the value lost by not simulating*?" (Coyle 2001), continue to haunt system dynamists.

Six specific research questions pertaining to the current research are listed below. These questions will be established in the next chapter (Chapter 2 – Literature Review).

RQ1. Do people trained in either (i) qualitative system dynamics, (ii) quantitative system dynamics or (iii) a combination of qualitative and quantitative system dynamics perform better in simple and complex tasks as compared to people not trained in system dynamics?

RQ2. Do people trained in qualitative system dynamics perform better in simple and complex tasks as compared to people trained in quantitative system dynamics?

RQ3. Do people trained in qualitative system dynamics alone perform better in simple and complex tasks as compared to people trained in both, qualitative system dynamics and quantitative system dynamics?

RQ4. Do people trained in quantitative system dynamics alone perform better in simple and complex tasks as compared to people trained in both, qualitative system dynamics and quantitative system dynamics?

RQ5. Do people trained in system dynamics retain learnt concepts for simple and complex tasks after a few months have elapsed since the initial training?

RQ6. Does re-familiarization with quantitative system dynamics software assist people in improving performance in simple and complex tasks after few months have elapsed since the initial training?

This research concludes that qualitative system dynamics tools are adequate for performance in simple stock/flow tasks. Quantification does not add significant value in these cases. For complex tasks however, quantitative system dynamics tools are necessary. This is primarily because we are unable to mentally forecast the behaviour of a complex system and are unable to incorporate and appreciate the ramifications of feedback loops.

It is also shown that after a one-off system dynamics intervention, quantitative system dynamics concepts are largely forgotten; but qualitative system dynamics concepts seem to be largely retained.

1.3 Methodology

Experimental research methodology is adopted for this research. The experimental research method is a common method applied to various fields such as experimental economics and marketing research. Moore and McCabe (1993, p202) describe the experimental method as

"*...the best method—indeed the only fully compelling method—of establishing causation is to conduct a carefully designed experiment in which the effects of possible lurking variables are controlled*".

With the exception of Doyle et al. (1998), the studies that have experimentally evaluated system dynamics efficacy have not employed a rigorous scientific method. They lack one or more essential elements such as not using a pre-test/post-test design, measuring changes of groups rather than of individuals, using small sample sizes etc. These characteristics along with the methodology and design used in this research are discussed in detail in Chapter 2 and also in empirical chapters (Chapters 3, 4 and 5).

The experimental design enables the manipulation of the degree and type of intervention (no intervention, qualitative, quantitative or both). System dynamics interventions and task complexity were used as independent variables. Performance, understanding and confidence in simple and complex tasks were used as dependent variables. Three experiments were conducted in order to answer the research questions. A randomised pre-test/ post-test design was used to measure dependent variables, both before and after system dynamics interventions. The details of the experimental design, procedure followed, tasks and method of analysis are discussed in the empirical chapters (Chapters 3, 4 and 5).

The complexity of the tasks was operationalized by varying the number of elements that contribute to dynamic complexity such as stocks and flows and feedback loops. For the simple task, a one-stock system was used with no feedback. The system in the complex task consisted of two stocks. In addition to the stocks, two feedback loops contributed towards the complexity of the task.

1.4 Summary of Empirical Chapters

As previously indicated, the literature does not address the value added by qualitative and quantitative stages, individually and combined with each other, as well as long-term system dynamics effectiveness. Therefore, the empirical chapters in this thesis are concerned with investigating these controversial areas.

The first experiment, reported in Chapter 3, examined the value added by quantification over and above of the understanding gained through the qualitative phase. It also measured the efficacy of qualitative system dynamics and combined system dynamics with respect to performance without these tools. This experiment answered research questions RQ1 (parts (i) and (iii)) and RQ3. The experiment used a repeated measures design. A set of participants were first trained in qualitative system dynamics and later in quantitative system dynamics. Their responses were measured once before the training, once after qualitative training and then finally after the quantitative training. Participants were administered a simple task (1-stock system) and a complex task (2-stock system with feedback). The same tasks were used in each of the three tests.

The second experiment, reported in Chapter 4, tested the efficacy of all the three system dynamics' interventions, individually and relative to each other. The second experiment was used to answer RQ 1, RQ 2, RQ 3 and RQ 4. Three analogous simple and complex tasks were created. At each stage of measurement, participants were given the simple and complex tasks, though with different cover stories. This ensured the removal of 'learning effect' that might influence performance due to familiarity of the task. The design of this experiment enabled us to measure the efficacy of quantitative system dynamics and compare it with qualitative and combined system dynamics.

In the first two experiments (reported in Chapters 3 and 4), participants' responses were recorded immediately after the interventions. These experiments did not measure the long-term impact of the interventions. The final experiment, reported in Chapter 5, presents a longitudinal study that tests the long-term effectiveness of system dynamics training. This experiment was conducted five months after the second experiment. Participants in the final experiment were a subset of those who participated in the second experiment. During the five-month period, participants were not involved in any system dynamics training or application. Participants were also provided time to re-familiarise themselves with system dynamics software ,which they might have forgotten. The third experiment answered research questions RQ 5 and RQ 6.

1.5 Applications of the research

The potential applications of the research findings are widespread. First, an experimental investigation will provide concrete evidence of the efficacy of the methodology in general and may confirm anecdotal claims. A rigorous study of this sort will not only clarify the extent to which each of the tools contribute to understanding, but will also help to build a strong foundation for use of system dynamics tools. Secondly, the research will contribute to the ongoing debate between the relative efficacies of qualitative versus quantitative tools. This is potentially useful for practitioners who may like/have to decide between the two tools due to time and monetary constraints. Finally, this study will provide potential insights regarding the effectiveness of system dynamics' interventions in the long-term. This research will test whether one-off system dynamics training makes fundamental changes to participants' understanding. It will also shed light on which elements of the training are retained and which are forgotten after a substantial period of time has elapsed since the initial training occurred.

1.6 Outline of this thesis

The thesis consists of six chapters including this introduction (Chapter 1). The next chapter (Chapter 2) reviews literature. It focuses on widespread applications of system dynamics and anecdotal claims of its effectiveness. The chapter then highlights the debate between qualitative and quantitative system dynamics, paucity of experimental research to test the effectiveness of system dynamics and also discusses previous work in this area. The aspect of experimental design that is common to all three experiments is also described in Chapter 2. The chapter culminates with specific research questions. The three chapters that follow Chapter 2 (Chapters 3, 4 and 5) describe the three experiments. Methodology that is specific to each experiment is discussed in the respective chapters. The first study is detailed in Chapter 3. It discusses the benefit of quantification over qualitative analysis. The next chapter (Chapter 4) describes the main experiment. This chapter answers research questions pertaining to the relative usefulness of qualitative and quantitative system dynamics. The experiment reported in Chapter 5 is the longitudinal study. The results of this study are discussed with respect to those reported in Chapter 4. The last part of the thesis contains a summary of major findings, limitations and avenues for further research (Chapter 6 - Conclusion).

Chapter 2
Literature Review

2.1 Introduction to the literature review
2.2 Foundation of system dynamics
2.3 Mental models
2.4 Components of dynamic complexity
2.5 Methodology of system dynamics
2.6 Two distinct phases: qualitative and quantitative
2.7 Recent trends in system dynamics – a review of articles published in *System Dynamics Review*
2.8 Model building versus model interaction
2.9 Decision making in a dynamically complex environment
2.10 Claims of system dynamics as being useful
2.11 Lack of empirical research in system dynamics
2.12 Previous experimental studies
2.13 Identification and relevance of gap in literature
2.14 Task complexity
2.15 Deriving research questions
2.16 Longitudinal Research
2.17 Research questions – part II
2.18 Experimental design
2.19 Participants
2.20 Contribution

2.1 Introduction to the Literature Review

System dynamics (SD) is a methodology used to study and manage complex feedback systems. System dynamics tools, both qualitative and quantitative, claim to enhance our mental models by enabling us to understand the elements of dynamic complexity, thereby improving performance in decision making tasks. System dynamics has been extensively used in the past to understand and solve complex problems. An underlying assumption of these applications is that the system dynamics process will create increased understanding of a complex situation. However, to date, claims of the usefulness of system dynamics have not been subject to rigorous experimental testing. This has been highlighted by various authors (e.g. Cavaleri and Sterman 1997; Doyle et al. 1998; Sweeney and Sterman 2000; Maani and Maharaj 2004). Furthermore, the relationship between qualitative and quantitative system dynamics tools with respect to their application and their respective usefulness to the understanding of components of dynamic complexity has been somewhat controversial (Coyle 2000). Qualitative system dynamics is characterised by the use of unquantified word-and-arrow diagrams such as causal-loop diagrams that aid in describing a system by highlighting feedback loops. Quantitative system dynamics, on the other hand, includes building quantified models that are captured in diagrams and equations and can be simulated by a computer program. Some authors claim that the use of qualitative tools alone might be sufficient (e.g. Coyle 2000; Coyle 2001) whereas others stress that the use of quantitative tools is necessary (e.g. Homer and Oliva 2001) in solving common business problems.

Pointing to the importance of evaluating system dynamics at a generic level, Cavaleri and Sterman (1997) assert:

> *"Rigorous follow-up research is essential to build a strong foundation for the refinement and wise use of the tools of system dynamics and systems thinking—a goal of both academics and practitioners". (p171)*

Furthermore, some studies have explored the effectiveness of system dynamics tools with respect to performance without these tools. However, the *relative* effectiveness of qualitative tools versus quantitative tools has not yet been explored. There has been a long standing debate on the efficacy of qualitative versus quantitative system dynamics (Howick et al. 2006; Luna-Ryes and Andersen 2003). Such back-and-forth arguments are evident in papers published in the *System Dynamics Review* in the years 2000-2001 (Coyle 2000; Coyle 2001; Homer and Oliva 2001). These debates have called for evaluation of claims that are based on anecdotal evidence alone (Richardson 1999, Coyle 2000).

Four distinct research areas are drawn together to carve out specific research questions for the current research. These are: (i) research in decision-making in the presence of dynamic complexity; (ii) the long standing debate on the efficacy of qualitative mapping versus quantitative modelling; (iii) the claims of system dynamics being a useful aid in dynamically complex situations; and (iv) the lack of experimental work on the long-term efficacy of system dynamics interventions.

Poor human understanding in situations that involve time delays, nonlinearities and feedback motivated to research in decision making in dynamically complex scenarios. This research showed that our natural ability to understand dynamic complexity is quite poor, leading to sub-optimal behaviour (e.g. Dorner 1980; Sterman 1989a; Sterman 1989b; Brehmer 1992; Kleinmutz 1993). System dynamics on the other hand claims to be a panacea for such situations. There exists anecdotal evidence of the usefulness of system dynamics through commercial applications of the methodology. As previously stated, these claims are not yet fully experimentally tested (highlighted in Cavaleri and Sterman 1997; Doyle et al. 1998; Sweeney and Sterman 2000; Maani and Maharaj 2004). Furthermore, the relative efficacy of qualitative and quantitative system dynamics tools is yet to be rigorously evaluated. This provides motivation to test the efficacy of the qualitative and quantitative phases of system dynamics method, both applied individually and when applied together. An experimental investigation is required that tests the effect of system dynamics interventions in the short-term, using a cross-sectional study and in the long-term, using a longitudinal study.

This chapter first discusses the state of the art of system dynamics. It covers the origin and history of the field; its methodology; two key phases in the methodology—the usage of qualitative tools and the usage of quantitative tools; and the pros and cons of both these tools. In the subsequent section prominent experimental studies in dynamic decision-making are discussed. The next section is devoted to a variety of anecdotal claims for the usefulness of system dynamics. The lack of rigorous controlled experimentation to test system dynamics' effectiveness is then emphasised. This is followed by a discussion of the relevant experimental studies that have in some way tested the usefulness of system dynamics training in a laboratory setting. The gap in the literature is then revealed. Specific research questions are formulated. A method to test the relative efficacy of system dynamics tools is discussed. Finally, the findings of the literature review are summarised.

The scope of current literature review does not extend to studies on 'group model building'[1] (Richardson and Andersen 1995; Vennix 1996; Vennix 1999; Rouwette et

[1] Typically, in group model building, one or more system dynamics experts works with a group of people to develop a system dynamics model. In such a process, participants develop a 'shared' understanding of the system.

al 2002). Group model building studies do not allow the measurement of individuals' understanding, which is one of the aims of this research.

2.2 Foundation of System Dynamics

In this section, the fundamental concepts of system dynamics are discussed. First, the roots of the field are described. This is followed by descriptions of the terms "system" and "systemic thinking".

2.2.1 General Systems Theory (GST) and contemporary systems approaches

The GST (sometimes referred to as "systems theory") is an interdisciplinary field to study complex systems and is the precursor of many contemporary systems approaches including system dynamics. It was developed by Hungarian biologist Ludwig von Bertalanffy in 1936 and was further developed by Ross Ashby. Bertalanffy (1950) who remarked:

> *"General Systems Theory is a logico-mathematical discipline, which is in itself purely formal, but is applicable to all sciences concerned with systems". (p139)*

GST applies to all types of systems. In one way the use of GST could unify science, as systems are ubiquitous. The concepts of 'wholeness', 'system', 'the whole is more than the sum of its parts', were all known before the theory came into being. The contribution of GST was to formalise this concept (Bertalanffy 1950). The theory proposed that the behaviour of systems cannot be determined merely by studying their parts in isolation. This is because some of the system's properties arise from the relationship between system elements and can only be understood from a holistic point of view (Ackoff 1971). Apart from system dynamics, GST has been a precursor of popular approaches like Soft System Methodology, Management Cybernetics and Critical Systems Thinking (see Appendix for a detailed representation). These popular approaches are discussed in turn. The concept of holism is prominent among all of these approaches.

Soft System Methodology is a qualitative learning methodology that was developed by Peter Checkland and his associates at Lancaster University, U.K. (Checkland 1981). It gained mass appeal as it could be used by people without technical backgrounds (Mingers and Taylor 1992). Soft System methodology enquiry can be expressed as a seven stage approach:

1. Investigation of the problem situation
2. Expressing the problem situation through Rich Pictures
3. Forming root definitions of relevant systems
4. Building conceptual models
5. Comparison of the conceptual models with the real world
6. Identify feasible and desirable changes
7. Recommendations for taking action to improve the problem situation

Oliva and Lane (1998) study the strengths and weaknesses of Soft system methodology and system dynamics. A key difference between soft system methodology and system dynamics is that the system dynamics process generally includes computer modelling and simulation, which helps in the inference of dynamic behaviour over time. They argue that both methods have much to learn from each other and suggest the use of the two techniques together. They christen it "Holon Dynamics".

Management Cybernetics was introduced by Stafford Beer (1959). It is the application of natural cybernetic laws to organisations and the interactions with and within them. Beer developed the Viable Systems Model (VSM) (Beer 1984), a model to represent the organisational structure of any autonomous system. According to VSM, a viable system requires five key systems in place for effective functioning. These are: Implementation, Co-ordination, Control, Intelligence and Policy. The VSM assists members of an organisation to take a systemic view of their communication processes.

Critical Systems Thinking is a relatively new systems methodology developed by Michael Jackson of the University of Hull, U.K. Jackson claims that critical systems thinking is growing more rapidly than any other form of systems thinking (Jackson 1991). Critical Systems thinking embraces five major commitments (Jackson 1991). These are:

1. Critical awareness
2. Social awareness
3. Complementarism at the level of methodology
4. Complementarism at the theoretical level and
5. Dedication to human emancipation

System Dynamics was developed by Jay Wright Forrester and his colleagues at the Massachusetts Institute of Technology (MIT) in the 1960s. Forrester's seminal book *Industrial Dynamics*, first published in 1961, marks the beginning of this discipline. Like general systems thinking, system dynamics also highlights the interdependencies within the system, rather than pinpointing the right solution. It reveals a variety of alternatives that could be taken. Keeping in mind that making a change to one part of the system might have an adverse effect on the other, it is up to the decision maker to recognise the tradeoffs of these actions (Senge 1996). The

System Dynamics Society, the prime body for all system dynamists, describes system dynamics as *"a methodology for studying and managing complex feedback systems, such as one finds in business and other social systems"*. They emphasise that though the word 'system' can be associated with many methodologies, system dynamics focuses on the relationship between two entities that influence each other (feedback). The underlying concept of system dynamics implementation is the concept of feedback. The theory of complex feedback systems has been developing for the past 150 years. Richardson and Pugh (1981) describe feedback as the transmission and return of information. Forrester (1961) describes the occurrence of this phenomenon as when the environment leads to a decision that results in action which affects the environment and thereby influences future decisions. Over the years, system dynamics has been widely used for solving and understanding complex management problems. One of the aims of this methodology is to create a virtual world for a management scenario through the use of modelling. Since its advent, the field has been successfully applied to various spheres including organizational settings, environment, innovation management and economics, to name a few. The successful understanding of various complex systems that eventually leads to improved decision making is an indicator of how this methodology has proven useful, especially to businesses.

In the following sub-sections, fundamental concepts that apply to all systems thinking approaches including system dynamics are discussed.

2.2.2 System

The word 'system' originates from the Greek word *sunistanai*, which originally meant "to cause and hang together". Kauffman (1980) describes a system as a collection of parts which interact with each other to function as a whole. According to Ackoff (1971),

> *"...a system is an entity which is composed of at least two elements and a relation that holds between each of its elements and at least one other element in the set".* (p662)

We come across systems in our daily life; we ourselves may be part of one or more of them. Common examples of systems are automobiles, industries and organisations. Even within a system, parts of it interact continuously. Hence everything that we see around us including ourselves is interacting as part of some system. All our actions and reactions are governed by not only what we think and do but also by what the effect of our actions is on the system that we are part of. According to Kirkwood (1998), almost everything in a business is part of one or more systems and it is impossible to isolate each and every part of an organisation and see it as a separate entity. Systems may be classified as either static or dynamic. Static systems do not change over time. For example a piece of furniture may be conceptualised as a static system. On the other hand, a system whose state changes over time is a dynamic system, for example, an automobile.

2.2.3 Systemic thinking versus linear thinking

Ison (2008) clarifies that linear thinking is related to 'step-by-step' thinking which deals with the parts of a whole, whereas systemic thinking involves 'the understanding of a phenomenon within the context of a larger whole' and this involves putting things 'into a context, to establish the nature of their relationships'. Ison and Russell (2000) contrast the epistemological basis of systemic thinking and linear thinking. They observe that in systemic thinking, the properties of the whole are said to emerge from their parts and thus systems cannot be understood merely by analysing their parts. Linear thinking, on the other hand, suggests that analysis is linear and that the entire system can be understood by step-by-step analysis of each part separately. Further, systems are characterised by feedback loops. According to Espejo (1994), systemic thinking includes more than merely learning about dynamic loops. He argues that systemic thinking consists of six distinct elements. These are:

1. How parts of a system relate to each other and constitute larger wholes
2. Understanding interactive processes in the system
3. Understanding the mechanisms that govern working of a system
4. Understanding likely effects in the whole of local behaviours and vice-versa
5. Understanding language and emotions, and
6. Understanding situational complexity.

Senge (1996) argues through one of the most popular works on systems thinking—*The Fifth Discipline*, that relationships are, in a genuine sense, more fundamental than things, and that wholes are primordial to parts. The primacy of the whole can be illustrated by thinking of any mechanical system, for example a computer. The various input and output devices of the computer cannot work in isolation. It is the relationship between these parts that makes the entire system work. According to Senge (1996), organisations are patterns of interaction rather than a 'thing'. This concept of giving importance to the system itself rather than its parts forms the fundamentals of systems thinking. On similar lines, *The Systems Thinker* newsletter observes that instead of analysing a problem in terms of an input and an output, for example, systems thinking tools look at the whole system of inputs, processes, outputs, feedback, and controls. They claim that this view provides more useful results than traditional decision making approaches.

A change in one part of a system brings about a change in the other and sometimes leads to a change in itself. This concept of having a holistic view of a complex system where one entity affects the behaviour of the other and is affected itself in turn by feedbacks has been illustrated in various disciplines including servo mechanics, cybernetics and systems thinking. Ackoff (1994) argues that the essential properties of a system can be captured and its behaviour explained when parts of a system are considered as a whole.

Systems generally are so complex that many times our well-intentioned efforts of solving a problem create unforeseen side effects. In fact, the solution to the perceived problem often causes the problem to amplify. This phenomenon, that causes an opposite effect than the one expected, has been termed 'policy resistance' (Sterman 2001). Sterman (2000) observes that one cause of policy resistance is our tendency to interpret experience as a series of events. He explains the event-oriented worldview with the following diagram (Figure 2.1).

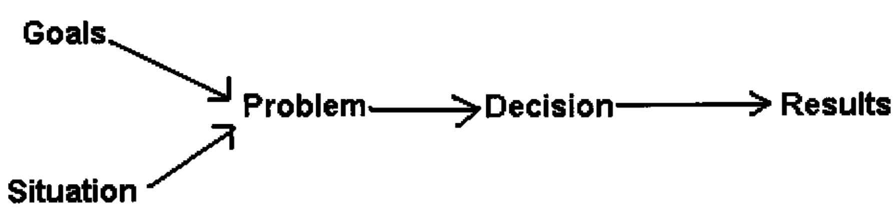


Figure 2.1: Traditional way of thinking and decision making; Sterman (2000)

Sterman further argues that policy resistance occurs because we fail to understand the full range of feedback in the system, and therefore many of our actions have unanticipated side-effects. This ignorance of feedback loops, nonlinearities and time-delays in the event driven view of causality leads to the formation of mental models that are unable to view the relationships between entities. Senge (1996) asserts that typical construction of Western languages is based on subject-verb-object structure that makes it hard to understand situations involving feedback. He describes 'the primacy of the whole' as one of the three guiding ideas for learning organisations. He observes that some of us tend to assume that parts are primary, existing somehow independent of the wholes within which they are constituted. Kirkwood (1998) describes human beings as hasty problem solvers who quickly determine the cause for any event that we think is a problem. And when we face a management problem we tend to assume that some external event caused it. Kirkwood (1998) argues that in the systems view, it is often the internal structure of the system which is more important than external events in generating the problem. Kirkwood (1998) further suggests that this tendency of finding the cause of each event might work for simple problems, but does not work for complex problems, especially those which are cross-functional or strategic in nature. It has been argued that the systems thinking approach is different from the way we generally think and there needs to be a paradigm shift in the way people think about complex problems (Kirkwood 1998). This change in thinking cannot be made easily and is dependent on the individual's perception of the problem—his/her own "mental model". At the superficial level, all problems seem to have been caused by a single event. For example, if sales of a company drop, it is blamed on an unmotivated sales force. And the cause of this cause too could be singled out as 'poor incentives' and so on. The traditional approach leads us from one cause to the other, in the process curing the symptoms of the immediate problem and temporarily alleviating the entire issue. At this stage we fail to understand that the problem lies with the structure of the system rather than entities which are part of it. In attribution theory, the fundamental attribution error (sometimes referred to as the actor/observer bias) is the tendency for humans to over-emphasise dispositional, or personality-based, explanations for behaviours observed in others while under-emphasising the role and power of situational influences on the same behaviour. In other words, people tend to have a default assumption that what a person does is based more on what "kind" of person she is, rather than the social and environmental forces at work on that person. This default assumption leads to people sometimes making erroneous explanations for behaviour (Heider 1958; Kelley 1967). As we go deeper to analyse the actual cause of the problem, we are able to find that the current event is part of a longer term 'pattern of behaviour' (Kirkwood 1998). According to Sterman (2000), one of the fundamental principles of system dynamics states that the structure of a system gives rise to its behaviour. Four common patterns of behaviour which could occur in business settings, individually or in combination with others are: exponential growth, goal-seeking, s-shaped growth and oscillation (Figure 2.2 (a) – 2.2 (d)).

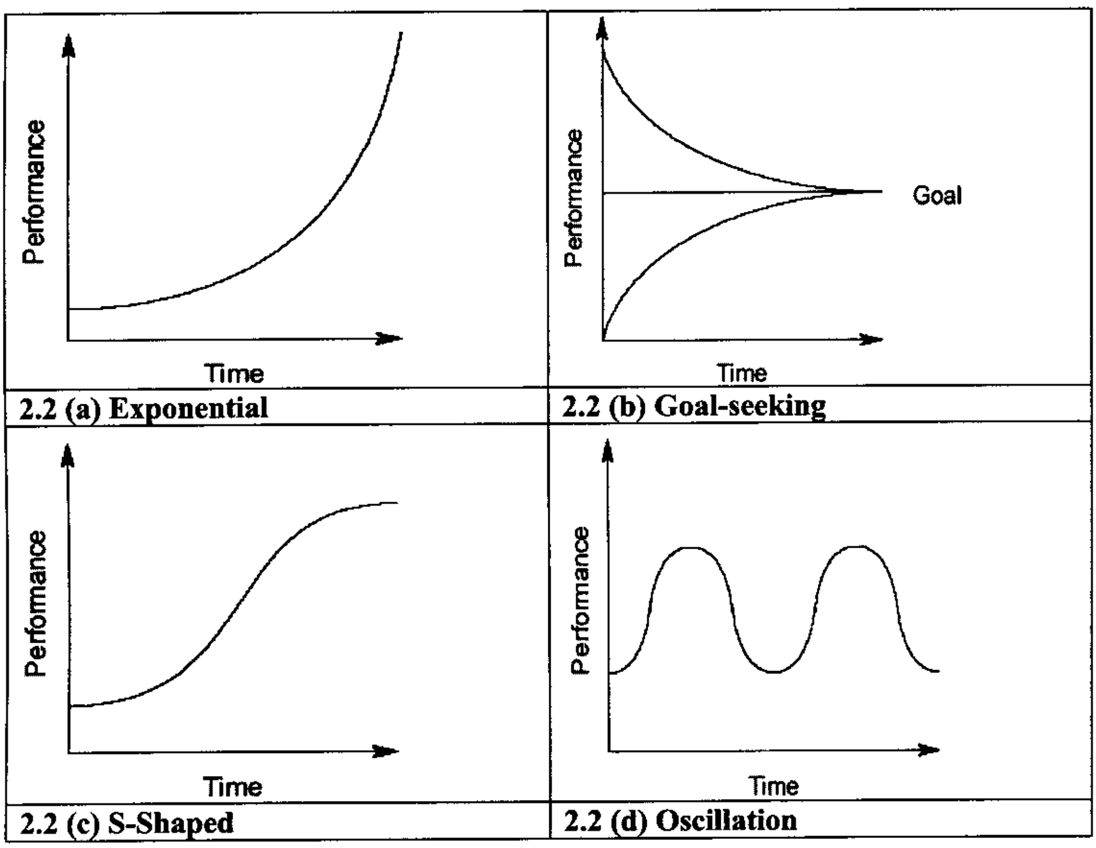


Figure 2.2: Common patterns of behaviour; source: Kirkwood (1998)

2.3 Mental models

Mental models are the psychological aspect of system dynamics. The word 'model' in psychology has been defined as any physical or chemical system which has a similar relation-structure to that of the process it imitates (Craik 1943). According to Doyle and Ford (1998), the concept of mental models has been vitally important to the field of system dynamics since its inception. Even though this concept is of prime importance there wasn't a specific definition of the term "mental model" in system dynamics literature up till 1998. The concept of mental models is abstract and many authors have defined and described mental models in different ways. This is because describing mental models is inherently difficult, as they are not directly observable and can change during procedures designed to assess them (Doyle and Ford 1998). Moreover, these definitions have been conflicting at times. Forrester (1961) described mental models as a mental image or verbal description in English and these substitute in our thinking for the real system to be represented. Later, Forrester (1971) also described mental models as fuzzy, incomplete and imprecise as the number of variables that the human mind can process at a given time is limited. Richardson and Pugh (1981) too describe mental models as fuzzy and unable to handle complexity. According to Sterman (1994), mental models are "vastly simplified compared to complexity of the systems themselves" and "dynamically deficient". Unfortunately, these mental models and decision habits are so deep-seated that they do not change just because of a logical argument (Forrester 1995). According to Senge (1990), there are two types of complexities exhibited by mental models—detailed and dynamic. Detailed complexity relates to the content of mental models whereas dynamic complexity reflects feedback thinking. System dynamics claims to change the latter. Doyle and Ford (1999) define the mental model of a dynamic system as:

"A relatively enduring and accessible but limited internal conceptual representation of an external system (historical, existing or projected) whose structure is analogous to the perceived structure of that system". (p414)

The fundamental concept behind all these definitions is that mental models are those models that we make in our own mind for a particular situation. In other words, it is what each individual perceives of a particular situation. These mental models help us choose from various alternatives, make decisions, etc. Researchers who study mental models in cognitive science and related fields agree on the shortcomings of mental models (Doyle and Ford 1998). According to Sterman (1992a), mental models are not explicit and cannot be examined by others. He stresses that mental models are ambiguous and subject to various interpretations. Sometimes people have difficulty in interpreting their own mental models. Improving mental models has been the focus of system dynamics methodology.

2.4 Components of dynamic complexity specific to system dynamics

Though there are many concepts that are common between system dynamics and other systems approaches, the two characteristics that differentiate system dynamics from the others are (i) the focus on feedback loops and (ii) the focus on stocks and flows. These are discussed below.

2.4.1 Feedback

Feedback is one of the main causes of complex behaviour in systems (Sterman 2000) and a major use of system dynamics is to identify feedback loops in a system (Wolstenholme 1990). Feedback refers to a closed loop of action and information. The patterns of behaviour of any two variables in such a closed loop are linked, each influencing, and in turn responding to, the behaviour of the other (Richardson 1991). The concept of feedback has been widely used in various fields. These include, but are not limited to, servo-mechanics, econometrics, control, corporate management, policy resistance, economic self-regulation, predator-prey interactions, homeostasis, organisational studies and policy analysis (Richardson 1991). Though the concept of feedback can be applied to understand the behaviour of almost any system, interestingly all feedback processes can be classified into two categories—reinforcing loops and balancing loops.

2.4.2 Stocks and flows

Stocks are accumulations such as inventory of goods in a warehouse, population of a city and water in a bathtub. A 'stock' is increased by inflows and decreased by outflows. Stocks and flows are ubiquitous in situations that change over time.

2.5 Methodology of system dynamics

To highlight the use of qualitative and quantitative phases in the system dynamics methodology, it is worthwhile to look at the method from the perspective of key contributors to the field. The methodology has been described slightly differently by various authors in the last fifty years. It has primarily been based on Forrester's initial findings and has been further refined by individuals' personal experiences in the field. However, the process has by and large remained constant since its inception in the early 60s. It involves identification of the problem, describing the system, creating and testing a simulation model and then devising policies based on the outcome of the model. What has changed is the emphasis on various stages and the use of tools to achieve desired outcomes. For instance, of late there has been widespread usage of qualitative tools for system description. Also, the emphasis on formal model validation is greater than earlier before (e.g. Forrester and Senge 1980; Peterson and Eberlein 1994).

Forrester (1961) laid out a ten step approach to enterprise design in his seminal book *Industrial Dynamics:*

1. Identifying the problem.
2. Isolating the factors that appear to interact (to create the observed symptoms).
3. Tracing the cause-and-effect information-feedback loops that link decisions to action to resulting changes and to new decisions.
4. Formulating acceptable formal decision policies that describe how the decisions result from the available information streams.
5. Constructing a mathematical model of the decision policies, information sciences, and interactions of the systems components.
6. Generating the behaviour through time of the system as described by the model.
7. Comparing results against all pertinent available knowledge about the actual system.
8. Revising the model until it is acceptable as a representation of the actual system.
9. Redesigning, within the model, the organisational relationships and policies which can be altered in the actual system to find the changes which improve system behaviour.
10. Altering the real system in the directions that model experimentation indicates.

The above approach was based on various principles. That include the theory of information-feedback; the possibility of model experimentation due to the advancement in computing; the availability of sufficient information for model-building without great expense and delay; industrial systems structured in such a fashion that they themselves create problems, which are most of the time treated as exogenous; a change in the structure of the system that can bring about long lasting change and can solve problems from their roots. As pointed out earlier, Forrester's work does not include the use of any causal loop diagrams which are native to qualitative system dynamics.

Morecroft (1985) described a two-phase process for system dynamics modelling for business policy and strategy. These steps involve first, 'business structure analysis and then 'simulation modelling'. Morecroft's approach was later extended by Lyneis (1999) to a four-phase approach. The four phases were:

1. Business structure analysis,
2. Development of a small, insight-based model
3. Development of a detailed, calibrated model
4. On-going strategy management

Both Morecorfts' and Lyeneis' approach emphasises on model building, rather than the use of qualitative tools.

The System Dynamics Society (Society Website, accessed 1/1/2005) defines system dynamics as a six step process, that involves:

1. Identification of a problem
2. Developing a dynamic hypothesis explaining the cause of the problem
3. Building a computer simulation model of the system at the root of the problem
4. Testing the model to be certain that it reproduces the behaviour seen in the real world
5. Devising and testing in the model alternative policies that alleviate the problem, and

6. Implementing the solution

Luna-Ryes and Andersen (2003) compare system dynamics methodology proposed by four authors—Richardson and Pugh (1981), Roberts et al. (1983), Wolstenholme (1990) and Sterman (2000). They assert that even though the number of steps in the methodologies put forward by the four authors varies, core activities involved in these steps are constant. Luna-Ryes and Andersen (2003) conclude that the common feature among all these methodologies is an iterative process of the development of a dynamic hypothesis based on feedback theory that generates different behaviours over time. This process improves people's understanding of the system that eventually leads to the design or redesign of policies. The above analysis reveals that the core system dynamics method has remained the same. It includes first understanding the problem, then visually describing the system followed by constructing a simulation model to test various alternative solutions.

2.6 Two distinct phases: qualitative and quantitative

This section defines and describes the stages of the system dynamics methodology. The methodology can be divided into two distinct phases. The first includes understanding the system by using qualitative tools such as causal loop diagrams. The qualitative phase is generally followed by a quantitative computer modelling phase. The terms, qualitative system dynamics and quantitative system dynamics do not have unambiguous definitions. Many authors see them as overlapping. For the context of this study we adopt the conventions used by Richardson (2001). Qualitative system dynamics is defined as the use of "*unquantified word-and-arrow diagrams—causal-loop diagrams, influence diagrams, maybe with representations of stocks*". Qualitative system dynamics is also referred to as 'mapping' or as systems thinking. These definitions are widely accepted within the system dynamics community and also used elsewhere (Lyneis 1999). Quantitative system dynamics, on the other hand, includes building "*formal, quantified representations, captured in diagrams and equations, which can be simulated by computer*". Quantitative system dynamics is also referred to simply as 'modelling'. Though many approaches/tools have been used to apply qualitative and quantitative system dynamics, causal loop diagramming for the former and stock and flow modelling for the latter have been the most popular (Scholl 1995; Lane 2000).

In order to understand the origin of the debate between qualitative and quantitative system dynamics, it is important to shed light on the use of these tools from a historical perspective. System dynamics was mainly perceived and used as a quantitative simulation technique when it was first introduced in 1961 (Coyle 2000). For instance, the first work on system dynamics in book form by J.W. Forrester (Forrester 1961) did not contain any causal loop diagrams and feedback structure was depicted by equations or stock-and-flow diagrams (Lane 2000). Quantitative system dynamics allows us to map a system's behaviour over time by allowing computer-based simulation. Its obvious benefit is that it combines human thinking and computational power to allow for a significant extension to qualitative system dynamics. Limitations of the quantitative method include not having enough or valid data (Wolstenholme 1999), incorporating soft variables (Coyle 2000), generation of misleading results owing to uncertain variables (Coyle 2000), generating more complexity than required, and the necessity for expert use of such a system (Wolstenholme 1999).

For the first time Wolstenholme (1982) presented a revised perception of system dynamics, shifting from a purely quantitative perspective to a methodology with a mix of qualitative and quantitative methods. Qualitative system dynamics tools have often been used to describe a system before formulating a simulation model. They provide a foundation for constructing and understanding formal simulation models. It has been well established that causal loop diagrams, one of the tools of qualitative system dynamics, may facilitate the way in which we can externalise mental models by introducing circular causality (Richmond 1997). They provide a way to simulate mental maps, modes of behaviour, and our assessment of a changed system (Wolstenholme 1999). In turn, mental models benefit from the output of formal modelling that provides for enhanced causal loop diagrams. Hence, the use of qualitative and quantitative methods for system dynamics may be perceived as iterative in nature, one benefiting from the other and vice-versa. The approach of combining the two methodologies is most common in the field today with most authors incorporating both to describe the complete system dynamics process (e.g. Wolstenholme 1983; Wolstenholme 1990; Maani and Cavana 2000). Howick et al. (2006) demonstrated that a combination of qualitative modelling tools with a simulated system dynamics model produces additional insights for clients.

Although the use of a combination of qualitative system dynamics and quantitative system dynamics is popular for inferring the behaviour of a system through its feedback structure, the last twenty-five years have also seen enthusiastic support for the use of qualitative system dynamics alone. Some authors believe that qualitative system dynamics may be sufficient for solving problems of dynamic complexity (e.g. Wolstenholme 1999). Further, Wolstenholme and Coyle (1983) argue that there could be value in simply using rigorous qualitative approaches to facilitate system description. Some of the significant studies using qualitative system dynamics alone include those by Wolstenholme (1983) and Cavana et al. (2004). Another example of a popular work on qualitative system dynamics that has become popular even outside the system dynamics community is the book by Peter Senge – *The Fifth Discipline* (Senge 1990).

Despite the fact that qualitative system dynamics alone is gaining popularity, many experts have doubted its efficacy and its adequacy in dealing with complex problems (e.g. Homer and Oliva 2001). Richardson (1996) acknowledges an increase in the use of qualitative system dynamics tools alone, however, he expresses concern over the usage of these tools by those who lack the knowledge of quantitative modelling. It has also been argued that causal loop diagrams may not make any distinction between information links and rate-to-level links, and the standard characterisations of positive and negative polarities may be wrongly interpreted (Richardson 1986/1976). In his presidential address at the International System Dynamics conference, Richardson (1997) warned practitioners that stretching the use of qualitative tools beyond their capability might be dangerous as it might produce flawed results. Taking a balanced view, he also points towards the shortcomings of the quantitative phase. Warren (2004) argues that qualitative tools like causal loop diagrams have been responsible for the lack of use of system dynamics amongst management.

The following section sheds more light on the use of qualitative system dynamics in the industry when compared with the use of quantitative system dynamics only and with studies that combine both qualitative and quantitative tools.

2.7 Recent trends in system dynamics: a review of articles published in **System Dynamics Review** *from 2002 to 2006 (inclusive)*

As previously discussed, the frequency of usage and the type of tools for quantitative, qualitative and combined system dynamics has changed over the years. A review of all studies published in the *System Dynamics Review* (SDR) during the last five years (2002 to 2006; inclusive) was conducted in order to get a clearer picture of these trends in recent times. The reason for choosing SDR is that it publishes the greatest number of system dynamics related articles in a year (Tignor 2007). It is also cited as the primary source of information about system dynamics (Scholl 1995). This analysis was done to learn about recent trends in system dynamics applications. Other relevant literature beyond these five years and from other journals has been included in other sections of this chapter.

The focus of the analysis was:

1. To analyse the number of studies that applied (i) qualitative system dynamics alone (ii) quantitative system dynamics alone and (iii) both qualitative system dynamics and quantitative system dynamics.
2. To analyse the type of tools used to apply these (e.g. causal loop diagrams, archetypes etc.)
3. To learn about the nature of industries where system dynamics is currently being applied

Out of the 106 items published in the journal in this period, 60 were found to be relevant to this research (Figure 2.3). Those eliminated include editorials, book reviews, republished papers, notes and insights and other works not deemed directly relevant to the current research. These constituted 43% of published works in SDR. The 60 publications were further classified into those that reported application of system dynamics tools to an industry problem; those that reported experimental studies and those that reported ways to improve system dynamics methodology and literature reviews (Figure 2.3).

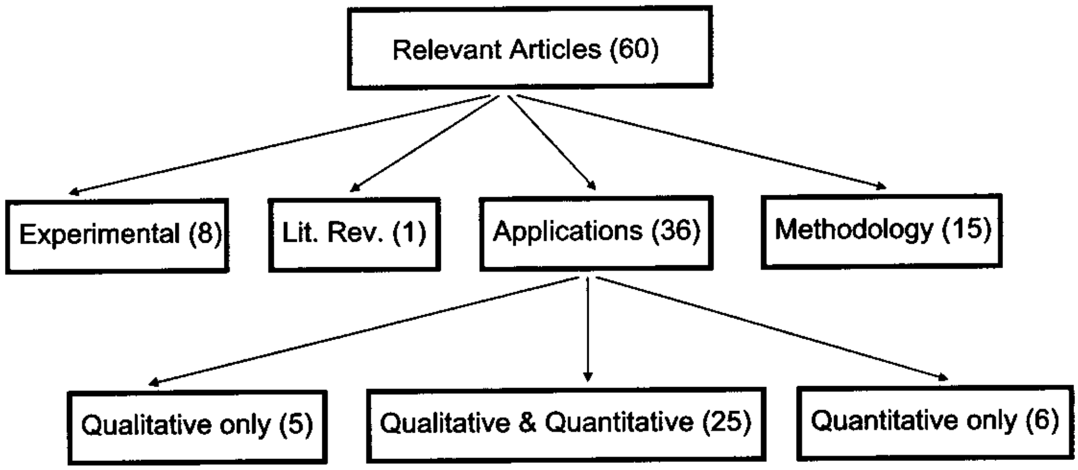


Figure 2.3: Articles from *System Dynamics Review*, 2002-2006 (inclusive) classified by type of work and type of tools used

The analysis reveals that the majority (60%) of the articles published in the last 5 years pertain to applications in industry. 15% are works that deal with the experimental evaluation of the system dynamics method in some way. The articles that discuss laboratory experiments are discussed in detail in later sections. A quarter of all articles discuss ways to improve the system dynamics methodology. One meta-literature review was published in this period.

Application-based articles were further classified on the basis of the kind of system dynamics tools used—qualitative, quantitative or both (Table 2.1). Those that made references to and/or used causal loop diagrams or other similar tools such as archetypes or influence diagrams, were classified under 'qualitative only'. Those studies that used computer modelling and simulation tools such as Vensim, iThink, Powersim etc., and did not report the use of qualitative tools, were classified under 'quantitative system dynamics only'. Those studies that used the combination of qualitative and quantitative tools were categorised under 'qualitative and quantitative'.

Table 2.1: Articles published in *System Dynamics Review* (sorted by date of publication)

	Author	SD Type	Evidence	Industry Sector
1	Cavana and Clifford (2006)	Both	CLD and SFD	Customs service
2	Friedman (2006)	Both	CLD and SFD	Urban dynamics/traffic
3	Bayer and Gann (2006)	Both	CLD and SFD	Project Management/workload fluctuations
4	Cooke and Rohleder (2006)	Both	CLD and SFD	Organization/disaster management
5	Saysel and Barlas (2006)	Both	CLD and SFD	Agriculture/Irrigation
6	Salge and Milling (2006)	Both	CLD and SFD	Nuclear reactor
7	BenDor and Metcalf (2006)	Both	CLD and SFD	Biology/parasite on trees
8	Taylor and Ford (2006)	Both	CLD and SFD	Project Management
9	Arquitt et al. (2005)	Both	CLD and SFD	Aquaculture/Shrimp
10	Gonçalves et al. (2005)	Both	CLD and SFD	Supply chain/high tech
11	Anderson Jr et al. (2005)	Quant only	SFD	Supply chain
12	Strohhecker (2005)	Both	CLD and SFD	Banking/short-term problems
13	Lee et al. (2005)	Qual only	CLD	Urban dynamics/construction
14	Otto and Struben (2004)	Both	CLD and SFD	Fisheries/feasibility of a new project
15	Homer et al. (2004)	Quant only	SFD	Health/chronic illness
16	Cavana and Mares (2004)	Qual only	CLD	Customs service
17	Pavlov and Saeed (2004)	Both	Archetype/CLD and SFD	Internet/P2P technology
18	Dudley (2004)	Both	CLD and SFD	Natural Resources/Forest
19	Fernández and Selma (2004	Both	CLD and SFD	Agriculture/Irrigation
20	Faust (2004)	Quant only	SFD	Population/endangered species
21	Arquitt and Johnstone (200	Both	CLD and SFD	Environment/Algae
22	Liddell and Powell (2004)	Both	QP influence diagrams	Health/Patient access to healthcare workers
23	Georgantzas (2003)	Qual only	SFD	Tourism/hotel value chain
24	Park and Peña-Mora (2003)	Both	CLD and SFD	Construction/Roadways
25	Oliva et al. (2003)	Both	CLD and SFD	E-commerce/Amazon.com
26	Cooke (2003)	Both	CLD and SFD	Industrial safety/mining
27	Pejic-Bach (2003)	Both	Archetype/CLD and SFD	Small business/financial indiscipline
28	Abdel-Hamid (2002)	Quant only	SFD	Health/Obesity
29	Trost (2002)	Both	CLD and SFD	Engineering/ship building/quality
30	Bianchi (2002)	Both	CLD and SFD	SME
31	Weber and Schwaninger (20	Both	CLD and SFD	Agriculture/trade organizations
32	Bianchi and Bivona (2002)	Quant only	SFD	E-commerce
33	Stave (2002)	Both	CLD and SFD	Transportation
34	Jones (2002)	Qual only	CLD	Natural Resources/Forest
35	Fiddaman (2002)	Qual only	CLD	Energy/Climate Change
36	Piattelli et al. (2002)	Quant only	SFD	Transportation

CLD: Causal Loop Diagrams – Qualitative system dynamics
SFD: Stock and Flow Diagrams – Quantitative system dynamics

Though system dynamics started as a purely quantitative methodology, Table 2.1 reveals that in recent years, the trend has been different. Only 6 (17%) studies report the usage of quantitative tools alone. The results reveal that qualitative tools have now become an integral part of the system dynamics methodology. These are generally used in conjunction with quantitative tools. Out of the 36 application-based articles evaluated, the majority used the combined system dynamics approach (69%).

Interestingly, qualitative system dynamics alone was used in 5 (14%) studies. The increase in the usage of qualitative system dynamics tools, alone and otherwise, has been acknowledged in the literature (see Richardson 1999).

Another perspective is to look at these application-based articles from the point of view of how the benefit of system dynamics was evaluated. Surprisingly, none of these studies used a rigorous method to evaluate the changes affected by the system dynamics process. Most studies primarily report through a case study, the successful implementation of a system dynamics model to a given problem.

The review of these studies also lets us explore the range of disciplines system dynamics has been applied to in the recent past. System dynamics, initially christened as "Industrial Dynamics", started with applications in industrial engineering. Such applications are still common (e.g. Trost 2002). Traditionally, practitioners have applied system dynamics to business and corporate policy and also to public and social policy (Scholl 1995). As the methodology gained momentum in the 60s and 70s, it started being used to understand complex systems in almost all disciplines. The interdisciplinary nature of system dynamics modelling draws together academics and practitioners from various fields such as management, engineering, information systems and economics. Further, system dynamics models can be applied to understand complex feedback processes in practically any domain. Forrester (1991) argues that system dynamics "provides a common foundation that can be applied wherever we want to understand and influence how things change through time". The above mentioned factors seem to have contributed to the widespread usage of system dynamics. Table 2.1 reveals similar trends in the recent past. Project management has been a typical example where system dynamics has been applied in the past. In recent years, studies by Bayer and Gann (2006) and Taylor and Ford (2006) have used system dynamics for such applications. System dynamics also finds applications in understanding technology, for example in understanding peer-to-peer networks (Pavlov and Saeed 2004) and e-commerce (Oliva et al. 2003; Bianchi and Bivona 2002). Urban dynamics has historically been a key area of system dynamics application since Forrester's seminal work in this area (Forrester 1969). From 2002 to 2006, various studies dealing with urbanisation issues such as transportation (Piattelli et al. 2002; Stave 2002), construction (Lee et al. 2005; Park and Peña-Mora 2003) and traffic (Friedman 2006) have been prevalent. Apart from applications in basic science and engineering, system dynamics has also been applied to aid in the understanding of natural resource and environmental issues. Common examples in the recent past include those of irrigation (Saysel and Barlas 2006; Martínez-Fernández and Selma 2004), aquaculture (Arquitt et al. 2005), fisheries (Otto and Struben 2004); algae (Arquitt and Johnstone 2004), forests (Jones et. al 2002; Dudley 2004) and trade organisations (Weber and Schwaninger 2002). Applications in Management have been diverse and range from banking (Strohhecker 2005), finance (Pejic-Bach 2003) and supply chains (Gonçalves et al. 2005; Anderson Jr. et al. 2005), to disaster management (Cooke and Rohleder 2006) and small- and medium-sized enterprises (Bianchi 2002). Other applications of system dynamics include healthcare, such as modelling of chronic illnesses (Homer et al. 2004), patient access to healthcare workers (Liddell and Powell 2004) and obesity (Abdel-Hamid 2002). As can be seen, the range of application areas of system dynamics continues to be quite diverse.

In summary, the analysis has revealed that qualitative and quantitative tools are generally used together in majority of system dynamics applications. The usage of qualitative or quantitative tools individually is scarce. Hence, the applications indicate towards greater benefit when the two tools are used together than when used individually.

2.8 Model building versus model interaction

Computer simulations that support learning in complex systems can be broadly classified under two categories—modelling-oriented simulations and gaming-oriented simulations (Maier and Grobler 2000). The former relates to model building and the later focuses on model interaction. It is argued that the system dynamics method pertains to model-building rather than model exploration and hence efficacy of the system dynamics method should involve active model-building.

Those relevant to system dynamics within the model-building category are various commercially available modelling environments such as Powersim, iThink, Vensim and others. These environments facilitate learning by *building* simulation models. In the context of system dynamics, this implies building qualitative models such as causal loop diagrams and/or building quantitative simulation models using stocks and flows. Gaming-oriented simulations on the other hand support learning by *interacting* with an existing simulation. These include simulation games and are popularly known as Management Flight Simulators (MFSs), Interactive Learning Environments (ILEs) and Microworlds.

Traditionally MFSs have presented a black-box view of the complex system. Players are able to see only inputs and outputs but not the underlying model. They make numerical decisions based on some initial information provided to them. These parameters are then simulated by the underlying simulation model. The output of the model is then presented back to the user. The player relies on the result of the simulation in the form of numbers and/or graphs (outcome feedback). The player learns about the system from this outcome and makes subsequent decisions. Davidsen (2000) asserts that the debate on the usage of pre-made models as a substitute to model building has been going on since the 80s. Historically system dynamics has been associated with model building rather than using a pre-made model. A closer look at this debate reveals that there are more than just these two variants. In between these two extremes of model-building and black box models are those MFSs which reveal the underlying structure of the system. In such cases, users still make their decisions based on a pre-made model, but they have access to part of or the entire underlying model. This could be achieved by exposing the stock and flow structure, causal loop diagram or the equations representing the system. In some cases verbal explanation of the model and visualisation of the interaction between the elements of the model may also be releaved (Alessi 2000). Whether one, all or combinations of the above mentioned are exposed to the user, varies from one MFS to the other. Hence, the degree of transparency depends on the needs and objectives of a situation (Alessi 2000). However, different ways to reveal the underlying model also presents some challenges (Davidsen 2000). For instance, it is not known which representations work better than the others; and whether combinations of these representations are more effective than when presented individually (Davidsen 2000).

The experiments designed in the current research deal with model-building as it is seen as a core component of the system dynamics methodology. Model-building in the context of the current research not only includes quantitative stock/flow computer models but also both qualitative models such as causal loop diagrams.

2.9 Decision making in a dynamically complex environment

2.9.1 Introduction

System dynamics tries to model situations with dynamic complexity. According to Sterman (2000), there are various reasons for this complexity. He observes that complexity arises as systems are: constantly changing, tightly coupled, governed by feedback, nonlinear, history-dependent, self-organising, adaptive, characterised by trade-offs, counterintuitive, and policy resistant. This dynamic complexity could arise even in simple systems with low combinatorial complexity. Hence, researchers have been fascinated to investigate whether the human mind is capable of understanding tasks from a systems perspective. It is believed that this human inability is evident especially when solving dynamically complex problems. In fact research conducted over several years shows that we take an event-oriented approach rather than incorporating feedback loops. For example, findings of Axelrod (1976), Hall (1976) and Dorner (1980, 1996), among others, confirm evidence of event-based decision making in such scenarios. Further, studies conducted by Wagenaar and Sagaria (1975) and Wagenaar and Timmers (1978) show that people significantly underestimate exponential growth and tend to extrapolate linearly. These authors have argued that the 'systemic view' of the problem has been absent from peoples' analysis. A number of studies have been undertaken to examine people's ability to understand and perform in tasks that involve dynamic complexity (e.g. Dorner 1980; Sterman 1989a; Brehmer 1992; Kleinmutz 1993). In this section, key dynamic decision making studies from the field of system dynamics have been evaluated. An evaluation of these studies reveals that there is an overall consensus on the fact that our ability to understand dynamic complexity is poor. Furthermore, system dynamics has been proposed as an antidote for this deficiency. In the context of this research, the studies discussed in this section reveal elements of dynamic complexity that are poorly understood. The analysis also reveals potential dependent and independent variables for the experimental design.

A summary of these studies is presented in Table 2.2. The analysis reveals that three main instruments have been used to test people's decision-making abilities. Some studies have used static tasks such as "bathtub dynamics tasks" (e.g. Dangerfield and

Roberts 1995; Sweeney and Sterman 2000; Cronin and Gonzalez 2007, Cronin et al. 2007; Sweeney and Sterman 2007). Secondly, "The beer distribution game" is a classic example that has been used by studies such as Sterman 1989b and Croson and Donohue 2006. Thirdly, several studies have used simulation games (such as management flight simulators) are discussed (Gary and Wood 2005; Moxnes 1998a; 1998b; Jensen and Brehmer 2003; Diehl and Sterman 1995; Paich and Sterman 1993).

Table 2.2: Dynamic decision-making experiments (sorted by date of publication)

Citation →	Sterman (1989a)	Sterman (1989b)	Paich and Sterman (1993)	Diehl and Sterman (1995)	Dangerfield and Roberts (1995)	Moxnes (1998a)	Moxnes (1998b)	Sweeny and Sterman (2000)	Jensen and Brehmer (2003)	Croson and Donohue (2006)	Cronin et al. (2007) - from web	Cronin and Gonzalez (2007)	Assuad and Moxnes (2007-forthcoming)
Characteristics													
Task Description	STRATEGEM-2 MFS	Beer Game	Market strategy MFS	Stock management task	Making a decision on adoption of one of the technologies based on four demand scenarios	MFS to manage virginfish stock	Predator-prey model for reindeer and lichen MFS	Bathtub tasks	Predator-prey model for rabbits and foxes MFS	Beer distribution game - computer-based	Bathtub task (department store and 2 new cover stories)	Bathtub tasks	MFS for making decisions for ways to decrease Greenhouse Gas Emissions
Dependent Variable(s)	1. Task performance (cost)	1. Task performance (cost)	Task performance (profit)	1. Task performance (cost) 2. Efforts for decision making (amount of decision time)	1. Task performance (cumulative profit over 12 years)	1. Task performance (Maximise infinite horizon profit)	1. Task performance (Maximise infinite horizon income)	1. Task performance (drawing correct BOT)	1. To make the system reach equilibrium	1. Order of beers with upstream supplier	1. Task performance (identifying inflows and outflows, time period of largest value of stock)	1. Task performance (identifying inflows and outflows, time period of largest value of stock)	1. Task performance (reduction of gas emissions)
Key Findings	1. Average performance was sub optimal 2. Participants misunderstood delays 3. Participants misperceived feedback loops	Misperception of feedbacks (MOF) - inability to manage dynamic tasks due to failure to identify positive feedback loops	1. Confirm MOF hypothesis 2. Task performance deteriorated with time delays	1. Confirm MOF hypothesis 2. Task performance deteriorated with feedback and delays 3. Time spent on tasks did not vary with task complexity	1. Performance is poor both with and without feedback 2. People cope with feedback by reducing the problem to one without feedback 3. In the absence of a model people have difficulties in drawing future system behaviour	1. Participants overinvested and this led to overutilization of fishing fleets even in the absence of commons issue. 2. Confirm misperceptions of feedback	1. Subjects capture only 9% of potential horizon 2. No participant combined appropriate dynamic mental model with appropriate analysis. 3. Confirms MOF hypothesis	1. Inability to respond properly to feedbacks 2. Poor understanding of delays 3. Confusion about definition of stocks and flows and the relationship between the two	1. Only half participants could find the optimal solution. 2. None applied an idea feedback approach	1. Bullwhip effect exists when retail demand is stationary and commonly known 2. occurs because participants underweight the supply line 3. Bullwhip exists, underweighting exists even when inventory information is shown 4. upstream members benefit more with shared info	1. Repeat bathtub - get similar results as prior studies 2. Performance did not improve on a simpler version of the task with fewer data points; when the presentation was in text or graph or text; context of the cover story was changed; increased motivation and feedback; with cues 3. Signs of overconfidence	1. Neither effort nor the representation suggested by the cover story had any effect on the correct perception of stocks and flows 2. Having fewer points or a distrator on the graph has no effect on performance 3. People focus at difference of flows at a single point in time	1. The median tax was significantly lower than the optimal tax 2. This is possible as subjects underestimated time delays and 3. Their inability to respond properly to feedback over time

2.9.2 Dynamic decision-making studies that involve static environments

As previously mentioned, static environments are those in which information about the task is provided to the decision maker and the decision maker is required to make just one decision. An example of a study that tested the effect of feedback complexity on performance within a static environment was conducted by Dangerfield and Roberts (1995). In this study, participants were supposed to make a capital investment decision. Four demand scenarios were used. The task performance was measured as cumulative profit using two questionnaires. The first condition did not involve any feedback effect on demand while the second involved feedback. Results showed that participants had difficulty in understanding cyclical demand when compared to linear demand. In fact performance was even worse in the situation involving feedback. The authors infer that this was due to the fact that participants coped with feedback by reducing the situation to one without feedback. The authors assert that in the absence of a formal model, people find it difficult to forecast the behaviour of a system. They suggest that a questionnaire with feedback could be used in a pre-test/post-test design where participants are trained in system dynamics between tests. Such an arrangement could then test the effectiveness of model building to forecasting system behaviour.

In a highly influential study, Sweeney and Sterman (2000) demonstrated that people's native ability to understand stocks, flows and the relationships between them, is poor. Using a battery of two tasks, popularly known as the bathtub dynamics task, they evaluated performance of students at MIT. The tasks provided a means to test fundamental system dynamics skills such as understanding of stocks and flows, feedback loops and time delays. The first task—bath tub/cash flow, was used to test participants' understanding of stock and flow relationships. The task consists of two analogous cover stories—bath tub condition and cash flow condition. Inflow and outflow were provided on a graph. Participants were required to draw behaviour of the stock over time. The system described was simple—as it consisted of only one stock, one inflow, one outflow and no feedback. Two different (square wave and saw toothed) inflow and outflow patterns were used. The second task— a manufacturing case study, was used to test participants' understanding of stocks and flows in the presence of time delays and feedback loop. Participants were asked to draw the behaviour over time. In general, the participants' performance in these tasks was poor. In the bath tub/ cash flow task, participants' responses reflected their poor understanding of basic principles of accumulation. They did not understand the definitions of stocks and flows and the relation between the two. Performance regarding time delays and feedback was even worse. In fact, several other researchers have since reported similar findings on the bathtub tasks (e.g. Ossmitz 2002; Quaden and Ticotsky 2003; Heinbokel and Potash 2003; Kubanek 2003; Fisher 2003; Zaraza 2003 and Lyneis and Lyneis 2003). Bathtub tasks provide a great tool for assessing peoples' understanding of a simple dynamic system—with one stock, one inflow and one outflow. In particular, Sweeney and Sterman (2000) claim that these tasks could be used to "measure the impact of various types of systems thinking training".

Recently, Cronin and Gonzalez (2007) made an attempt to investigate the causes of cognitive deficiency that lead to the misperception of stocks and flows. The authors argued that, in previous studies, the cover story might not have been conducive to form a correct representation of the problem. Their first experiment was used to test whether people can interpret graphs of dynamic systems, when the system represented in bathtub tasks is commonly understood in terms of stocks and flows. Since understanding the system might not be sufficient to arrive at the correct answer, participants also needed to apply the knowledge to the graph in order to obtain the highest level of stock. The authors assert that, in previous studies, participants may not have put in an effort to solve the task. They hypothesise that an increased thinking effort will improve performance. Hence, the study used a 2X2 experimental design (varying familiarity of the cover story and thinking effort). A more familiar cover story was designed using the context of a bank account. The unfamiliar story was the original department store task (Sweeney and Sterman 2000). In the low effort condition, participants were paid $5 for the task. In the high effort condition, in addition to being paid $5, participants were told that their answers would be personally inspected by the researchers and would be used to compare their performance with top-ranked universities. The results show that neither a different cover story nor putting more effort into thinking resulted in higher performance. It should be noted that different cover stories had been previously used by some authors yielding similar results (e.g. Kainz and Ossimitz 2002). Also, in some previous studies, participants were examined on these tasks as part of university courses (e.g. Pala and Vennix 2005). Thus the incentive in these cases was even higher than that provided by Cronin and Gonzalez (2007). Even in situations with a greater incentive, performance was sub-optimal. In their second experiment, Cronin and Gonzalez (2007) tested the influence of a graphical representation of the problem to task performance. They hypothesised that the number of data points on the graph was directly proportional to the cognitive load. A higher cognitive load might have led to poor performance. They also suggested that a graph that might mislead participants to identify the incorrect response (for e.g. with a distractor point such as peak) would result in lower performance than a simpler graph. Using again a 2X2 design (varying data point of the graph and the shape of the graph) the authors demonstrated that neither of the two factors influences performance. In the third experiment, the authors explored the most common mistake participants make. They found that the majority of the incorrect respondents choose the point based on the momentary difference in the net flow. Overall, the authors found that the graphical representation of the task has a significant impact on the way people interpret the system.

In a similar study to the above, Cronin et al. (2007) conducted a series of five experiments to test the influence of various factors on performance in the bathtub task. Their first experiment evaluated the effect of cognitive burden and data display on task performance. Their findings suggest that having fewer data points did not improve performance. These results corroborate those of the study reported above (Cronin and Gonzalez 2007). In addition, the results also suggest that performance did not vary when participants were provided with a task data bar or a line graph. Further the format of the task (tabular, graphical or textual) did not influence task performance. These results contradict those obtained by Kainz and Ossimitz (2002), who found that tabular representation leads to higher performance when compared to graphical. The second experiment evaluated the content of the cover story. Results show that a more familiar cover story did not make any difference in task performance. These results reinforce previous findings by Cronin and Gonzalez, 2007. Furthermore, using continuous rather than discrete quantities of data made no difference. The authors then examined the effect of motivation and feedback on performance in these tasks (third experiment). Participants in the experiment group were supposed to bring their responses to the instructor, who indicated whether the answer was correct or incorrect (feedback). Participants with incorrect responses then had to go back and think about the question again. They could leave the test when they achieved the correct response (motivation). There were no significant differences found when results of the experiment group and control group were compared. Using a different form of motivation, Cronin and Gonzalez (2007) reported similar results on the same task. In the fourth experiment, instead of providing inflows and outflows and inferring the value of stock, participants were provided with the value of the stock and one of the flows. The findings suggest that task performance was similar to those reported for other experiments. In the fifth experiment, the authors report that even providing cues that might encourage people to notice the presence and behaviour of stock and flows did not improve task performance on the stock/flow task. Taken together, the results of these experiments show that an understanding of stocks and flows is more fundamental than previously thought. The authors compare these tasks with insight problems—which are analytically easy once a proper frame of use is recognised.

Bathtub tasks have also been used to show misconception of stock and flow principles. For example Sweeny and Sterman 2007's experiment on climate change evaluated people's understanding of the fundamental relationship between the flows of Greenhouse Gases (GHG) and its stock in the atmosphere. Their results indicate that participants misunderstand stocks and flows and mass-balance relationships. Further the authors assert that although improving these misconceptions about GHGs requires knowledge of climate change as well as of stocks and flows, it is the latter that is more fundamental to the understanding of such dynamic scenarios.

On one hand, bathtub tasks might seem simple, once one knows the answer. On the other hand, however, the studies above point towards the misconceptions people have of the definition of stocks and flows and the relationship between the two. These studies have tried numerous variations of the original bathtub tasks in order to observe changes in task performance. None of these variations have produced encouraging results, leading to the conclusion that such misconceptions are fundamentally difficult to solve. This indicates that training people on the concepts of stocks and flows might compensate for this deficiency.

To summarise, this section has showed that decision-making abilities in the presence of dynamic complexity are poor, even when all information is provided and participants have to make just one decision. The following two sub-sections discuss studies that use multiple decision environments.

2.9.3 Dynamic decision-making studies that use the beer game

The beer distribution game is a classic example that shows people's poor understanding of dynamic complexity. It is essentially a typical production and distribution system of goods (beer). The role-playing game was developed by the System Dynamics Group at Massachusetts Institute of Technology in the early 1960s. Since then it has been played by thousands of people. The game is generally played on a board. Lately, computerised versions of the game have been available. The figure below shows the layout of the game (Figure 2.4).

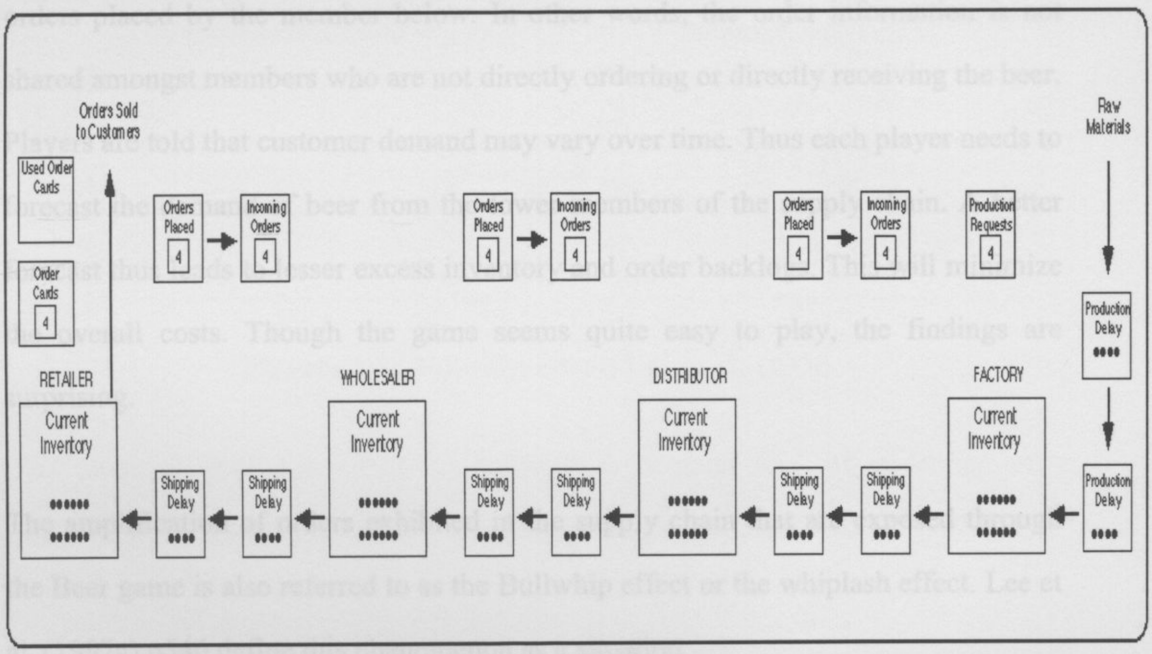


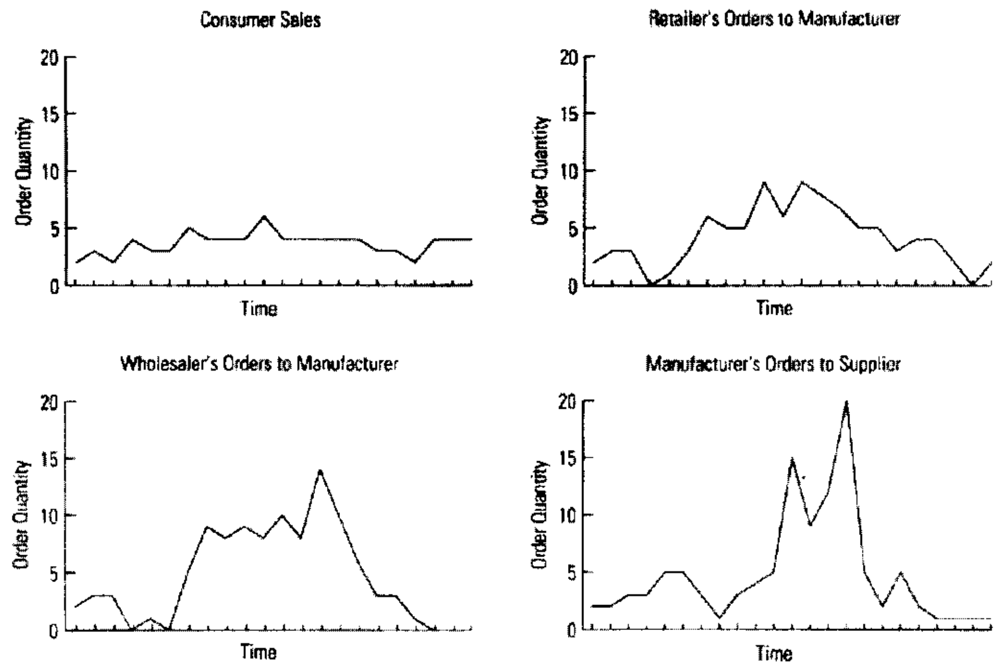
Figure 2.4: Beer distribution game; Sterman (1992b)

The beer game has been used in numerous studies to understand human response to dynamic complexity. It is therefore worth spending some time to describe the game in detail. As can be seen in Figure 2.4, the game comprises four sectors—customers, retailers, wholesalers and factory. Each sector is represented by one player or more (team). As in any business setting, orders are placed from the lower member of the supply chain to the one just above it. Customers place orders of beer to the retailers. Retailers respond to this request by shipping the order out to their customers. Retailers also order beer from the wholesalers. Wholesalers ship beer to the retailers. The wholesalers have the same relationship with the factory. The factory brews the beer. There are time delays (order-processing delays and shipping delays) at each stage. Participants are penalised for holding excess inventory and for having a backlog of orders. The inventory backlog cost is typically double that of excess inventory cost. The goal of each player/team is to minimise total costs. The game starts with all players in equilibrium. Each player has the same amount (typically 12 cases) of beer at this stage. The initial order from the customers to retailers is for 4 cases. It should be noted that only the next member in the chain knows the actual orders placed by the member below. In other words, the order information is not shared amongst members who are not directly ordering or directly receiving the beer. Players are told that customer demand may vary over time. Thus each player needs to forecast the demand of beer from the lower members of the supply chain. A better forecast thus leads to lesser excess inventory and order backlogs. This will minimize the overall costs. Though the game seems quite easy to play, the findings are surprising.

The amplification of orders exhibited in the supply chain that are exposed through the Beer game is also referred to as the Bullwhip effect or the whiplash effect. Lee et al. (1997a) p546 define this phenomenon as a situation,

> *"...where orders to the supplier tend to have larger variance than sales to the buyer (i.e., demand distortion), and the distortion propagates upstream in an amplified form (i.e., variance amplification)"*

J.W. Forrester was the first to study these amplifications in the supply chain (see Forrester 1961) and hence the phenomenon is also referred to as the Forrester effect (e.g. Moyaux and Chaib-draa 2007). The term 'bullwhip' is claimed to have been coined by Procter and Gamble when they observed the phenomenon in the orders of their products (Lee et al. 1997b). A typical representation of the bullwhip effect is depicted in the Figure 2.5 below.



Figures 2.5: Bullwhip effect; Lee et al. 1997b

The cause of Bullwhip effect is attributed to demand forecast updating, order batching, price fluctuation and rationing and shortage gaming (Lee et al. 1997b).

In a seminal study, Sterman (1989b) showed that people are unable to understand feedback when playing the beer game. Participant's orders showed 'oscillations' even though the rate at which customers ordered (exogenous input to the system) was constant. Sterman (1989b) infers that these were produced endogenously due to the interaction of participants' decisions and the structure of the system. People adopt an open-loop, event-oriented view of causality, ignoring feedback loops. Sterman termed this phenomenon as the misperception of feedback (MOF). The author explains that participants order beer cases for one period, but fail to account for them while placing orders for the next period. This leads to underweighting—a cognitive bias. The bias often leads to the bullwhip effect—defined as the tendency of orders to increase as one moves up a supply chain (Forrester 1961). Average cost per participant/team is $2000. The optimal cost is $200—one-tenth of the average (Sterman 1992b). Though most participants blame a fluctuating customer demand, they are shocked when it is revealed that the demand had been constant except a step increase from 4 to 8 cases. Sterman asserts that the participants' output pattern of orders *always exhibit* oscillations, amplification and phase lag. The beer game serves as a great teaching and learning tool for the study of dynamically complex systems.

More recently, Croson and Donohue (2006) used the computerised version of the beer distribution game to study the behavioural causes of the bullwhip effect. The authors conducted two studies. In the first, they demonstrated that underweighting and the bullwhip effect persist even when operational causes of these phenomena are removed. Unlike in Sterman (1989b), in this experiment, demand distribution was known and was stationary. The authors focussed on these two aspects as these have been commonly cited as causes of the bullwhip effect by participants. Knowledge of these however does not seem to remove bullwhip effect. Results show that this occurs because participants still underweight the supply line. In the second study, authors tested whether sharing inventory information might reduce the bullwhip effect. Their results suggest that sharing inventory information did not decrease the amplification of orders between supply levels and participants continued to underweight the supply line. Information sharing led to a greater reduction in order oscillations for manufacturers and distributors than for retailers and wholesalers. Overall, their results suggest that a limitation of our cognition to understand dynamic complexity is the cause of the bullwhip effect.

To summarise, beer game based studies reveal the bullwhip effect and the misperceptions of feedback and time delays in simple supply chain scenarios. These situations are common in business settings and hence add to the motivation of testing system dynamics in situations that involve dynamic complexity.

2.9.4 Dynamic decision-making studies that involve simulation games

The studies that fall under the last category are those that used a simulation game or management flight simulators. These tasks are also known as dynamic decision-making tasks. Dynamic decision-making is described as "decision making under conditions which require a series of decisions, where the decisions are not independent, where the state of the world changes, both autonomously and as a consequence of the decision maker's actions, and where the decisions have to be made in real time" (Brehmer 1992).

The misperception of feedback (MOF) hypothesis has been further confirmed by a few studies. Paich and Sterman (1993) used a market strategy simulation game to test participants' understanding of the system structure. They directly varied the strength of the feedback to test the MOF hypothesis. The authors suggested that higher feedback complexity will produce unanticipated behaviour in the system. Those participants who do not account for feedback would not be able to understand this behaviour. This would eventually lead to lower task performance (cumulative profit). During the experiment, participants managed a new product from launch to maturity using an interactive MFS. The underlying system dynamics model comprised of three parts: market sector, firm sector and competitor structure and strategy. The strength of feedback was controlled through the 'word of mouth' positive loop in the system. Participants made decisions on the price and capacity of the product. The task was performed by the subjects repeatedly. The MFS provided outcome feedback to participants after each round of decisions. Thus, participants had the potential to learn from one trial to the next. The findings of the experiment show that participants improved with experience. The mean performance increased by 50% in five trials. However, task performance deteriorated, relative to potential, when the strength of feedback was increased. The learning did not help participants in understanding the dynamics of the system that would result in better performance under high feedback complexity. The results of this experiment are consistent with the MOF hypothesis. The authors suggest that learning can be accelerated when participants have knowledge of representing the feedback structure rather than merely playing the simulation game.

In another study, Diehl and Sterman (1995) used a general stock management task to look at the effects of feedback complexity and time delays on task performance. They used three time delay and five feedback gain levels to design an experiment with 15 conditions. Task performance was compared with two benchmarks. In addition to this, time spent on each decision was also evaluated. As observed in Paich and Sterman (1993), outcome feedback was not sufficient to improve understanding of dynamic complexity. Participants' performance was far below optimal results. The results show that participants produce results that are characteristic of systems without feedback loops and time delays. This is in support of the MOF hypothesis. The findings also show that the time spent on making decisions did not vary with the complexity of the task. The authors conclude that people's mental models are dynamically deficient and exclude the influence of dynamic complexity. They suggest that this misperception can be overcome by training about feedback loops and through an effort to model the structure of the system. However, an understanding of dynamic complexity alone may not result in the ability to forecast the behaviour of such a system. The authors assert that computer simulation is essential for such tasks—an assertion that is directly tested in this research.

The two studies discussed above evaluated the effect of varying task complexity on task performance. Further, Gary and Wood (2005) tested the effect of task complexity on mental models that eventually influence task performance. They argue that task complexity is proportional to the amount of cognitive load. When this load exceeds working memory capacity, people tend to reduce problems with higher complexity to ones with manageable complexity. A similar finding was observed by two other studies (Diehl and Sterman 1995; Paich and Sterman 1993). The authors report an experiment where participants used the same MFS, as reported in Paich and Sterman (1993). The goal in this case was to maximize profits. Participants spent some time familiarising themselves with the MFS (learning phase). Following the learning phase participants completed a questionnaire. Soon after this, participants played the same simulation game (immediate transfer phase). Fifteen weeks after the initial phase, participants again played the game. In each of the three phases, participants completed 3 sets of 40 trials. Results indicate that a more accurate mental model led to higher task performance. The findings also suggest that the participants working on the simpler task developed more accurate mental models than those working on the complex task. Hence, just like the other studies discussed above, which propose and confirm the MOF hypothesis, Gary and Wood (2005) also found that overall task performance was far below optimal. In conclusion, this study contributes to the understanding of mental models—an important missing link between task complexity and task performance.

Misperceptions of feedback have also been observed in situations outside business settings. Using two studies, Moxnes (1998a, 1998b) showed that these misperceptions existed in the management of renewable resources. In the first study, participants were asked to manage the stock of cod fish using a MFS. They made decisions on ordering of new vessels and percentage utilisation of the fleet. Their goal was to maximize infinite horizon profits. The authors hypothesised that misperceptions of feedback would lead to overinvestment and overutilisation of fishing fleets even in the absence of a common property problem. The findings show that most participants overinvested. They built larger fleets than the benchmark. Hence, the results confirm the hypothesis. The authors suggest that an investigation is required into pedagogical devices that could correct such misperceptions of feedback in renewable resources. In the second study that focuses on overexploitation of renewable resources, Moxnes (1998b) explored the role of misperceptions of feedback on bioeconomics. Describing a specific case of reindeer herding and over-grazing of lichen, the study focused on crisis management in cases where overexploitation was a fact. Participants used a MFS to set yearly quotas for reindeer over a 12 year simulated period. The underlying model of the MFS represented a predator-prey scenario for reindeer and lichen. The goal for the participants was to maximize the present value of infinite horizon income. Results showed that the median participant captured only 9% of the infinite horizon. The authors suggest that this may have been due to inappropriate dynamic mental models and analysis, pointing towards the misperceptions of feedback in participants' decisions.

Poor task performance is not only a characteristic of complex task. Such deficiencies have been observed even in simple tasks. In a study involving a simple dynamic system consisting of two causal relationships, Jensen and Brehmer (2003) used a prey-predator example. The results show that only half (8 out of 15) of the participants were able to find the optimal solution. Out of these 8 participants, none applied an ideal feedback approach, though two of them understood the system thoroughly. The authors suggest that rather than understanding the system, participants paid attention to how to use the 'software' to achieve the desired result.

2.9.5 Summary

The main finding from the studies above is that dynamic complexity is poorly understood in the absence of a decision aid. This is primarily due to our inability to mentally simulate complex scenarios. This deficiency is seen in both static and dynamic decision-making tasks. Furthermore, feedback loops and stocks and flows are not understood well. People do not generally account for these and manage the task by reducing the problem to one without dynamic complexity. This leads to poor task performance. A lengthier discussion of such studies is done elsewhere (e.g. in Hsiao and Richardson 1999). The studies discussed also indicate that, a decision aid is required to perform well in dynamic situations. In addition, mere interaction with a simulation game might improve performance in some situations. Finally further research is needed to explore training programs that could improve the deficiencies.

In the context of the current research, findings of previous studies lead to two objectives. First, the revelation that our intuitive ability to understand dynamic complexity is poor and claims of system dynamics to improve this deficiency provides motivation to test the efficacy of qualitative and quantitative system dynamics, individually and combined with each other. Secondly, these studies provide useful information with respect to the experimental design, especially regarding the choice of the independent variable (task complexity) and type of tasks.

2.10 Claims of system dynamics as being useful

The previous section revealed that people are prone to making errors in the presence of dynamic complexity. System dynamics has been seen as a method that improves these abilities. This section provides an overview of the success of system dynamics methodology and applications from different perspectives.

Several researchers have evaluated the usefulness of system dynamics techniques in a variety of applications using case studies. For instance, Abdel-Hamid (1989) uses a case-study for a software project staffing problem to gain useful insights into the policies for managing human resources and to assess their impact on project behaviour. Another study by Lyneis et al. (2001) in the area of project management uses a system dynamics model to assess risks in an air defence system project. Along similar lines, Repenning's (2000) system dynamics model gives insights regarding "under allocation" of resources in multi-project development environments. Oliva and Sterman (2001) develop a dynamic model of a service organisation to understand how service quality could persistently deteriorate. Fiddaman (2002) uses a behavioural climate-economy model to explore policy options. Ford's (1990) case-study to estimate the uncertainty in price and demand of electricity claims to culminate in net savings of $250 million in option costs. System dynamics has also been applied to policy development programs such as in Cavana and Clifford (2006) to address the influence of price on the use and consequence of tobacco. All the above-mentioned applications clearly demonstrate the usefulness of system dynamics tools in making better decisions in complex situations.

System dynamics has been widely and successfully used in the industry to solve various problems. It has repeatedly been demonstrated as an effective analytical tool in a wide variety of situations, both academic and practical and is being currently used widely (Sterman 1992a). Stata (1989), chairman of Analog Devices, argues that one of the major causes of the decline of the US industry is the lack of management innovation. According to him, management innovation depends on new technology which depends on new knowledge, tools and methods. He further argues that systems thinking and specifically system dynamics are the tools that will help in innovation management which in turn will help organisations to change their mental models and learn continuously. Analog Devices applied system dynamics tools and consequently understood the functioning of the organisation better than before, which eventually lead to greater profits. The Boston Indicators Project (www.bostonindicators.org), a joint effort of The Boston Foundation and the City of Boston, Massachusetts, used systems thinking in their 2002 report—Creativity & Innovation: A Bridge to the Future. The Foundation worked with systems thinking consultants (Daniel Aronson, Four Profit Inc; Phil Clawson, Community Matters Group; and Brendan Miller and Osamu Uehara of the Massachusetts Institute of Technology) to help find a core theme in the changes in 200 indicators related to the greater Boston area's economic strength, civic life, community fabric, health status, diversity, and other areas. As a result, the report highlights the connections between economic innovation, transportation, the cost of living, diversity, demographics, and many other areas. System dynamics has also been used by major consulting firms such as McKinsey and Company, Accenture, Booz Allen and Hamilton and A.T. Kearney (Scholl 1995; Richardson 1999)

The growth of system dynamics can also be gauged by the strength of the system dynamics community around the world. The system dynamics society was formed at the MIT in 1983. Along with it was its journal the *System Dynamics Review*. The society has been going strong ever since and has around a thousand members from over eighty countries (System Dynamics Society). The society holds an annual conference that is attended by more than five hundred delegates. There are also worldwide chapters in the society, which hold conferences, seminars and meetings regularly. Subscription to the academic, product-oriented journal *System Dynamics Review* has increased over the years (Lyneis 1999). Another similar journal which is more process-oriented is *The Systems Thinker* which was founded in 1990 and has more than five thousand subscribers. Its annual conference attracts over one thousand attendees (Lyneis 1999).

System dynamics courses are being introduced at undergraduate and postgraduate level at various universities and short term courses are organised for the benefit of executives. The field attracts practitioners, academics, and students alike from across the globe to interact via various online networks. These networks are primarily used for sharing system dynamics information of mutual interest, and for advertising conferences and job positions in the field of system dynamics.

2.11 Lack of empirical research in system dynamics

As explained in sections 2.8 and 2.10, there is rich anecdotal evidence and reported observations that reinforce the contribution of system dynamics to understanding complex dynamic processes. However, 'observations' and 'anecdotes' are prone to errors and biases such as 'experimenter bias' (Rosenthal 1966) and 'the Hawthorne effect' (Roethlisberger et al. 1939) and hence are not sufficient to scientifically test the usefulness of systems interventions in understanding dynamic complexity (see Doyle [1997] for a detailed discussion). Observations and anecdotes also do not allow for the control of the many influences that would seem to drive results. An experimental approach allows us to do that. Experimental control is crucial in answering questions about the effectiveness of system dynamics tools on mental models (Doyle et al. 1996). Cavaleri and Sterman (1997) argue for the importance of experimental rigour in building a strong foundation for the refinement and use of systems tools.

In general, the effectiveness of decision support systems (DSS) on task performance has been previously tested under controlled experimental settings (e.g. Sharda et al. 1988; Mackay et al. 1992; Webby and O'Connor 1994). However, this has not been the case for system dynamics. The claims and popularity of system dynamics in academic and consulting work significantly outweigh the paucity of research that focuses on empirical evaluation of system dynamics' efficacy in a controlled setting. The last ten years have seen an increase in the number of experts raising such concerns. Sweeney and Sterman (2000) claim:

> *"The use of systems thinking and system dynamics is increasing dramatically, yet there is little evidence, or even systematic research, to support educators' and consultants' faith in its efficacy".* (p249)

Cavaleri and Sterman (1997) assert:

> *'Although various intervention techniques that fall under the rubric 'systems thinking' have become quite popular, little is known about their efficacy in enhancing organizational effectiveness or productivity. With few exceptions the relationship between the use of systems thinking and organizational performance remains the province of anecdote rather than rigorous follow up research.'* (p171)

Maani and Maharaj (2004) concur with these comments:

> *'...there is little empirical evidence to support the notion that systems thinking is indeed effectual in dealing with complexity'.* (p21)

Doyle (1997) focussing on system dynamics claims on improving mental models stresses:

> *"Many claims have been made concerning the ability of systems thinking interventions to change the nature and quality of thought about complex systems, ... [yet] important questions about the relationship between systems thinking and basic cognitive processes such as learning, memory, problem solving, decision making, and updating mental models remain unanswered".* (p253)

On similar lines Pala and Vennix (2005) have observed:

> *"Even though these purposes [improving understanding of dynamic complexity and the ability to recognize stocks, flows, time delays, and feedback relationships as well as to identify patterns of dynamic behaviour of a system over time] are widely accepted by system thinking/dynamics educators, studies showing that such courses have achieved these aims have been lacking in the literature".* (p147)

In summary, although it is widely accepted that experimental evaluation of the system dynamics method is essential, there is little research in the area that focuses on empirical assessment in a controlled setting.

2.12 Previous experimental studies

2.12.1 Introduction

This section discusses experimental studies where participants were provided with system dynamics intervention in order to measure the influence of the intervention on components of dynamic complexity. Table 2.3 summarises these experiments across seven key characteristics. It should be pointed out that all these studies are cross-sectional studies, i.e. where the dependent variable(s) was measured soon after the intervention. Longitudinal studies that evaluate the usefulness of the system dynamics intervention after a significant time has elapsed are discussed separately (section 2.17). The main reason for this segregation is that the time period between training and measurement in longitudinal studies might play a significant role in the retention of system dynamics concepts.

The summarisation process revealed that these cross-sectional studies were quite different from each other, especially with respect to experimental design, instrument used to measure dependent variable(s), amount of system dynamics training provided and the type of training provided to the participants. At the most fundamental level, some studies evaluated task performance on static tasks and others on dynamic tasks, i.e. using a MFS. In the discussion below, we first discuss all those studies where participants made decisions without the aid of a MFS (2.12.1). Subsequently the studies that use MFS are discussed (2.12.2). Within these two broad categories, the studies are then classified according to the type of system dynamics intervention provided to participants. Those that provided the least amount of intervention are discussed first, followed by those with larger interventions.

Table 2.3: Comparison of experimental studies that test the efficacy of system dynamics interventions (sorted by date of publication)

Citation → Characteristics ↓	Pala and Vennix (2005)	Pala and Vennix (2005)	Pala and Vennix (2005)	Qudrat-Ullah (2005)	Maani and Maharaj (2004)	Grosser (2005)	Hachr (2004)	Brekke and Moxnes (2003)	Moxnes (2004)	Kainz and Ossimitz (2002)	Cavaleri et al. (2002)	Barros et al. (2002)	Doyle et al. (1998)
Intervention	Qualitative and Quantitative	Qualitative and Quantitative	Qualitative and Quantitative	Debriefing and outcome feedback from MFS	Qualitative and Quantitative and outcome feedback from MFS	Qualitative and Quantitative and outcome feedback from MFS	Qualitative and Quantitative	Outcome feedback from MFS	Outcome feedback from MFS	Qualitative	Qualitative and outcome feedback from MFS	Qualitative and outcome feedback from MFS	Debriefing and outcome feedback from MFS
Content of Intervention	Introductory SD course at Uni	Introductory SD course at Uni	Introductory SD course at Uni	Debriefing only structured around the MFS	Participants trained in SD during university course prior to experiment; outcome feedback from MFS	2 groups were presented explicit representation of stock/flow structure of the model behind the MFS	Drawn from various ST/SD texts	During MFS only numerical advice from 2 different (simulation and optimization) models	Outcome feedback through MFS	Stocks/Flows paper based training	Causal loop diagrams, behaviour over time graphs, structure-behaviour pair, surfacing and testing assumptions, causal tracing	Interaction with MFS during task	Debriefing session structured around the MFS. Involved CLDs and discussion on results
Participants involved in drawing CLDs and/or constructing a simulation model to solve the task	Pre test/Post test repeated measures, no control group	Pre test/Post test repeated measures, control group	Pre test/Post test repeated measures, no control group	Single factor, completely randomised, pre test/post test, control group	No pre test, No control group	2X2 in between subjects (information treatment vs. participants' experience	2X2 in between subjects, pre test/post test	3X3 factorial;pre test/post test	Repeated measures, 3 trials of MFS	Pre test/Post test repeated measures, no control group	No pre test; repeated measures	No pre test; control group	Pre test/Post test repeated measures, control group
Task Description	Bathtub tasks (CO2) - subjects asked to draw behaviour of atmospheric CO2 when an input CO2 emission is reduced to 0	Bathtub tasks (Department Store) - subjects asked to identify the time period at which the stock reaches its maximum	Bathtub tasks (Manufacturing) - subjects asked to draw behaviour of inventory in the presence of one stock and time delay	MFS (Fish Banks II F) - built around 'tragedy of commons' phenomenon	MFS (Computech)	MFS (some hypothetical business)	Bathtub tasks (Manufacturing) - subjects asked to draw behaviour of inventory in the presence of one stock and time delay AND other tasks	MFS (Virtual Reality)	MFS (Simple - Like Bathtub, but using MFS. Complex - builds upon simple with 2 stocks)	Bathtub tasks (parking lot) - subjects asked to identify the time period at which the stock reaches its maximum	MFS (Luigi's Pizza and B&B Enterprises) decisions based on the games	MFS (simulation environment, project model & scenario models) + Project simulator	MFS (STRATEGEM-2) game of the Kondratiev cycle
Participants involved in drawing CLDs and/or constructing a simulation model to solve the task	No	No	Yes	No	No	No	No	No	No	No	No	No	No
Number of participants	70	220	107	78 (reduced from 87)	6 (reduced from 10)	10	156	64	simple - 34 (16+18, 2 runs) complex 31 (16+15, 2 runs)	64 (reduced from 94)	45	13 (reduced from 18)	46 (reduced from 64)
Key Findings	1. Intervention was not able to improve performance	1. Intervention was able to significantly improve participants' ability to discern between stocks and flows 2. Experimental group performed significantly better than control group	1. Intervention was not able to improve performance, in fact performance significantly deteriorated post intervention.	Debrief resulted in a significant improvement in 1. Task performance; 2. structural understanding, 3. heuristics knowledge, 4. cognitive effort, compared to those who did not receive the debrief	1. No direct relation between intervention and performance 2. Certain types of systems thinking are more useful than others. 3. High performers show a general pattern of behaviour when making decisions	1. Little SD knowledge is worse than no SD knowledge, on those provided with additional info do not perform better. 2. Experienced professionals perform better than novices.	1. ST skills increase the propensity to consider unintended consequences of implementing redundant controls. 2. Formal ST training helps participants in recognising unintended consequences regardless of their background	1. Both, the simulation model and optimization model were equally useful, but in different ways 2. optimization helped in identifying appropriate target stocks 3. simulation had a stabilizing effect, due to its rich dynamic structure 4 tools complement, not substitute each other	1. In both simple and complex tasks, participants unable to reach maximum sustainable herd size in 15 years 2. there is significant improvement from 1 trial to the second and third in both tasks. 3. In complex task, there is still potential for improvement	Intervention was able to significantly improve participants' ability to discern between stocks and flows	1. Worse before better behaviour observed. Participants perform better for the second MFS. 2.more the perceived value of tool, more is the frequency of use	Intervention was a useful decision aids for performance in software project management tasks	Intervention able to produce reliable changes in content and size of mental models as well as degree of feedback thinking. No significant impact on detailed and dynamic complexity

2.12.2 Studies that tested the role of system dynamics interventions in static environments

This section discusses those experimental studies in which system dynamics interventions were provided to measure performance on static tasks. The first experiment evaluated the role of qualitative intervention on task performance (Kainz and Ossimitz 2002). The other two studies (Pala and Vennix 2006 and Hetch 2004) evaluate the role of combined system dynamics intervention.

Kainz and Ossimitz (2002) conducted an experiment on sixty-four undergraduate students in which participants' performance on bathtub dynamics tasks (Sweeney and Sterman 2000) was measured. The study used five different tasks, based on the original bathtub tasks. For each of the tasks, two different cover stories were used, to avoid any learning due to familiarity with the same cover story. Task performance was recorded before, and after a one and half hour training on stocks and flows. The training comprised a lecture on the basics of stocks and flows (It was limited to qualitative concepts and did not include system dynamics computer modelling). Pre-test results show that performance in this simple stock/flow task is initially quite poor. This is evident by the fact that in Task 1, only 3% participants drew inflow and only 11% drew outflow correctly. Post-intervention, performance increased to 31% and 34% respectively. In Task 2, 56% identified the time period at which the quantity of the stock reached its maximum, 95% identified the time period at which outflow reached its maximum and 39% identified the time period at which the stock reached its negative maximum. After stock/flow training, 80% participants identified the time period at which the quantity of the stock reached its maximum though the percentage of participants that identified the time period at which outflow reached its maximum remained the same. In Task 3, 14% identified the largest value of stock from a graph showing inflow and outflow and 95% identified the correct outflow. Post-intervention, performance increased to 67% and 97% respectively. In Task 4, participants were required to describe the behaviour over time of a stock by using inflows and outflows. The mean result for seven items that together determined the correctness of the response was 0.36, which increased by 0.18 after stock/flow training. Task 5 was used to check participants' understanding of stock and flow relationships. Specifically participants were asked to depict the net rate of flow as behaviour over time graph of the stock. The mean result of seven criteria was 0.77, that increased by 0.15 later on. These results reinforced previous findings—that people do not completely understand even basic stock and flow structures (Sweeny and Sterman 2002; Ossimitz 2002). Interestingly, Kainz and Ossimitz (2002) demonstrated that this deficiency could be significantly overcome after their stock/flow training. Specifically, task performance increased. Overall, the study showed that qualitative system dynamics training was successful in assisting the participants understand stocks and flows that led to an improved performance in all the five tasks. Although, this study was the first of its kind to show that deficiencies in peoples' understanding of stocks and flows could be compensated with system dynamics training, its main drawback was that the design lacked a control group

On similar grounds, Pala and Vennix (2005) also evaluated participants' performance on bathtub dynamics tasks. However, unlike the training provided in Kainz and Ossimitz (2002), in this case, participants were provided training on qualitative as well as quantitative system dynamics. Three separate experiments were conducted on undergraduate students who were enrolled in an introductory system dynamics course. In the first experiment, participants were administered the 'department store' task (Sweeny and Sterman 2000), both before and after the intervention. Participants were presented with the inflow and outflow of people on a graph. They were required to identify the time period at which 1) the inflow was the largest; 2) stock reaches its maximum; 3) outflow was the largest; and 4) at which the stock was at its negative maximum. This task was used during the pre-test as well as in the post-test. In between the two tests, the experimental group was trained in system dynamics, whereas the control group was not. The results of this study followed the same trend as those reported by Kainz and Ossimitz (2002). Initial results on participants' understanding of stocks and flows were dismal. Only 32% and 20% participants were able to arrive at the time period when the stock reaches its highest and lowest value. These results increased significantly after the system dynamics intervention (to 54% and 39% respectively). The experimental group also outperformed the control group in the post-test (by 28% and 25% respectively). The authors suggest that the increase in performance was because of the system dynamics training. The second experiment of the same study used the 'manufacturing' task (Sweeny and Sterman 2000) to test the effectiveness of the system dynamics intervention. As in the previous experiment, the same task was administered before and after the intervention. There was no control group. Participants were required to draw the behaviour of the inventory of a one-stock (inventory) system in the presence of a time delay. They were provided with necessary information such as inflows, current stock and desired stock. Responses were coded using eleven different criteria. Interestingly, participants' performance did not improve even after a one-semester long system dynamics course at the university. Average performance fell from 81.2% in the pre-test to 72.5% in the post-test. To add to this surprise, authors reveal that participants were involved in the modelling of the same task during the intervention. The authors argue that the unexpected decrease in performance might have been due to a different way of perceiving the time delay. In the third experiment, the experimental group was administered the 'CO2' task (Sweeny and Sterman 2000) before and after the intervention. This experiment did not have a control group. In this task, participants were required to draw the behaviour of atmospheric carbon dioxide and global mean temperature when input (CO2 emissions) is reduced to zero. In this experiment as well, the intervention was not able to improve performance. The number of participants who correctly drew the CO_2 trajectory increased in the post-test (by 4.3%). However, 7.1% fewer participants drew the global mean temperature correctly in the post-test. The authors argue that participants could have been overwhelmed by the complexity of the task as there was not enough time for them to fully understand fundamental concepts discussed during the training. For a simple task, participants performed significantly better after the intervention but their performance deteriorated for a relatively complex task. Although, the system dynamics intervention resulted in mixed effects on task performance, this study suffers from a few drawbacks. Firstly, there were differences between the experimental group and the control group. The experimental group comprised 2^{nd} year students and the control group comprised 1^{st} year students, and hence were not comparable. The incentives for both groups were also different i.e. where completing the tasks was part of the curriculum of the course in which participants of the experimental group were enrolled, this wasn't the case for the control group. Secondly, the same tasks were used for recording performance before and after the intervention in all the three experiments. The familiarity with the task might have contributed to the improved performance in this study.

Like Pala and Vennix (2005), Hecht (2004) used a combined qualitative and quantitative system dynamics intervention to test whether participants' ability to recognise unintended consequences of implementing redundant controls in accounting improved after the intervention or not. The study used an in-between subjects design. Results were recorded both, before and after the intervention. Participants were administered the 'manufacturing' task (Sweeney and Sterman 2000) as well as other tasks. By using three groups from different academic backgrounds (those from accounting but not medicine, those from medicine but not accounting, those from neither accounting nor medicine) the author tested the usefulness of system dynamics “skills” and of formal system dynamics “training”. Results showed that (i) system dynamics skills increased the inclination to consider unintended consequences of implementing redundant controls and (ii) formal system dynamics training helped participants in recognising unintended consequences regardless of their background.

These studies show that qualitative and combined system dynamics interventions are successful in improving performance in simple tasks such as the bathtub dynamics tasks. However, there is no consensus on the efficacy of system dynamics interventions to relatively complex tasks.

2.12.3 Studies that tested the role of system dynamics interventions in dynamic environments

All the three studies that were discussed in the previous section measured static decision-making. In other words, information was provided to the participants and they had to make only one decision (e.g. to identify the time period of the largest value of stock). This decision had no implication on further decisions—a characteristic of dynamic decision-making. In dynamic decision-making, a series of decisions are made in real time. Here the outcome of the previous decision influences the decision-making environment, thus influencing future decisions. Simulation games or management flight simulators capture this process well using computer software. Here we discuss some of the significant studies that test the usefulness of system dynamics to task performance and/or understanding. The studies are discussed on the basis of the level of intervention provided to participants. Decision aids can be provided using various means. As seen before, one dimension of a decision aid lies in the different amount of system dynamics training provided. This may range from a 90-minute stock/flow ‘crash course’ (Kainz and Ossimitz 2002) to a full-fledged university course on system dynamics (Pala and Vennix 2005). Another dimension of the decision aid comes into the picture when we consider model transparency in simulation games; in other words, the support provided by the MFS by revealing the underlying structure. At the bare minimum, the MFS environment provides an interface that has already identified key variables of the system. It presents a pre-made, error free system dynamics model that is ready to be simulated. Apart from the basic features, in some experimental designs, participants may be able to view the underlying stock and flow structure of the system. They may even have access to causal relationships between factors of the system. These models may facilitate participants' understanding of the system. In other experiments, the researchers might explicitly explain these underlying models with respect to the task in question. Therefore, there can be numerous permutations and combinations of decision aids across the two dimensions discussed above. In the subsequent paragraphs, 7 studies are discussed in decreasing order of the level of decision aid provided to participants. In some studies, participants were first provided system dynamics training and then results were recorded using a MFS. This way, participants were not only trained in system dynamics, but their decision-making process was facilitated by outcome feedback from the MFS. In some of these cases, the underlying stock and flow model/causal loop diagram was also made available.

In a study conducted by Maani and Maharaj (2004), participants underwent a full semester of system dynamics course at university. Similar to Pala and Vennix (2005), the curriculum of this course consisted of training on qualitative as well as quantitative system dynamics modelling. Each participant, acting as a CEO of a company, was required to make decisions on five variables—sales force headcount, sales compensation, marketing spending, average price per unit and capacity order. The task for the participants was to maximise 'revenue', 'profit' and 'market share' using the MFS—Computech. Participants' responses were classified under one of the five categories of systems thinking as defined by Richmond (1993). These results were analysed through Verbal Protocol Analysis (VPA). Measures from the MFS were also taken into account. Participants' understanding of the structure of the task was operationalised by categorising responses into four categories—basic one-to-one relationships, complex one-to-one relationships, three-way relationships and big picture. The results indicated that there was not a direct relationship between intervention and task performance. The authors suggested that certain types of system dynamics skills are more useful than others. Better performers undertook greater 'forest thinking' (seeing the big picture) than others. Also, the higher systems thinking skills—forest thinking, closed-loop thinking and operational thinking contributed more towards task understanding and performance. Their results also indicate that high performers show a general pattern of behaviour when making decisions and term it the CPA (Conception-Planning-Action) cycle. Although the study by Maani and Maharaj (2004) offers insight into system dynamics effectiveness, it suffers from some serious limitations. First, only six participants were able to successfully complete the experiment questioning the generalisation of these results. Second, the absence of a control group does not allow comparison with participants who did not take the course but played the game. Finally, the absence of a pre-test meant that an option to compare participants' ability to perform on the same task before the system dynamics intervention was not available.

In a similar study, Barros et al. (2002) evaluated the efficacy of system dynamics tools as decision-aids in software project management tasks. Participants, each role-playing a project manager, were required to make decisions such as allocating resources to projects and deciding the number of hours developers should work per day. The goal was to minimize the total time taken and minimize the cost of the project. Half the participants underwent a thirty-minute qualitative system dynamics training and used a system dynamics-based MFS to make decisions (experimental group). The experimental group relied mainly on the outcome feedback from the simulator. They were also able to access the underlying system dynamics model during the experiment. The other half did not undergo any system dynamics training and were not provided with the MFS/system dynamics model (control group). They made decisions based only upon their personal experience. The results of this study show that those who used the system dynamics model took significantly less time than those who did not use the model. The authors suggest that the intervention (system dynamics training, outcome feedback from MFS and availability of the system dynamics model) is a useful decision-aid for task performance in software project management.

In yet another study that combined system dynamics training with MFS as decision-aid, Cavaleri et al. (2002) tested the efficacy of five qualitative system dynamics tools in relation to task performance. The tools that were evaluated are causal loop diagram, behaviour over time graphs, structure-behaviour pair, surfacing and testing assumptions and causal tracing. Performance was recorded using two MFSs—Luigi's Pizza and B&B Enterprises. Participants first played the Luigi's Pizza MFS and then played the B&B Enterprises MFS. These were played one after the other. Results show that participants performed poorly in the first MFS, but significantly improved upon their performance during the second MFS. Their findings also suggest that the frequency of the use of a system dynamics tool is directly proportional to participants' perception of that tool.

In some studies, task performance was recorded using an MFS but no explicit system dynamics training was provided to the participants. In such situations, either outcome feedback alone or combined with the availability of the underlying system dynamics model was provided as a decision-aid to participants. An example of a study where only outcome feedback was provided to participants has been reported in Moxnes

(2004). In this study, participants were required to manage a reindeer rangeland. Their performance was recorded over three trials of a MFS. The study used two tasks of varying complexity. The simpler task was similar to the bathtub dynamics task (a one-stock system). Participants received exact information about the flows, stock and other variables in the system. The complex task was an extension of the simple task into a two-stock system. In both the tasks, participants were required to find the optimal solution. Four experiments were conducted—two runs for the simple task and two runs for the complex task. Participants made decisions over multiple time periods. Each new decision was supposed to be made based on the output of the previous simulation of the model. The underlying stock and flow model was not available to the participants. Unlike the studies previously discussed, no explicit system dynamics training was provided to participants. The only information available was the outcome feedback from the underlying system dynamics model. The results from this study show that participants were unable to reach the maximum sustainable herd size in 15 years, in both simple and complex scenarios. Even precise outcome feedback was not enough to compensate for misperceptions of feedback. There was significant improvement in performance from the first trial to the second and from the second trial to the third. There was still potential for improvement in the complex task.

Another example where only outcome feedback was available to participants is reported in Brekke and Moxnes (2003). The study compares the usefulness of a system dynamics-based simulation model and an optimisation model on the management of stocks of fisheries. Participants made decisions using the MFS. Numerical advice was provided to the participants. Pre-test/post-test design was used to measure changes prior to and after participants were provided with numerical advice. The results of this study suggest that both the simulation model and the optimisation model were equally beneficial. In particular, the authors argue that the rich dynamic structure of the system dynamics simulation model led to improved task performance.

There are instances of experiments where participants were allowed to access the underlying stock/flow structure. In one such study, Grosser (2005) used a 2 X 2 in-between-subjects design to test the influence of availability of information (underlying stock/flow model) and experience in system dynamics to task performance. The results showed that those who were provided with the stock/flow structure performed worse than those who were not provided with this information. The author infers that a little knowledge of system dynamics is worse than no knowledge of system dynamics. The results of the same study also reveal that those with substantial system dynamics experience (greater than 1.5 years) perform significantly better than those with lesser experience (0-1.5 years).

There have also been experiments where participants were provided information on the task through debriefing supported by MFSs. The difference between the studies reported above and these types of studies is that whereas system dynamics training in the former was of generic nature, in the latter, the debriefing is catered specifically towards the understanding of the task on which performance is measured. In one such study, Qudrat-Ullah (2005) reports the benefits of such a practice. Their study used a pre-test/post-test design with a control group. Performance was measured using the FishBankILE MFS. Participants, as fishing fleet managers, made decisions on ordering of ships and on the percentage utilisation of the fleet. Task performance was measured by cumulative profits over a 30-year period and the remaining fishes in the final year. Half the participants took the pre-test, then were debriefed about the task and later took the post-test (experiment group). The other half of the participants also took the pre-test and the post-test, but they were not part of the debriefing activity (control group). During the debriefing, participants' performance charts were discussed. The structure of the system and its behaviour were also discussed explicitly. Participants were allowed to ask questions during this period. The results show that those who underwent the debriefing performed significantly better than those who did not. Specifically, the experimental group performed better on the task; exhibited better structural understanding and knowledge of heuristics; and made decisions faster than the control group. According to the author, the debriefing activity helped participants to reflect on their understanding of the system structure and gain clarity on the misconceptions of the task. This eventually led to an improved understanding of the system. Doyle et al. (1998) report another study that combined debriefing and MFS outcome feedback to facilitate task performance. Participants were required to understand the dynamics of the economic long wave using the STRATEGEM-2 MFS. Similar to the process reported in Qudrat-Ullah (2005), the debriefing was provided to participants between the pre-test and the post-test. The debriefing consisted of a detailed description of causal loop diagrams and discussion of results obtained in the pre-test. Unlike in Qudrat-Ullah (2005), participants in this study had access to detailed information about the structure of the system. This information was available as a structure diagram, behaviour over time graphs for key variables and information on changes over time in all variables during the game. The results from this study suggest that the intervention was able to produce reliable changes in the content and size of mental models as well as the degree of feedback thinking. The intervention did not have any significant impact on the level of detail or dynamic complexity.

Although all these studies report positive changes as a result of system dynamics interventions, in none of these studies participants were involved in qualitative or quantitative model building—a key component of the system dynamics method.

2.13 Identification and relevance of gap in literature

The experimental studies discussed in section 2.12 contribute to the interest generated in efficacy system dynamics methodology. They suggest that in general, system dynamics training assists people to understand/perform in the presence of components of dynamic complexity, when compared with their understanding without any system dynamics training. The results of MFS-based studies also contribute to the motivation of the current research. Some important questions regarding the efficacy of system dynamics tools remain unanswered. For instance, the *relative* efficacy of qualitative, quantitative and a combination of qualitative and quantitative system dynamics tools on task performance for both simple and complex tasks has not been addressed so far.

Richardson (2001) asserts that qualitative mapping could provide 'structural' insights, such as knowing about internal relationships between the parts of a system. However, according to him, 'structure and behaviour' or 'dynamic' insights, such as forecasting the behaviour of a variable can only be provided through quantitative modelling. This assertion has been concurred with by many others (e.g. Peterson and Eberlein 1994). Sterman (2002a), while acknowledging the difficulty in creating formal models and highlighting the benefits of qualitative mapping, argues that qualitative system dynamics is not sufficient even when the purpose is mere 'insight'. He believes that computer simulations are far more reliable than the mental simulations that one might use with qualitative system dynamics.

The relationship between qualitative and quantitative system dynamics raises sufficient interest (Wolstenholme 1999; Coyle 2000; Coyle 2001; Homer and Oliva 2001) but has not yet been tested in the context of decision-making under dynamic complexity. The current research sheds light on some crucial research questions raised such as

> *"The field must address the relationships between qualitative mapping and quantitative modelling—in short, when to map and when to model"* Richardson (1999) (p8)

And,

> *"How can we* measure *the value added by the extra effort of simulation or the value lost by not simulating?"* Coyle (2001) (p359)

Richardson, in his 1997 (Richardson 1997) presidential address, suggested that collaboration between the two approaches needs to be worked out to get the best out of the two approaches. A different argument on the same lines was put forward by Peterson and Eberlein (1994). They argue that both qualitative system dynamics and quantitative system dynamics have much to learn from each other. Qualitative system dynamics requires quantification to enable practitioners to precisely define structure and subsequently forecast behaviour of the system. On the other hand, the understanding gained through qualitative system dynamics could assist in validating quantitative models.

2.14 Task complexity

The investigation of previous studies in section 2.12 (Previous experimental work on evaluation of system dynamics interventions) and section 2.9 (Decision making in a dynamically complex environment) revealed the use of task complexity as an independent variable. In the context of the current study, it is proposed that qualitative system dynamics and combined system dynamics have differential effects on performance depending on the level of complexity of the task. One may envisage that, for simple tasks, the incremental contribution of combined system dynamics will be quite small, but for complex tasks combined system dynamics should be quite useful. To test the usefulness of system methodologies in dynamically complex situations, people's ability to understand fundamental concepts of dynamic complexity (stocks and flows, time delays, non-linearity, feedback loops and the dynamics that arise from them) needs to be tested before and after qualitative system dynamics and combined system dynamics interventions. Though it is possible to test the effectiveness of these tools on the same task, the additional information produced from the quantitative phase will help in understanding components of dynamic complexity. Predictors related to task-complexity have been reported to have significant effects on task-performance (Hsiao and Richardson 1999). For instance, feedback-loops that are commonly used in system dynamics are used also as indicators that reflect task complexity (Hsiao and Richardson 1999). The strength of the feedback loop has been shown to be directly proportional to task-complexity (Hsiao and Richardson 1999) and hence to task-performance (Sterman 1989a; Sterman 1989b). Similar discussions have also been witnessed in the data modelling community (Bajaj 2004 and Batra and Wishart 2004).

2.14.1 Forecasting and performance in complex tasks

An important aspect of measuring performance in a complex task is to measure people's ability to forecast. Forecast accuracy was measured as forecasting is a critical tool for decision-making and planning in today's business environment. For example, forecasting plays an important role in areas such as scheduling, acquiring resources and determining resource requirements. Forecasting situations are diverse and differ widely with regards to time horizons, types of data patterns, external factors etc. To deal with the diversity of applications, several statistical techniques have been developed. Statistical techniques can be applied in conditions where information about the past is available, is quantifiable and it can be assumed that certain features of the past pattern will continue into the future. Since statistical techniques assume that existing patterns or relationships in the data will continue and not change in the forecasting phase, human judgment can play an important role in recognising changes that can and do occur. In these cases, the human forecaster provides judgmental modifications to the forecast obtained using a statistical method (Lim and O'Connor 1996). In other cases the forecast is entirely based on the judgment of the forecaster, with no statistical intervention. It is well known that systems that incorporate both – statistical and judgmental approaches are more effective (Webby & O'Connor 1994), especially for short-term forecasting (Goodwin 2002). Another approach is to the use of "decision-aids" to support judgmental forecasting (Goodwin 2005). In these cases, the decision aid does not produce a statistical forecast but provides quantitative information that aids in better judgmental forecasting. Although there have been attempts in the past to create decision support systems to aid judgmental forecasting (e.g. Rajadhyaksha and Dwivedi 2003), there is a need to develop improved methods that particularly support judgmental forecasters (Lawrence et al. 2006). Forecasting has been applied in system dynamics as an extension where parameters are changed in combination with utility functions (e.g. variable to be forecasted) to assess the behaviour of the variable over time or to measure the value of the variable. Examples of studies that have used forecasting through system dynamics include Coyle 1985; Moxnes et al. 2001; Graham and Ariza 2003. Hence, it may be optimistically argued that system dynamics modelling may provide an improvement in judgmental forecasts, thus making the decision-maker more confident about their judgment. For this reason, a component of this thesis evaluates the role of system dynamics in forecasting.

2.15 Deriving research questions

The literature review set out with fundamental and general questions about the efficacy of system dynamics methodology. A comprehensive review of the system dynamics literature has pointed towards some specific questions.

The first research question addresses a fundamental concern about the efficacy of the two stages of system dynamics and the combined system dynamics process.

RQ1. Do people trained in either (a) qualitative system dynamics, (b) quantitative system dynamics or (c) combination of qualitative system dynamics and quantitative system dynamics perform better in simple and complex tasks when compared with people who have not been trained in system dynamics?

To explore the relative efficacy of the qualitative, quantitative or combined stages, the performance after each intervention needs to be compared.

RQ2. Do people trained in qualitative system dynamics perform better in simple and complex tasks when compared with people trained in quantitative system dynamics?

To explore the relative efficacy of the qualitative system dynamics and quantitative system dynamics with respect to the combined approach, performance after individual and combined intervention needs to be compared.

RQ3. Do people trained in qualitative system dynamics perform better in simple and complex tasks when compared with people trained in both qualitative system dynamics and quantitative system dynamics?

RQ4. Do people trained in quantitative system dynamics perform better in simple and complex tasks when compared with people trained in both, qualitative system dynamics and quantitative system dynamics?

2.16 Longitudinal research

Even though many studies claim to have made changes to mental models, very few have gone a step further to conduct longitudinal experiments to test the efficacy of the system dynamics training after some time has elapsed (e.g. Cavaleri and Sterman 1997; Huz et al. 1997). It is also important to evaluate whether system dynamics intervention made "fundamental" changes to participants' mental models. Specifically, which components of the methodology were retained and which ones were forgotten a few months after initial training. The need for rigorous evaluative research is widely accepted. For example, Stake (1967) writes of educational innovations,

> "*...folklore is not a sufficient repository. In our data banks we should document the causes and effects, the congruence of intent and accomplishment, and the panorama of judgments of those concerned*" (p539).

Moreover, it is well known that studies that involve observation at a single point in time (cross-sectional studies) are deficient in accounting for a high percentage of variance (Heller et al. 1977). A longitudinal study aims to consider any possibility of events that might have occurred in the elapsed time period (Heller et al. 1977). Hence, conducting a longitudinal experiment becomes imperative in situations where the long-term effectiveness of an intervention needs to be tested. In the context of system dynamics, the need for a rigorous longitudinal study arises from the fact that most evidence that exists to date in this area is anecdotal and lacks systematic evaluation (Huz et al. 1997). Further, Cavaleri and Sterman (1997) argue that rigorous longitudinal research is essential to build a strong foundation for the enhancement and optimal use of system dynamics methodologies. It needs to be proven whether exposure to system dynamics training produces lasting changes in the way people think and analyse complex problems.

Generally, conducting longitudinal research has been problematic due to the cost associated and the relative paucity of adequate methodology to carry out such studies (Heller et al. 1977). The lack of commitment by clients to carry out longitudinal research in real applications has been an additional challenge for system dynamists (Huz et al. 1997).

In general, memories of business education are short-lived. Anderson (2000) argues that decay, interference and absence of retrieval cues are the three prime factors that are responsible for forgetting what has been learnt in class. In this study, we assume that a time gap of five months would have resulted in decay, participants' involvement in other activities might have constituted interference and their abstinence from using system dynamics constituted an absence of retrieval cues. Intuitively one would expect that concepts that were taught five months ago would now have been forgotten.

Contrary to this belief, longitudinal studies in system dynamics have demonstrated that people do retain system dynamics skills. Two such studies are described in detail below. However, an important aspect should be noted. Participants in these studies were using system dynamics skills during the retention interval. Also, the first study was not conducted as a controlled experiment.

In a first of its kind study, Cavaleri and Sterman (1997) evaluated the change in thinking, behaviour and performance of two groups of people (managers and non-managers) who were subjected to different levels of systems thinking intervention. Non-managers merely played a board game that helped in understanding inefficiencies of supply chains—the beer game (Sterman 1992). Managers also played the beer game, but in addition to that, played another simulation game that was specific to the work they were involved in and also attended a seminar on systems thinking. The results of this retrospective study suggest that there were significant differences in the amount of systemic thinking in the behaviour of managers, but not in the case of non-managers. Both managers and non-managers rated the systems thinking training as 'moderately valuable'. There was no evidence that the system dynamics made any difference in performance during the six years following the initial training. Organisational performance was measured by using four standard measures that were specific to the insurance industry. The data collected did not suggest any pattern of improvement in any of these four measures. Given the difficulty in conducting longitudinal experiments in industry, this study is a pioneering research which sets the stage for further research in the area. However, there are some serious limitations. Firstly, this study wasn't performed in a controlled setting and hence it is difficult to ascertain if the difference in systemic thinking and behaviour was due to systems thinking training alone or attributed to other factors such as intrinsic differences between the participants (managers and non-managers). Secondly, the majority of the data collected from participants was in the form of self-reporting questionnaires. However, as pointed out by the authors themselves, these are prone to 'demand effects'. In this situation managers would have reported positive impact of systems thinking training as they believed that this was what the researchers were looking for.

In another example, that of a project that focuses on the integration of mental health and vocational services, Huz et al. (1997) evaluated the impact of systems thinking interventions over a six-month period. They conducted a controlled experiment using a pre-test/post-test design in which they measured changes in participants' mental models to assess the impact of systems thinking interventions. Similar to Cavaleri and Sterman (1997), this study too measured participants' mental models and organizational performance. In addition to these two measures, the modelling team evaluated their own performance as well. These key measurements were operationalised using ten 'domains' to measure changes prior to and after the intervention. The change in participants' mental models was operationalised using five domains of measurement and analysis. These were: participants' perceptions of the intervention, shifts in participants' goal structure, shifts in participants' change strategies, alignment of participants' mental models, shifts in understanding of how the system functions. The first four were measured using pre-test/post-test surveys and the last was operationalised via follow-up interviews, formal meetings' minutes, and informal reflections by the modelling team. Participants' perceptions of the intervention were evaluated using the 'model building evaluation questionnaire'. Results suggest that participants perceived the intervention as productive. Specifically, building a formal model proved beneficial to their understanding of the system. To measure shifts in participants' goal structure, participants rated the level of importance they associated with 20 goal statements, before and after the intervention. The ratings were recorded on a five point Likert scale. Paired T-test analysis revealed that there was a significant shift in the importance of shared services and common intake between two departments. This was an important aspect as the fragmentation between the two departments was a concern for the efficient working of the client system. Shifts in participants' change strategies were measured in a similar way as for the previous domain. Paired T-tests in this case show that there was significant shift in the importance of four change strategies. Alignment of participant mental models was measured using two questionnaires. Results show that participants were more aligned in their perceptions of system goals. However, there were no significant changes in their alignment in their perceptions on strategies for change. Using data from archives, informal reflection and observation in meetings, the authors concluded that the formal model also facilitated participant understanding of system structure and behaviour. Organisational performance was defined as shifts in the network of agencies that support services integration, changes in policies and changes in outcomes for clients. These were measured via follow-up interviews and analysis of project archives and administrative data. The modelling team's reflection of the process was analysed using minutes of sessions, archival analysis and informal reflections of the team members. Huz et al. (1997) raise important research questions based on their follow-up, such as "...what is the contribution of specific sub-components of the overall intervention..." (p150)

2.17 Research questions – part II

The initial research questions test the efficacy of system dynamics tools, both individually and combined with each other. They do not test however, whether the changes in participants' understanding and performance would last for long. The following two research questions test this element of system dynamics efficacy.

RQ5. Do people trained in system dynamics retain these concepts and perform equally well in simple and complex tasks after a few months have elapsed since the initial training?

RQ6. Does re-familiarisation of system dynamics software assist in participants in simple and complex tasks after a few months have elapsed since the initial training?

2.18 Experimental design

To adequately address the research questions, three experiments were designed using a pre-test/post-test setup. This design enabled the measurement of performance before, and after the system dynamics intervention(s). Though fundamentally, the design was similar for all the experiments, it varied to an extent in each experiment depending on the demands of the comparison. For instance, the first experimental design compared the efficacy of combined system dynamics and qualitative system dynamics with each other and with baseline performance. Testing these was possible using a repeated measures design. In the second experiment, quantitative intervention too needed to be compared with qualitative, combined and baseline performance. Further, the design needed to restrict any learning that might occur from using the same cohort for each test and/or from using the same questions. Thus, a more complex design was required. The experimental designs for each experiment are discussed in detail in the three empirical chapters. The table below (Table 2.4) shows which research questions were addressed through which experiments.

Table 2.4: Experiments and Research Questions

Experiment → Research Questions↓	Experiment 1	Experiment 2	Experiment 3
RQ 1 (a)	•	•	
RQ 1 (b)		•	
RQ 1 (c)	•	•	
RQ 2		•	
RQ 3	•	•	
RQ 4		•	
RQ 5			•
RQ 6			•

Participants' performance and logic to solve the simple and complex task was evaluated. In addition to these measures, their ability to make accurate forecasts was also measured. Since the main purpose of this research was to measure task-performance with respect to system dynamics in a controlled setting, we applied guidelines from Doyle et al. (1996) to set up the experiments. Doyle et al. (1996) proposes eight "goals" that should be incorporated to conduct the study with rigour and robustness. These are highlighted below and were incorporated in the experimental scheme to ensure a robust design.

We detail the experimental design using the eight criteria (Doyle et al. 1996) mentioned below:

1. *Attaining a high degree of experimental control*: This was achieved by using a pre test/post-test design in which participants' baseline scores were first measured. The experiments were held in a classroom setting and hence it was relatively easy to control the variables under examination, keeping others constant. It was necessary to mitigate the effect of superfluous variables such as monetary incentives, job-pressure, market-influences etc. to eliminate any biases that that might hamper our understanding of the variables being examined. To further enforce control, it was made sure that participants who

appeared in the "pre-test" had no prior training or experience related to system dynamics.

2. *Separation of measurement and improvement*: We recorded changes in participants' performance, confidence and understanding, pre- and post-intervention. The interventions themselves were the means by which we aimed to improve mental models, hence ensuring that measurement and improvement occurred at different times of the experiment and through distinct and separate procedures.

3. *Collecting data from individuals in isolation*: All tests were performed individually by all participants in isolation from each other.

4. *Collecting detailed data from the memory of each individual*: According to Doyle et al. (1998), a participant's mental models may be temporarily influenced by statements made by other participants in group discussions which might lead to eliciting mental model that are not necessarily their own. Mental models were elicited from individuals in isolation and not in groups. The interventions in this experiment discuss generic concepts of the methodologies. The tests were based on a case study and answering the questions meant that participants were required to 'apply' systems concepts rather than merely reiterate what was taught to them.

5. *Measuring change rather than perceived change:* Conducting a pre-intervention test and comparing the results of this test with ones conducted after systems thinking and system dynamics interventions helped in measuring 'actual' change rather than 'perceived' change.

6. *Obtaining quantitative measures of characteristics of mental models*: To achieve this we obtained quantitative measures of characteristics of mental models by defining the characteristics of dependent variables, a priori, along with quantitative measures associated with each of them. Some of the measures were quite straightforward and quantitative by nature—such as identification of stocks and flows and the measurement of change in confidence from one test to the next. Inferring the behaviour of a key variable was measured by taking into account three significant factors associated with making the forecast.

7. *Employing a naturalistic task and response format*: The tasks pertained to common business problems such as forecasting of human resources, sales and population in cities in today's organizational setting. The response format was open-ended questions that are very likely to be encountered in any business setting—they did not lead participants to a predefined solution.

8. *Obtaining sufficient statistical power*: The three experiments were conducted on 31, 80 and 30 participants respectively.

2.19 Participants

All three experiments used postgraduate students as participants. It may be argued that since the experiment was conducted in a university setting, these results cannot be widely generalised. However, it should be noted that in the past, the majority of such experimental studies have relied on tertiary students (Rouwette, Größler et al. 2004). However, it should be noted that numerous studies in the past have used students as participants. In fact, students were used as subjects in all the studies discussed above.

In decision-making experiments, task-performance is not significantly affected by factors such as computing skills (Trees et al. 1996), cognitive style (Maxwell 1995), task expertise (Bakken 1993) and personality type (Georgantzas 1990), hence solidifying our assumption that participants were at an equitable level and limiting possible bias in our analysis.

2.20 Contribution

The main contribution of the current research is to establish a relationship between qualitative and quantitative system dynamics, as well as add to the literature on the overall efficacy of system dynamics methodology.

The research draws together:

- A rigorous experimental approach to measure the efficacy of system dynamics.
- Use of model-building, rather than merely exploration of a pre-made model. This approach tests the effect of the system dynamics *process* rather than the effect of playing a simulation game.
- The testing of the efficacy of individual components of the system dynamics process (qualitative and quantitative), as well as their combination against a baseline condition, i.e. performance without the aid of these tools.
- The testing of the relative efficacy of qualitative system dynamics, quantitative system dynamics and combined system dynamics with respect to each other.

- The evaluation of the long-term efficacy of the system dynamics method using a longitudinal experiment.
- Using task complexity as the independent variable.

Chapter 3

Experiment 1

3.1 Introduction

3.2 Background

3.3 Methodology

3.4 Results

3.5 Discussion

3.6 Conclusion

3.1 Introduction

"How much value does quantified [system dynamics] modelling add to qualitative [system dynamics] analysis?" (Coyle 2000)

Qualitative and quantitative system dynamics (SD) methodologies have been successfully applied in the area of dynamic decision-making. Though these have been generally used in conjunction with each other, more recently, the use of qualitative system dynamics alone has found support with some authors. Exclusion of the computer modelling phase has sparked a debate on the relative usefulness of qualitative system dynamics alone as compared to a combination of qualitative system dynamics and quantitative system dynamics. The objective of the current study is to evaluate the relative efficacy of the combined approach versus that of qualitative system dynamics only, as well as to add to the literature on the overall usefulness of qualitative and combined system dynamics tools.

This chapter has been organised in the following way. The second section (Background) discusses some of the arguments for and against the use of qualitative system dynamics. It also describes the results of similar experimental studies that have previously been conducted. The Methodology section details the experimental design and the tasks used. Qualitative system dynamics and combined system dynamics interventions were administered to test the above-mentioned hypothesis for both simple and complex tasks. Subsequently, the results are presented and contrasted to those obtained by previous studies. Finally, the last section (Conclusion) presents an overall summary of the findings and discusses limitations.

3.2 Background

System dynamics was mainly perceived and used as a quantitative simulation technique when it was first introduced in 1961 (Coyle 2000). However, qualitative system dynamics tools have often been used to describe a system before a simulation model is formulated. Though the use of a combination of qualitative system dynamics and quantitative system dynamics is popular for inferring the behaviour of a system through its feedback structure, the last twenty-five years have also seen enthusiastic support for the use of qualitative system dynamics alone, which some authors believe may be sufficient for solving problems of dynamic complexity (Wolstenholme 1999).

The relative effectiveness of these tools is an area that raises interest (Wolstenholme 1999; Coyle 2000; Coyle 2001; Homer and Oliva 2001) but has not yet been tested in the context of dynamic decision-making. For instance, in an interview (Keough and Doman 1992 p17), Forrester argues that *"some people feel they have learned a lot from the systems thinking* [qualitative system dynamics] *phase. But they have gone less than five percent of the way into understanding systems. The other 95 percent lies in the system dynamics* [quantitative system dynamics] *structuring of models and simulations based on those models. It is only from the actual simulation that inconsistencies within our mental models are revealed."* On the same lines, Homer and Oliva (2001, p353) stressed that "... *uses of stand-alone mapping* [qualitative system dynamics] *are, perhaps in most cases, better than nothing...* ". Lane (1997) termed qualitative system dynamics as 'system dynamics lite', substantiating the use of this term by pointing towards the deficiencies of both, causal loop diagramming and system archetypes. On the other hand, proponents of qualitative system dynamics have argued that the qualitative phase alone adds significant value to the combined system dynamics methodology (Coyle 2001). To date there has been little agreement on the usefulness of combined system dynamics versus qualitative system dynamics with regards to performance in situations that involve components of dynamic complexity. Several studies have reported anecdotal evidence of the use of these methodologies; however, no controlled studies have been found that compare the relative influence of qualitative system dynamics and combined system dynamics on performance. In order to resolve this controversy and gain clarity regarding the use of qualitative system dynamics and combined system dynamics, it becomes important to experimentally determine the relative effectiveness of these methodologies. The current research sheds light on some crucial questions such as *"the field must address the relationships between qualitative mapping and quantitative modelling—in short, when to map and when to model"* (Richardson 1999 p8) and "*how can we measure the value added by the extra effort of simulation or the value lost by not simulating*?" (Coyle 2001 p359). Since the usefulness of these techniques in dealing with dynamic complexity is to be measured, it is useful to categorise tasks as either "simple" or "complex". This classification is necessary as it is widely believed that system dynamics plays a pivotal role in solving complex problems especially when the stakes are significant. The use of qualitative system dynamics alone in these situations may be insufficient (Homer and Oliva 2001).

It is proposed that there are differential effects for both, performance of qualitative system dynamics and combined system dynamics, depending on the level of complexity of the task. One may envisage that, for simple tasks, the incremental contribution of combined system dynamics will be quite small, but for complex tasks combined system dynamics will be more useful. To test the usefulness of system methodologies in dynamically complex situations, people's ability to understand fundamental concepts of dynamic complexity (stocks and flows, time delays, non-linearity, feedback loops and the dynamics that arise from them) needs to be tested before and after qualitative system dynamics and combined system dynamics interventions. Though it is possible to test the effectiveness of these tools on the same task, the additional information produced from the quantitative phase will help in understanding components of dynamic complexity. Because it is proposed that qualitative system dynamics and combined system dynamics might have differential effects on task-performance, we intend to distinguish between tasks that are "simple" and tasks that are "complex". Therefore "task-complexity" was used as the independent variable to estimate differential effects. Predictors related to task-complexity have been reported to have significant effects on task-performance (Hsiao and Richardson 1999). For instance, feedback-loops that are commonly used in system dynamics are used also as indicators that reflect task complexity (Hsiao and Richardson 1999). The strength of the feedback loop has been shown to be directly proportional to task-complexity (Hsiao and Richardson 1999) and hence to task-performance (Sterman 1989a; Sterman 1989b).

Even though several studies in the past have tried to test the efficacy of system dynamics methodologies, none has tried to experimentally ascertain the *relative* impact of the combination of qualitative and quantitative system dynamics tools over qualitative system dynamics alone. Some examples are discussed below. However only studies of adequate scientific rigor that incorporate a pre-test/post-test design are described (Doyle 1997). Others that use anecdotal evidence may be prone to errors and biases such as the 'experimenter bias' (Rosenthal 1966) or the 'Hawthorne effect' (Roethlisberger and Dickson 1939) as the researcher might have a vested interest in the outcomes.

Doyle et al. (1998) conducted a rigorous experiment to measure change in mental models using the Strategem-2 simulation game. They found that a system dynamics-based debriefing on the simulation game produced positive changes in the size and content of mental models as well as the degree of feedback thinking, although there was no significant effect on the degree of detail or dynamic complexity. Kainz and Ossimitz (2002) conducted an experiment on sixty-four undergraduates in which participants' performance on 'bathtub dynamics' tasks (Sweeney and Sterman 2000) were recorded before and after training on stocks and flows. The training did not consist of system dynamics computer modelling. Their findings suggest that performance in simple stock/flow tasks is initially quite poor. This reinforces the findings reported by Sweeny and Sterman (2002). More importantly, Kainz and Ossimitz (2002) demonstrate that the intervention significantly increased participants' abilities to deal with such tasks. However, they did not test the effectiveness of system dynamics computer modelling on such tasks.

In yet another study, Pala and Vennix (2005) evaluated participants' performance on bathtub dynamics tasks. Unlike the training provided in Kainz and Ossimitz (2002), in this case, participants were provided training on combined system dynamics. Their intervention resulted in mixed effects on task performance. For a simple task, participants performed significantly better after the intervention but their performance deteriorated for a relatively complex task. Their results indicate that the combined system dynamics intervention is useful in simple stock/flow tasks but not in relatively complex tasks. However, the experiment conducted by Pala and Vennix (2005) suffers from a few drawbacks. First, in the simple task, the experimental group (2nd year students) and the control group (1st year students) were not comparable. The incentive for both groups was different as well. Whereas completing the tasks was part of the curriculum of the course in which participants of the experimental group were enrolled, this wasn't the case for the control group. Finally, the same tasks were used for recording performance before and after the intervention. The familiarity with the task may have contributed to the improved performance.

Even though the above studies answer some questions about the effectiveness of qualitative system dynamics or combined system dynamics training, they do not contribute to the debate regarding the relative efficacy of qualitative system dynamics vs. combined system dynamics for simple and complex tasks. Relevant comparisons with these studies are made when the results are discussed later in the chapter.

3.3 Methodology

3.3.1 Introduction

The tasks, detailed procedure and basis of the experimental design are described in the following sub-sections.

3.3.2 Participants

Thirty-one postgraduate students participated in the study. All students were in their final year at the University of Sydney. The experiment was part of a course (Management Information Systems) that participants had chosen to undertake as part of their degree. Further, participants had an incentive to be seriously involved in the study as it accounted for 50% of their total assessment. One participant was excluded from the data analysis as she missed part of the intervention.

3.3.3 Design

The experiment was spread over a period of six weeks and was based on a repeated measures pre-test/post-test design which is illustrated in Figure 3.1. Participants underwent a pre-test, followed by qualitative system dynamics intervention and then were administered the post-qualitative system dynamics test. Following this, they underwent quantitative system dynamics intervention and then the post-combined system dynamics test.

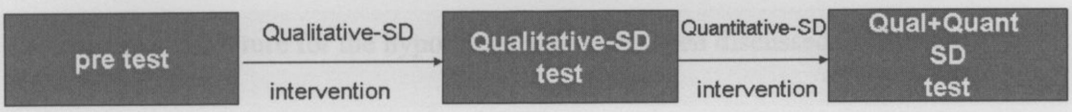


Figure 3.1: Design

3.3.4 Hypotheses

In order to answer the research questions that arise from the above discussion, the effectiveness of qualitative system dynamics tools alone and also the relative effectiveness of combined system dynamics tools over qualitative system dynamics tools need to be investigated. Hence, the following hypotheses were constructed:

H1. Participants' performance will significantly improve for both simple and complex tasks after they undergo qualitative system dynamics training, when compared with their performance without any system dynamics training, i.e.

H1.1 Qualitative system dynamics [1] >[2] Baseline[3] for simple tasks

[1] Qualitative system dynamics: Performance of participants after systems thinking intervention

H1.2 Qualitative system dynamics > Baseline for complex tasks

H2. Participants' performance will significantly improve for complex tasks but not for simple tasks after they undergo combined system dynamics training, when compared with performance after qualitative system dynamics training alone, i.e.

H2.1 Combined system dynamics[4] =[5] Qualitative system dynamics for simple task

H2.2 Combined system dynamics > Qualitative system dynamics for complex task

The empirical literature for the hypotheses above has been discussed in Chapter 2.

3.3.5 Procedure

The experiment proceeded in the following six stages:

(i) Prior to commencement: One week prior to the commencement of the experiment, participants were provided with the case study pertaining to the tasks. They were told that they would be quizzed on the scenario in the following week. They had ample time to read and re-read the scenario before the pre-test. However, they had no idea about the type of questions they would be asked.

(ii) Pre-test: All participants underwent the pre-test in the first hour. The pre-test consisted of a simple task and a complex task. Participants were seated at a sufficient distance from each other and the tests were invigilated by two instructors.

---

[2] Significant difference in results. E.g. combined system dynamics > qualitative system dynamics implies that the performance under system dynamics will be superior to performance under systems thinking

[3] Baseline: Performance of participants prior to any intervention

[4] Combined system dynamics: performance of participants after a combined system dynamics intervention

[5] No significant difference in results

(iii) Qualitative system dynamics intervention: Soon after the pre-test, participants underwent the qualitative system dynamics intervention[6]. The training included qualitative system dynamics concepts such as causal loop diagrams, feedback loops, system archetypes, stocks and flows and behaviour-over-time graphs. At the end of each session, participants were provided with handouts, notes and practice questions. The qualitative system dynamics intervention was spread over two weeks. The time devoted to face-to-face teaching on qualitative system dynamics was five hours.

(iv) Qualitative system dynamics test: At the end of the qualitative system dynamics intervention, participants were administered the post-qualitative system dynamics test. Participants were already aware that they would undergo this test. The test conditions were the same as those described in the pre-test.

(v) Quantitative system dynamics intervention: In the subsequent sessions, participants underwent the quantitative system dynamics intervention. These were taught as an extension of concepts they had been taught during the qualitative system dynamics intervention. This training included quantitative system dynamics concepts such as computer modelling and simulation. The training included hands-on training on a simulation software package Powersim Studio™. The package was used to facilitate participants' ability to create simulation models by developing, and testing 'what if' scenarios, and analysing how changes over time affect key factors—indicating which factors are important and which are not. The quantitative intervention was spread over two weeks and participants spent five hours face-to-face with the instructors.

---

[6] Interventions were taught by the researcher and the supervisor

(vi) Combined system dynamics test: At the end of the system dynamics intervention, participants were administered the post-combined system dynamics test. The conditions were the same as those in the previous two tests.

3.3.6 Tests

Each test consisted of two tasks (described below). The tests on all three occasions were similar. The inflow/outflow values and the context of the simple task were changed for each of the three tests. The complex task remained exactly the same in the three tests. However, participants did not have access to the questions after the test, nor were they aware that similar questions would be repeated in subsequent tests.

3.3.7 Tasks

Simple Task

Task 1 was a simple question on stocks and flows (see Appendix for detailed description). This task was an adaptation of a previously utilized task (Kainz and Ossimitz 2002). The aim of this task was to evaluate performance of participants in a simple task. The system involved a single stock (the number of people working on a project) with one inflow (new hires) and one outflow (people quitting). The question asked participants to specify the point in time when the number of people working on the project were highest. The task does not include other elements of dynamic complexity such as feedback, time delays or nonlinearities and hence is a typical example of a simple dynamic task.

Method of analysis – Simple Task

To test Hypotheses H1.1 and H2.1, the change in performance, strategy to solve the task and confidence of participants was measured. These three factors were assessed prior to and after each of the two interventions. To measure performance, participants' responses were analysed and categorized as either (i) correct (those that identified the correct time period) or (ii) incorrect (those who did not identify the correct time period. To get an insight into participants' strategy to solve this task, their explanation of the method of arriving at the result was assessed. There were two possible strategies by which the correct time period could be identified. The first one was to manually compute the net flow in each time period and then arrive at the largest value of stock by summing up the net flows. The second method was to merely note that the inflow exceeds the outflow until a particular time period and then this trend reverses. The largest value of the stock is when the trend reverses. Some of the participants that calculated the correct responses left the 'justification' section blank. Therefore, all correct responses are categorised under the three categories. All correct answers fell under one of these categories and this could be identified by looking at the space provided to them to justify their answers. The categories were:

1. Those that used computational strategy: These participants manually computed the result. The calculations they performed to obtain the result were done in the space provided to them.
2. Those that used visual strategy (identified the reversal of trend): These participants observed that after a certain time period the value of inflow exceeded the value of outflow. In the space provided, these participants wrote that they had identified this trend and based their answer on it.

3. Those that provided no justification (blank response): These participants arrived at the correct response but did not provide any justification for it.

Incorrect responses were analysed to reveal common errors and potential patterns. The two most common errors reported in literature are that either participant's confuse the net flow with inflow thereby choosing the time period with maximum inflow as the answer or they use the right method of arriving at the result but do not arrive at the correct result due to an arithmetic error (Kainz and Ossimitz 2002). Post-hoc analysis of the responses indicated other sources of confusion as well, such as participants confusing net flow with either the largest outflow or with the time period with largest outflow +/- inflow. An analysis of the responses led to their categorization, as below.

1. Those that used visual strategy correctly but arrive at the wrong answer
2. Those that used computational strategy correctly but arrive at the wrong answer
3. Those that identified the time period when inflow was largest
4. Those that identified the time period when outflow was largest
5. Those that identified the time period when (outflow + inflow) was largest
6. Those that identified the time period when (outflow – inflow) was largest
7. Those that identified the time period when (inflow – outflow) was largest
8. Those that added all inflows and all outflows separately and compare
9. Those that provided no justification

For each stage, participants' confidence was measured by averaging the confidence (percent) of the entire cohort.

Complex Task

The second task was also based on stocks and flows, but in addition contained feedback loops that contribute to the dynamic complexity. It is hence much more complex as compared to the first one. This task was based on a well-known dynamic problem of software projects (Abdel-Hamid and Madnick 1991). It described a scenario of a firm that provided consulting and IT services, and was experiencing human resources and revenue problems over a period of time. Participants were provided with a hypothetical situation and a graph on which they could draw their response (see Appendix for details). Specifically, participants were required to deduce the behaviour-over-time of a key factor (total workforce). Inferring the behaviour of a system from its feedback structure is considered to be one of the most fundamental uses of system methodologies (Winch 2000 and Richardson 1995). Determining the correct pattern of behaviour of the total workforce involved understanding the underlying feedback structure of the problem and subsequently determining the feedback loop that dominates over a period of one year. Figure 3.2 shows the causal loop structure of the problem.

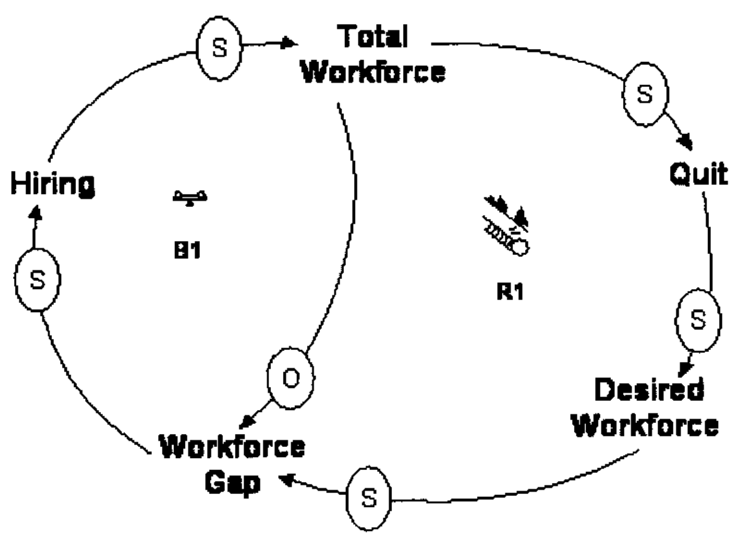


Figure 3.2: Causal loop diagram for the complex task

A hiring delay of forty days is assumed to be an average delay. The behaviour of the total workforce over one year comprises three distinct patterns. The correct pattern (Figure 3.3) was obtained by simulating the model described in Abdel-Hamid and Madnick (1991). As per the task, the total workforce on the project is ten for the first eighty days. During this time no hiring takes place. But due to the fixed outflow (quit rate) the total workforce keeps declining. This is represented by T1. As per the case study, the project manager decides to increase the total workforce to twenty on the eightieth day. This is represented as a step increase in the 'desired workforce'. This decision triggers the hiring process. Overall, the system is dominated by the hiring (balancing) loop due to which the behaviour of the total workforce follows a classic goal-seeking pattern. This is represented as T2. Once the stock reaches its desired level, the system returns to equilibrium. This phase is represented as T3. The aim of the completed task was to observe the changes in participants' performance, understanding and confidence in solving this task. This task allowed us to test hypotheses H1.2 and H2.2.

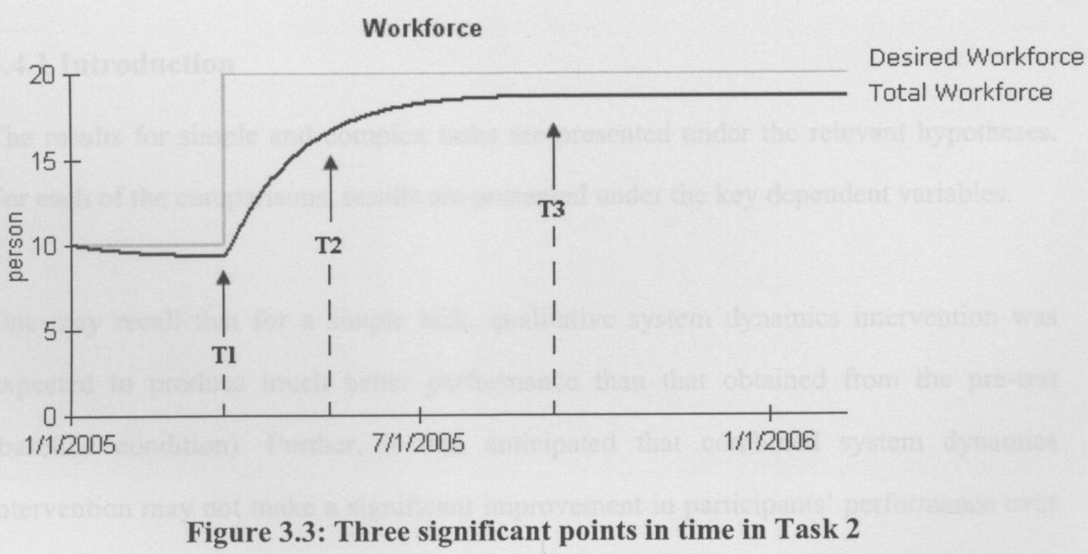


Figure 3.3: Three significant points in time in Task 2

Method of analysis – Complex Task

To measure participants' performance, answers were classified as correct if they represented the general 'pattern' across three time periods correctly—a dip in the total workforce, followed by upward goal-seeking behaviour (Figure 3.3). Average confidence was calculated in the same way as for Task 1. In addition, for this complex task, the forecasting accuracy of the variable in question was measured using the Mean Absolute Percentage Error (MAPE), which is a popular absolute error measure used by the forecasting community (Makridakis et al. 1982). Forecasting accuracy is measured across three points in the total workforce graph (T1 – T3). These three points represent significant behaviour: 1) lowest value of depreciating stock level, 2) mid-point of the goal-seeking period and 3) the point at which stock reaches equilibrium. The MAPE was used as the error measure to compare participants' responses with the ideal values.

3.4 Results

3.4.1 Introduction

The results for simple and complex tasks are presented under the relevant hypotheses. For each of the comparisons, results are presented under the key dependent variables.

One may recall that for a simple task, qualitative system dynamics intervention was expected to produce much better performance than that obtained from the pre-test (baseline condition). Further, it was anticipated that combined system dynamics intervention may not make a significant improvement in participants' performance over qualitative system dynamics. In particular, a significantly greater number of participants should be able to identify the correct time period of the stock level after the qualitative system dynamics intervention as compared to the pre-test results. However, combined system dynamics intervention should not result in a significant increase in this number. As Table 3.1 and subsequent statistical tests reveal, qualitative system dynamics intervention does significantly help participants to identify the correct time period in the simple task. Statistical tests show that combined system dynamics helps achieve further improvement, albeit not significant. Even though qualitative system dynamics intervention plays a significant role in assisting performance in a simple task, does it have a similar influence in complex tasks? Specifically, did participants infer the behaviour over time of the total workforce correctly after qualitative system dynamics intervention alone? Table 3.4 shows that they did not. Results reveal that it is the combined system dynamics intervention that helped participants in inferring the behaviour of total workforce in the complex task.

3.4.2 Simple Task

Table 3.1 presents an overview of the results from the three tests. Two key comparisons were made to test the hypotheses for the simple task. For each comparison, performance, strategy to solve the task and confidence are presented.

Table 3.1: Performance of Participants in the simple task

1. Pre-Test		**2. Qualitative System Dynamics Test**		**3. Qualitative and Quantitative System Dynamics Test**	
Performance	50.0%	Performance	80.0%	Performance	90.0%
Visual Strategy	17.0%	Visual Strategy	27.0%	Visual Strategy	70.0%
Confidence	76.0%	Confidence	88.0%	Confidence	95.0%
Ref: Baseline		**Ref: Qualitative System Dynamics**		**Ref: Qual+Quant System Dynamics**	

Legend

Performance	Percentage of participants identifying maximum value of stock	**Confidence**	Mean of confidence rating scale (0-100)
Visual Strategy	Percentage of participants using visual analysis		

3.4.2.1 Comparison: Efficacy of qualitative system dynamics as compared to baseline performance

Test for Hypothesis: H1.1

Performance: Performance of participants in the simple task before any system dynamics training was quite poor. Surprisingly, only half of the participants were able to correctly identify the answer (Table 3.2). However, qualitative system dynamics considerably helped participants in improving performance. After the introduction of this intervention, the number of participants that could identify the correct result increased by 30%. The results show statistically significant differences in performance after the qualitative system dynamics intervention, as compared with the baseline performance ($\chi^2 = 5.93$, df =1, p=0.01). This led to the acceptance of Hypothesis H1.1.

Strategy: Participants' responses in the pre-intervention test reflect fundamental flaws in their ability to understand stocks and flows. For instance, three quarters of these participants simply subtracted the inflow from outflow for each time period and compared net flows for each of them. A typical incorrect response was *"I used the number of departures to minus the number of arrivals, and then I got a new number. I compared these numbers for all periods and selected the biggest one"*. The remainder came up with other incorrect answers that were not substantiated with any justification. On the other hand, responses to the test after the qualitative system dynamics intervention reflected much better understanding. For instance, one participant wrote *"By accumulation, for example, January – March (0-0)=0; April – June (0 + 11(new) = 11 -1 (completed) = 10 (on hand); and so on"* and supported this with complete calculations to compute the time period at which the stock was largest. Further, only 17% of the participants in the pre-test adopted a visual analysis method. As seen in Table 3.2, this increased significantly by 36% after qualitative system dynamics intervention (χ^2 = 7.33, df=1, p=0.007). These results indicate that qualitative system dynamics not only improves performance in simple stock/flow tasks, it also induces a sophisticated way to deal with them.

Table 3.2: Participants' strategy to solve the simple task

	Baseline	Qual-system dynamics
Correct Responses	**50 %**	**80 %**
1. Use visual strategy	16.7 %	53.3 %
2. Use computational strategy	33.3 %	26.6 %
3. No justification	0 %	0 %
Incorrect Responses	**50 %**	**20 %**
4. Use visual strategy correctly but arrive at the wrong answer	6.6 %	3.3 %
5. Use computational strategy correctly but arrive at the wrong answer	3.3 %	0 %
6. Confound with largest inflow	0 %	0 %
7. Confound with largest outflow	0 %	3.3 %
8. Confound with largest (outflow + inflow)	0 %	3.3 %
9. Confound with largest (outflow – inflow)	3.3 %	0 %
10. Confound with largest (inflow - outflow)	30.0 %	6.6 %
11. Add all inflows and all outflows separately and compare	3.3 %	0 %
12. Blank answer	3.3 %	3.3 %
Total	**100 %**	**100 %**

Further analysis of the incorrect responses reveals that a significantly higher number of participants got confused with largest (inflow – outflow) in the pre-test (χ^2 = 5.46, df=1, p=0.02). This reduced by 23.4% after the qualitative intervention.

Confidence: Surprisingly, the overall confidence of participants was quite high prior to any intervention (Table 3.1). This was unexpected as it did not reflect participants' level of performance. This high level of confidence might have been due to the fact that the question was apparently simple and something that participants come across in their daily life. However, after the qualitative system dynamics intervention, participants' confidence increased by a further 12%. This increase might have been a consequence of the qualitative system dynamics training that helped participants formally understand the concept of stocks and flows.

3.4.2.2 Comparison: Efficacy of combined system dynamics as compared to qualitative system dynamics

Test for Hypothesis: H2.1

Performance: Participants' performance improved by 10% after combined system dynamics compared to the results obtained after qualitative system dynamics alone. This increase may not be noteworthy, but the scope for improvement was not great as the number of correct responses was already high (80%). The Chi-square results were not statistically significant ($\chi^2 = 1.18$, df=1, p=0.28). Hypothesis H2.1 is thus accepted.

Strategy: As seen in Table 3.3, combined system dynamics training led to an increase in participants using the sophisticated strategy (by 17%). However, this increase was not significant ($\chi^2 = 0.79$, df=1, p=0.37). The improvement can be attributed to the quantitative system dynamics intervention, in which participants practised various problems including identifying stocks and flows and analysing their behaviour-over-time using a computer-based modelling tool. This learning might have enhanced the ability to visualize stock/flow behaviour. For instance, one participant wrote: *"From the beginning to this time (answer) arrivals are more than departures, which means the employees are increasing. After October 2004, arrivals are smaller than departures, which means that the total number of employees is decreasing".*

Table 3.3: Participants' strategy to solve the simple task

	Qual-system dynamics	Qual+Quant-system dynamics
Correct Responses	**80 %**	**90 %**
1. Use visual strategy	53.3 %	70.0 %
2. Use computational strategy	26.6 %	20.0 %
3. No justification	0 %	0 %
Incorrect Responses	**20 %**	**10 %**
4. Use visual strategy correctly but arrive at the wrong answer	3.3 %	0 %
5. Use computational strategy correctly but arrive at the wrong answer	0 %	3.3 %
6. Confound with largest inflow	0 %	0 %
7. Confound with largest outflow	3.3 %	0 %
8. Confound with largest (outflow + inflow)	3.3 %	3.3 %
9. Confound with largest (outflow – inflow)	0 %	0 %
10. Confound with largest (inflow - outflow)	6.6 %	3.3 %
11. Add all inflows and all outflows separately and compare	0 %	0 %
12. Blank answer	3.3 %	0 %
Total	**100 %**	**100 %**

Further analysis of the correct responses reveals that there was no statistically significant difference between the usages of visual or computational method after the combined intervention (χ^2 = 2.94, df=1, p=0.086 for the visual strategy and χ^2 = 0.373, df=1, p=0.542 for the computational strategy).

Confidence: System dynamics computer modelling boosted participants' average confidence to 95% (an increase of 7%). The time spent in "self-modelling" of simple and complex issues may have led to an increased level of confidence in solving this relatively simple task.

3.4.3 Complex task

Table 3.4 presents an overview of the results from the three tests. Two key comparisons were made to test the hypotheses for the complex task. For each comparison, performance, forecasting accuracy, feedback thinking and confidence are presented.

Table 3.4: Performance of participants in the complex task

1. Pre-Test		**2. Qualitative System Dynamics Test**		**3. Qualitative and Quantitative System Dynamics Test**	
Performance	0%	Performance	3.0%	Performance	67.0%
Confidence	52.0%	Confidence	48.0%	Confidence	70.0%
Forecasting Accuracy	24.7%	Forecasting Accuracy	25.0%	Forecasting Accuracy	9.7%
Ref: Baseline		**Ref: Qualitative System Dynamics**		**Ref: Qual+Quant System Dynamics**	

Legend

Performance	Percentage of participants inferring the correct behaviour	**Confidence**	Mean of confidence rating scale (0-100)
Understanding	Percentage of participants identifying feedback loops	**Forecasting Accuracy**	Mean Absolute Percentage Error (MAPE)

3.4.3.1 Comparison: Efficacy of qualitative system dynamics as compared to baseline performance

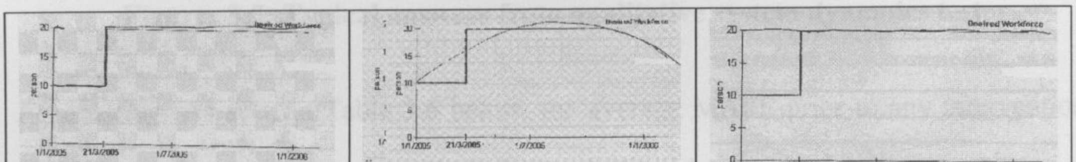
Test for Hypothesis: H1.2

Performance: As shown in Table 3.5, none of the participants were able to correctly infer 'total workforce', i.e. the three time periods as represented in Figure 3.3, prior to any system dynamics intervention.

Table 3.5: Correctness

	Baseline	**Qual-system dynamics**
Correct (Both segments correct)	**0 %**	**3.3 %**
Incorrect	**100 %**	**97 %**
1. Only one segment correct	0 %	10.0 %
2. No segment correct	96.7 %	80.0 %
3. Blank answer	3.3 %	6.7 %
Accuracy (MAPE)	**24.7%**	**25.0%**

At this stage, participants made fundamental errors, such as ignoring inflows/outflows, accumulation and feedback loops. For instance, the majority of those that got both segments incorrect infer the behaviour of the total workforce the same as that of desired workforce. Desired workforce was only one of the variables required to correctly infer total workforce. One of the reasons for this was that none of the participants took 'quit rate' (outflow) and average time delay into account which resulted in a step-like graph. This sort of response is common in complex decision-making problems, where participants are not able to process many variables together and eventually think that the key input variable will reflect the output. A few participants drew the total workforce as either linear or as linear growth followed by a linear decay. This may be due to the fact that they concentrated more on the positive reinforcing loop that produces exponential growth/decay. In the process they completely ignored the effect of the negative loop that produces goal-seeking behaviour which dominates this system. In fact, none of the responses reflected participants' understanding of 'goal seeking' behaviour. Typical responses from the pre-test are shown in Figure 3.4.

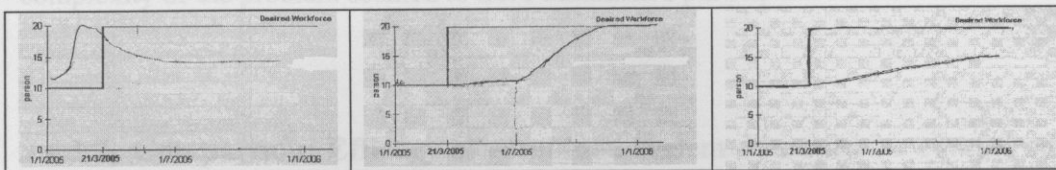


3.4(a)	3.4(b)	3.4(c)

Figure 3.4: Typical responses from pre-test

Participants were taught concepts of feedback, delays and stock/flows in the qualitative system dynamics training intervention and hence it was expected that performance would significantly improve even after qualitative system dynamics intervention. However, it did not. Only 3.3% participants recognized the correct behaviour of total workforce. The change was not statistically significant (χ^2 = 1.02, df=1, p=0.31). Therefore Hypothesis H1.2 is rejected. However, there were some improvements in participants' responses when compared to those in the pre-test. For instance, 80% participants drew both the segments incorrectly—17% less than in the pre-test. Many participants acknowledged the contribution of a time delay. This led to some participants drawing a goal seeking curve. For instance in the first segment, 10% of the participants acknowledged the effect of 'quit rate' (outflow) on the 'total workforce' (stock). In the second segment, 20% drew a goal-seeking curve. These graphs resembled the simulated graph more than responses from the pre-test. Some typical examples are shown in Figure

3.5 below. The results indicate that qualitative system dynamics did make some changes to participants' thinking that resulted in more participants representing the quit rate and goal seeking behaviour than those in the pre-test.



3.5(a)	3.5(b)	3.5(c)

Figures 3.5: Typical answers from qualitative system dynamics test

Forecasting: As seen in Table 3.6 below, the average MAPE prior to any intervention (24.7) was almost the same as that obtained after the qualitative system dynamics intervention (25.0). T-tests between MAPEs at time T1, T2 and T3 for pre-test and Qual-test reveal that there was no significant difference between forecasting accuracy between the two tests (for T1, t=0.8620 and p=0.3963; for T2, t= -1.8370 and p=0.0772; and for T3, t=-1.3005 and p=0.2044). Hence it can be concluded that the qualitative system dynamics intervention did not help in improving participants' ability to forecast the behaviour of the total workforce.

Table 3.6: Forecast accuracy

	Baseline	**Qual-system dynamics**
Accuracy (MAPE)	**24.7%**	**25.1%**
1. T1	36.8%	28.3%
2. T2	22.3%	26.8%
3. T3	15.1%	20.1%

Confidence: Qualitative system dynamics intervention did not help in boosting confidence—in fact, confidence fell by 4%, which was surprising as one would assume that 'understanding' had improved during this stage. Similar to the results reported for the simple task, the average confidence with which participants responded to this question was unexpectedly high—in spite of their poor performance. However, average confidence in the complex task was less than that reported in Task 1 (Table 3.1). The complexity of the problem seemed to have diminished participants' confidence.

3.4.3.2 Comparison: Efficacy of combined system dynamics as compared to qualitative system dynamics

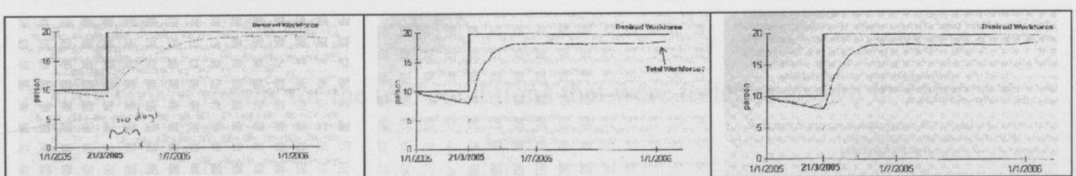
Test for hypothesis: H2.2

Performance: As expected, participants' understanding and performance increased significantly after the combined system dynamics intervention. Participants were trained in stock and flow modelling of problems and used simulation software to generate behaviour over time. Most participants in this task were able to identify important aspects of the behaviour of the total workforce. 67% percent recognized and represented outflow in the first time interval. In the subsequent segment, 70% drew a goal-seeking curve. This was a significant increase when compared to that of the qualitative system dynamics intervention (Table 3.7). The proportion of participants that represented the entire behaviour correctly over a period of one year was 67% (compared to 10% after qualitative system dynamics alone). The combined system dynamics intervention produced statistical improvement compared to the qualitative system dynamics intervention ($\chi^2 = 26.45$, df=1, $P<0.001$), leading to the acceptance of Hypothesis H2.2. The reason why there was a significant improvement was almost certainly due to improvement in forecasting the individual segments.

Table 3.7: Correctness

	Qual-system dynamics	**Qual+Quant-system dynamics**
Correct (Both segments correct)	**3.3 %**	**66.7 %**
Incorrect	**97 %**	**33.3 %**
1. Only one segment correct	10.0 %	0 %
2. No segment correct	80.0 %	33.3 %
3. Blank answer	6.7 %	0 %
Accuracy (MAPE)	**25.0%**	**9.7 %**

Chi-square tests on incorrect responses show that a significantly higher number of participants did not get even one of the segments correct in the Qual-test as opposed to those in the Qual+Quant-test (χ^2 = 13.30, df=1, p=0.0003). Typical responses from the combined system dynamics test are shown below in Figure 3.6.



3.6(a) | **3.6(b)** | **3.6(c)**

Figure 3.6: Typical answers from combined system dynamics test

Forecasting Accuracy: As compared to post-qualitative system dynamics, the average MAPE shows significant improvement after the combined system dynamics intervention. From a high 25.0, it is almost halved to 9.7 (Table 3.8). A paired T-test analysis reveals that the quantitative training resulted in a significant improvement in forecasts at time T1 (t=2.6032; p=0.0148) and very significant improvement at times T2 (t=4.0212; p=0.0004) and T3 (t=4.2762; p=0.0002). A possible explanation of this improvement is the fact that more participants were able to model the system correctly after combined system dynamics training which led to an accurate projection of the total workforce. This eventually reduced the forecasting error.

Table 3.8: Forecast accuracy

	Qual-system dynamics	**Qual+Quant-system dynamics**
Accuracy (MAPE)	**25.1%**	**9.7%**
1. T1	28.3%	8.2%
2. T2	26.8%	14.1%
3. T3	20.1%	6.7%

Confidence: Average confidence after the combined system dynamics intervention rose considerably compared to the level after the earlier qualitative system dynamics intervention (by 22%). It seems that the use of formal modelling, simulation and hypothesis testing led to increased understanding and hence helped in boosting the overall confidence of the group.

A summary of results for the two conditions that were tested is shown in Table 3.9.

Table 3.9: Summary of results

Question	Expectation	Result	Hypotheses
Efficacy of qualitative system dynamics as compared to baseline performance	Simple: Qualitative-SD* > Baseline	Simple: Qualitative SD > Baseline	H1.1 - Accepted
	Complex: Qualitative-SD > Baseline	Complex: Qualitative SD = Baseline	H1.2 - Rejected
Efficacy of combined system dynamics as compared to qualitative system dynamics	Simple: Qual+Quant-SD = Qualitative-SD	Simple: Qual+Quant-SD = Qualitative-SD	H2.1 - Accepted
	Complex: Qual+Quant-SD > Qualitative-SD	Complex: Qual+Quant-SD > Qualitative-SD	H2.2 - Accepted

SD* - System Dynamics

3.5 Discussion

Previous experimental research suggests that our native ability to solve simple stock/flow tasks is quite poor (Sweeney and Sterman 2000; Kainz and Ossimitz 2002; Pala and Vennix 2005). These studies further show that this shortcoming can be overcome by employing qualitative system dynamics interventions. However, these studies are lacking in two respects. One, these experiments did not measure the extra value added by quantitative system dynamics (combined qualitative and quantitative system dynamics) to performance, and second, they did not test the usefulness of qualitative system dynamics or the combined system dynamics for complex tasks (that are ubiquitous in businesses). The objective of the current study was to determine the influence of qualitative system dynamics and combined system dynamics on performance in tasks that involve components of dynamic complexity. Specifically, the experiment was designed to record participants' performance, understanding and confidence in both simple and complex tasks, before and after qualitative system dynamics and combined system dynamics training. Hence, the purpose of this study was two-pronged—first to determine the efficacy of qualitative system dynamics compared with no system dynamics training and second, to establish the relative effectiveness of combined system dynamics over qualitative system dynamics under two conditions of complexity.

The results of the current study suggest that in the absence of any system dynamics training, human understanding of even simple stock/flow systems that have no feedback is originally quite poor. This finding reinforces the results previously obtained by Sweeney and Sterman (2000). Specifically, participants were confused between stocks and flows, and ended up with incorrect responses. Those who were able to identify the correct response did so by using a time-consuming method.

Results suggest that this deficiency can be dealt with by qualitative system dynamics training. Specifically, those who underwent qualitative intervention performed significantly better than those who did not undergo any system dynamics intervention for the simple task. A possible explanation of this improvement is that during the qualitative system dynamics training, participants learnt the concept of accumulation which might have helped them in solving this task. The intervention led to fewer participants making logical errors in computing the answer. Further there was an increase in the number of participants that used the visual method instead of the time-consuming computational method. Overall, the results indicate that qualitative system dynamics tools are sufficient to significantly improve performance in simple stock/flow tasks.

The results reported corroborate with those obtained by Kainz and Ossimitz (2002). The level of expertise of the participants, task complexity and amount of training provided in Kainz and Ossimitz (2002) were similar to those used in the current study. The comparison of results is depicted in Figures 3.7 (a) and (b) below.

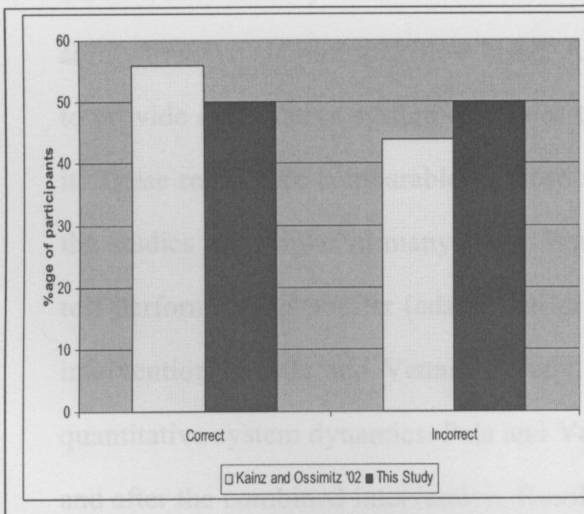


Figure 3.7(a): Pre-test comparison with Kainz and Ossimitz (2002)

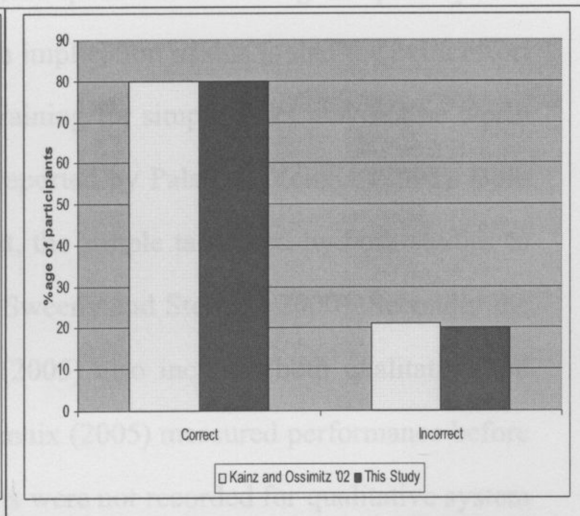


Figure 3.7(b): Qualitative comparison with Kainz and Ossimitz (2002)

Figure 3.7: Comparison with Kainz and Ossimitz (2002)

As indicated by the graphs above, the results of both studies bear sizeable resemblance. Chi-square analyses on the pre-test as well as Qualitative test reveals that there are no statistically significant differences between results from the two studies (χ^2 = 0.32, df=1, p=0.57 for pre-test and χ^2 = 0.001, df=1, p=0.97 for Qualitative test). This reinforces the fact that qualitative system dynamics skills assist in improving performance on simple stock/flow tasks. The improvement in performance and understanding due to qualitative system dynamics is encouraging and implies that qualitative system dynamics training alone is valuable.

As opposed to results reported for qualitative system dynamics intervention, combined system dynamics intervention did not lead to any significant improvement in either performance, or in the use of sophisticated methods or in confidence, on top of the results achieved through qualitative intervention alone for the simple task. A possible explanation of the insignificant increase is that performance and understanding were already quite high after the qualitative system dynamics intervention and hence the scope for improvement wasn't great. The small improvement in performance due to combined system dynamics training suggests that a weak link may exist between additional quantitative training and participants' ability to solve simple stock/flow tasks. An implication of this is that the extra effort to provide quantitative system dynamics training for simple tasks may not be worth it. These results are comparable to those reported by Pala and Vennix (2005). Both the studies are similar in many ways. First, the simple task used by both studies to test performance is similar (adapted from Sweeny and Sterman 2000). Secondly the intervention in Pala and Vennix's study (2005) also includes both qualitative and quantitative system dynamics. Pala and Vennix (2005) measured performance before and after the combined intervention. Results were not recorded for qualitative system dynamics intervention. Hence only the pre- and post-intervention results of their experimental group can be compared with pre-test and combined system dynamics test in the present study. These are depicted in Figure 3.8 below. As seen, pre-test results bear a resemblance in both studies ($\chi^2 = 1.51$, df=1, p=0.22). However, the results from the combined system dynamics test in both studies bear significant differences ($\chi^2 = 9.91$, df=1, p=0.002).

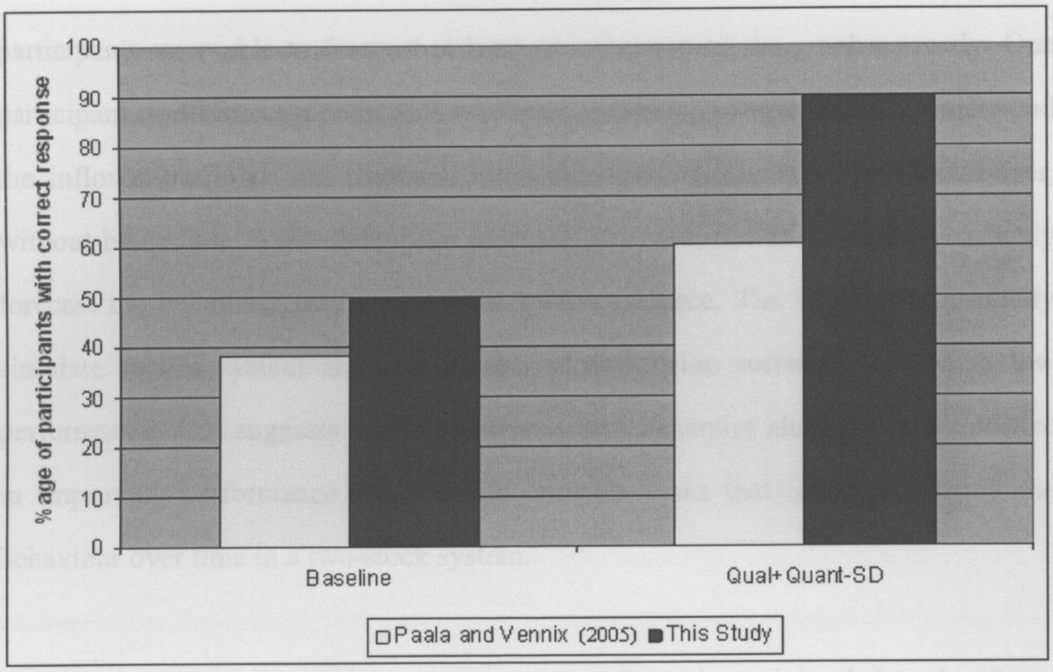


Figure 3.8: Comparison of results with those of Pala and Vennix (2005)

Both studies reinforce the fact that system dynamics training brings about significant changes to participants' abilities to deal with simple stock/flow tasks.

As in the case of the simple task, baseline performance for the complex task was also sub-optimal. For instance, before the intervention, participants drew graphs that had absolutely no resemblance to the correct output. The responses reflected their inability to grapple with the complexity of the task. Specifically, participants struggled with identifying stocks and flows and got confused between the two. In contrast to the results reported for the simple task, qualitative system dynamics intervention did not lead to a significant improvement in performance for the complex task. The system structure in the complex task could be represented by feedback loops, which many participants should have correctly identified. This can be attributed to the fact that participants might have learnt how to draw causal loops that were taught during the qualitative system dynamics training. Some of the graphs showed some similarity with that of the simulated output. For instance, 10% of the participants were able to forecast at least one segment of the graph correctly. One participant could forecast both. This was some indication that participants understood the inflows, outflows and feedback loops operating within the system. However, without being able to simulating the feedback structure, it was difficult to correctly forecast the behaviour over time of the total workforce. The inability to mentally simulate such a system and the absence of simulation software resulted in low performance. This suggests that qualitative system dynamics alone might not suffice in improving performance in relatively complex tasks that involve inferring the behaviour over time in a two-stock system.

As expected, the addition of quantitative system dynamics training led to significant improvement in performance and understanding of participants in the complex task. These results are consistent with anecdotal evidence that exists in support of system dynamics computer simulation. A possible explanation of the improved result is the ability of participants to be able to reinforce some qualitative system dynamics concepts and to simulate the system using Powersim. The conceptual understanding of the system gained through qualitative system dynamics, combined with computer simulation, plays a vital role in tackling complex tasks and inferring correct behaviour of key variables. These results suggest that the addition of quantitative modelling did add significantly to the understanding achieved by qualitative analysis. An implication of this result is that it might be worth investing time and money in quantitative system dynamics training/modelling when task complexity is high.

In contrast to the performance recorded prior to the system dynamics interventions, participants' confidence was quite high for the simple task. Participants remained equally (over) confident of this task even after the two system dynamics interventions. This inconsistency is reported in the literature (e.g. Cronin et al. 2007; Peterson and Eberlein 1994; Fischhoff et al. 1977; Lichtenstein and Fischhoff 1977). On the other hand, participants' confidence for the complex task was not as high as that reported for the simple task. There was no significant difference in average confidence after the qualitative intervention. It may be recalled that even performance had followed the same trend. Average confidence improved considerably (from 50% to 70%) after the combined intervention. Performance too, improved significantly at this stage. This indicates that at least for complex tasks, participants' feelings of confidence were a realistic reflection of their performance.

The forecasting accuracy of the graphs followed the same trend as that of performance. The MAPE for baseline performance and that recorded after qualitative test were essentially the same. The high MAPE was not surprising as the performance of participants was dismal. As a result of the quantitative system dynamics intervention, 64% more participants were able to get the correct output graph. This means that these graphs were more like the ideal graph. As a result, forecasting accuracy doubled, suggesting that the addition of quantitative training also led to a significant improvement in forecasting abilities of participants.

3.6 Conclusion

The first study set out to determine the efficacy of qualitative system dynamics tools compared with baseline performance and a combination of qualitative and quantitative tools. It was hypothesised that those who underwent qualitative system dynamics training would improve upon their performance for both simple and complex tasks. Further, it was expected that the additional quantitative system dynamics training would not significantly improve this performance for simple tasks but it would for complex tasks.

The most obvious finding to emerge from this study was that qualitative system dynamics tools alone are not sufficient to infer behaviour over time in complex tasks. Combined system dynamics tools are necessary to accomplish such tasks. On the other hand, results suggest that the additional effort spent on teaching and learning of quantitative system dynamics tools is not required to solve simple stock/flow tasks. It was shown that qualitative system dynamics tools are sufficient to solve these simple tasks. The results also reinforce that, in general, participants' native ability to deal with components of dynamic complexity is quite poor. The interventions did not make any significant changes to participants' confidence in the simple tasks. Overall, the findings from this study add to our understanding of the additional role of system dynamics computer modelling in solving simple and complex tasks.

Finally, a number of important limitations of the present study need to be considered. The most important limitation lies in the fact that the same cohort underwent the pre-test and subsequent post-tests, i.e. an experimental control was absent. Secondly, as discussed before, this experiment used a repeated measures design where the same participants underwent three almost similar tests. Since the same tasks were used to assess performance before and after the interventions, improvement in results could be partly attributed to repeated exposure to the two tasks. These gains are also referred to as learning effects or practice effects (Siders et al. 2006). Specifically, this improvement could be a result of an increased familiarity with the tasks, experience in solving these tasks and the development of strategies in solving the tasks. For instance, after the pre-test, participants could have remembered the presentation format of the two tasks. In the subsequent tests, they might have then felt comfortable understanding the question. Similarly, participants might have figured out a successful strategy to solve the tasks which they might have repeated in subsequent tests. Thirdly, it needs to be noted that the results obtained in this case after the qualitative system dynamics intervention reflect the performance of those who have never been exposed to system dynamics modelling. In real-life practice however, many experts who apply qualitative system dynamics tools, already have knowledge of system dynamics modelling, which changes their approach to problem solving. Finally, the design of this study did not separately measure performance after system dynamics computer modelling alone, i.e. without the knowledge of qualitative system dynamics tools. This test is important as many practitioners claim that the computer modelling phase alone might be sufficient for both simple and complex tasks.

Further research in the following chapter is required to formulate an experimental design that nullifies the learning effect and also allows the testing of quantitative system dynamics alone. A comparison of qualitative system dynamics alone, quantitative system dynamics alone, and a combination of the two is needed to add to the growing debate on 'when to model and when to map'. This forms the basis of the experiment discussed in the next chapter.

Chapter 4

Experiment 2

4.1 Introduction

"The field must address the relationships between qualitative mapping and quantitative modelling — in short, when to map and when to model" (Richardson 1999, p8)

In the previous chapter (Chapter 3), the usefulness of qualitative system dynamics (SD) tools and a combination of qualitative and quantitative tools was compared with each other and with baseline performance. The results showed that for simple tasks, performance after qualitative system dynamics was significantly better as compared to performance before system dynamics training. The addition of quantitative system dynamics training did not improve performance significantly over that achieved through qualitative system dynamics. For the complex task, however, performance significantly improved after quantitative system dynamics training was provided. The previous experiment lacked two important aspects. Firstly, the design did not allow the testing of quantitative system dynamics tools in isolation, i.e. comparison of quantitative system dynamics with (i) qualitative system dynamics, (ii) combined system dynamics and (iii) with baseline performance could not be made. Secondly, the repeated measures design used in the previous experiment and the repeated exposure to the same cover story and tasks in all the three tests might have resulted in a 'learning effect'. The current study addresses both these issues.

To compare the relative efficacy of the three system dynamics tools, a rigorous experimental study was conducted involving 80 participants. Qualitative and quantitative interventions were administered and responses were recorded for simple and complex tasks. In the following section (Background), first quantitative system dynamics is discussed. Subsequently, a section is devoted to describe the approach that was used to ensure the elimination of the learning effect. The experimental design, tasks and procedure are explained in detail in the Methodology section. The results are then presented and contrasted with those obtained by previous studies (Results). The final section (Conclusion) presents an overall summary of the findings and discusses the limitations of this study.

4.2 Background

As discussed in Chapter 2, system dynamics was mainly perceived and used as a quantitative simulation technique when it was first introduced. For instance, the first work on system dynamics in book form by J.W. Forrester (1961) did not contain any causal loop diagrams and feedback structure was depicted by equations or stock-and-flow diagrams. Quantitative system dynamics allows us to map a system's behaviour over time by allowing computer-based simulation. Its obvious benefit is that it combines human thinking and computational power to allow for a significant extension to qualitative system dynamics. Limitations of the quantitative method include not having enough or valid data (Wolstenholme 1999), incorporation of soft variables (Coyle 2000), generation of misleading results owing to uncertain variables (Coyle 2000), generation of more complexity than required, and the necessity for expert use of such a system (Wolstenholme 1999). The pros and cons of using qualitative system dynamics and a combination of qualitative and quantitative system dynamics have been discussed in the previous chapter (Chapter 3).

From being widely used as a quantitative technique, system dynamics evolved into incorporating the use of qualitative methods—to the extent that at times, qualitative methods alone are applied as a modelling tool (such as Wolstenholme 1983; Wolstenholme 1998; Cavana et al. 1999; J Coyle et al. 1999). The debate regarding the use of qualitative and/or quantitative methods has been highlighted in Coyle

(2000). The relationship between the qualitative and the quantitative phases has yet to be properly discerned (Richardson 1996). While some researchers support the use of qualitative methods as a first step towards quantitative modelling (Wolstenholme 1999), others maintain that qualitative and quantitative methods can be applied independently.

Pertinent questions arise regarding the use of qualitative system dynamics versus quantitative system dynamics owing to mixed results in the literature such as the efficacy of these methods, the extent of their use and the context of their use. These form the motivation to derive the hypotheses for the present study. In particular, we are interested in assessing how effective qualitative and quantitative methods are individually and when used together. The current study also examines how qualitative and quantitative compare with each other and with their combination. Finally, the study explores the usefulness of qualitative and quantitative methods in the context of simple and complex tasks. General perspectives in the area lead to the following suppositions: a) The use of qualitative and quantitative methods leads to increased performance when used individually and when used together, b) a combination of qualitative and quantitative methods leads to higher performance than the approach of quantitative methods applied independently, which is in turn more useful than the approach of qualitative methods applied independently to solve complex tasks and c) the above (b), may not be true for simple tasks that may not require the rigor of quantitative methods.

As stated earlier and discussed in detail previously, the design of experiment reported in Chapter 3 might have led to the improvement in participants' performance due to repeated exposure to the cover story and tasks. In the current study, three comparable cover stories were used to ensure that participants did not benefit from repeated exposure to the same cover story and tasks. Secondly, the design in this experiment used a different but comparable cohort to measure performance after each system dynamics intervention. Thirdly, the design allows the explicit testing of the learning effect at three stages. These are discussed in detail later in the chapter.

In addition to the above mentioned changes in design, the current study uses a comprehensive method of examining participants' responses from the complex task. It aims to measure the change in participants' mental models by assessing the level of feedback thinking prior to and after each intervention. Feedback thinking is a key element to understand such complex tasks and the concept of mental models has been vital to the field of system dynamics since its inception (Doyle and Ford 1998). However, as mental models are fuzzy, measuring them has always posed a challenge in the past (Richardson and Pugh 1981). According to Sterman (1994), mental models are "vastly simplified compared to the complexity of the systems themselves" and "dynamically deficient". System dynamics tools claim to improve mental models of decision-makers by making them more accurate, complete and able to reflect the dynamic nature of systems. Hence it is important to measure changes in mental models of participants prior to, and after, interventions. According to Senge (1990), there are two types of complexities exhibited by mental models—detailed and dynamic. Detailed complexity relates to the content of mental models whereas dynamic complexity reflects feedback thinking. system dynamics claims to change the latter.

4.3 Methodology

4.3.1 Introduction

The experiment was spread over a period of two weeks and was based on a pre-test/post-test design. Participants were divided into four cohorts that underwent qualitative and/or quantitative interventions. Their performance was recoded subsequently. The tasks, detailed procedure and basis of the experimental design are described in the following sub-sections.

4.3.2 Participants

Eighty postgraduate students enrolled at the University of Sydney participated in this study. Forty-eight of these were enrolled in Masters by coursework degrees and thirty-two in a PhD degree. All participants had at least one year of prior full-time work experience. Further, they had all studied mathematics at undergraduate level.

The study was voluntary and participants were paid for their participation. Sixty participants were required for two days and the remaining twenty for one day (see the experimental design explanation below). Those who participated on both the days were paid Australian $150. Those who participated for one day were paid $75. This amount was paid after the completion of the experiment. Participants were also offered additional monetary incentive based on their performance in the study. They were told that the top 50% of the group would be paid an additional $25. Once the experiment was conducted, participants' answer sheets were examined. The top ten (50%) of each group were identified. These participants were subsequently notified and payment was made to them.

4.3.3 Design

It may be recalled that the experiment was designed to test the effectiveness of qualitative and quantitative tools, both, individually and with each other, for performance in simple and complex tasks. The design of the experiment is explained in detail in Figure 4.1.

Participants were randomly assigned to one of four groups (A, B, C or D) each consisting of 20 participants. Each group had 8 participants enrolled in a Ph.D. degree and 12 enrolled in Masters by coursework. There was no significant difference in performance between those enrolled in PhD or Masters. These results are discussed later in the chapter.

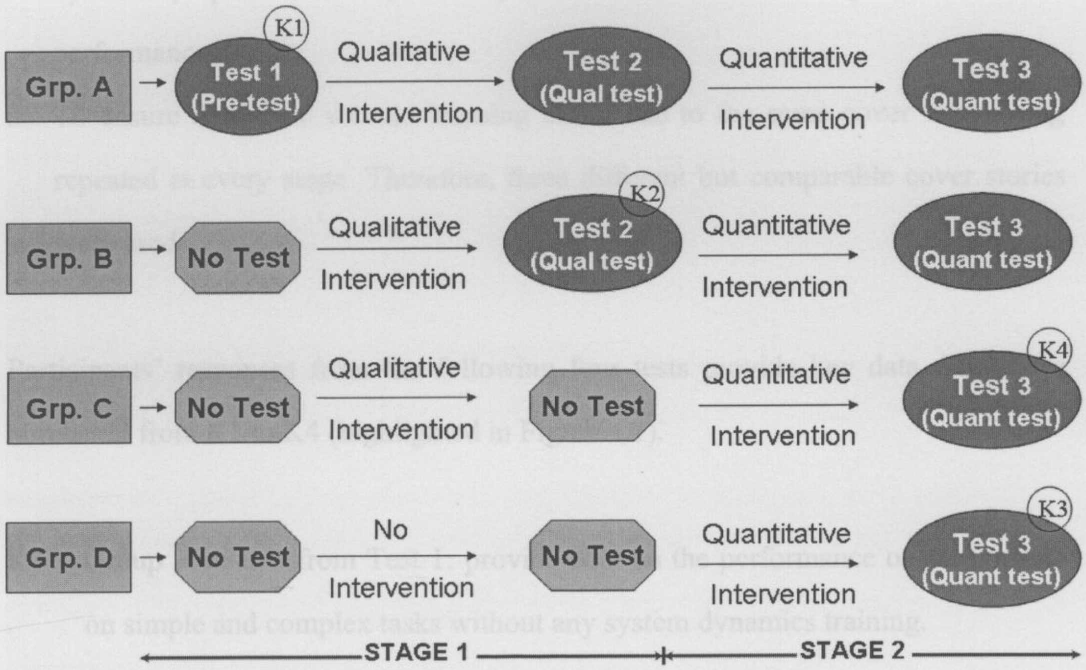


Figure 4.1: Experiment design

The experiment was divided into two stages, each lasting 8 hours. There was a gap of one week between the two stages. The first stage was used to test qualitative intervention and the second stage was used to test both quantitative and the combined intervention. Apart from re-testing the usefulness of qualitative and a combined system dynamics intervention, as done in Chapter 3, the design of this experiment was used for the following.

1. To test the usefulness of quantitative system dynamics intervention when compared with (i) baseline performance (ii) qualitative system dynamics intervention and (iii) combined system dynamics intervention.
2. To ensure that there is no learning effect when participants progress from one stage to the other, due to the same cohort being used repeatedly. Therefore, four different but comparable cohorts were used to test baseline, qualitative system dynamics, quantitative system dynamics and combined system dynamics performance.
3. To ensure that there was no learning effect due to the same cover story being repeated at every stage. Therefore, three different but comparable cover stories were used.

Participants' responses from the following four tests provide key data. These are numbered from K1 to K4 (highlighted in Figure 4.1).

K1. Group A results from Test 1: provide data on the performance of participants on simple and complex tasks without any system dynamics training.

K2. Group B results from Test 2: provide data on the performance of participants on simple and complex tasks with qualitative system dynamics training. Up till this stage Group B did not undergo any pre-test and hence there was no possibility of learning due to similarity of questions.

K3. Group D results from Test 3: provide data on the performance of participants on simple and complex tasks with quantitative system dynamics training. Up till this stage Group D did not undergo any pre-test and hence there was no possibility of learning due to similarity of questions.

K4. Group C results from Test 3: provide data on the performance of participants on simple and complex tasks with a combination of qualitative and quantitative system dynamics training. Up till this stage Group C did not undergo any pre-test and hence there was no possibility of learning due to similarity of questions.

4.3.4 Hypotheses

Based on the experimental design and the results from the previous experiment, the following hypotheses may be tested:

H1. Participants who attended qualitative system dynamics training will perform better on simple tasks but not on complex tasks when compared to participants who did not undergo any system dynamics training, i.e.,

H1.1 Qual[1] >[2] Baseline[3] for simple tasks

H1.2 Qual = Baseline for complex tasks

These hypotheses were tested by comparing results from K1 and K2.

H2. Participants who attended quantitative system dynamics training will perform better on both simple and complex tasks when compared to participants who did not undergo any system dynamics training i.e.,

H2.1 Quant[4] > Baseline for simple tasks

H2.2 Quant > Baseline for complex tasks

These hypotheses were tested by comparing results from K1 and K3.

H3. Participants who attended both qualitative and quantitative system dynamics training will perform better on both simple and complex tasks when compared to participants who did not undergo any system dynamics training, i.e.,

---

[1] Qual: Results obtained after qualitative system dynamics intervention

[2] Significant difference in results. E.g. Quant > Qual implies that the performance under quantitative system dynamics will be superior to performance under qualitative system dynamics

[3] Baseline: Performance of participants prior to any intervention

[4] Quant: Results obtained after quantitative system dynamics intervention

H3.1 (Qual, Quant)[5] > Baseline for simple tasks

H3.2 (Qual, Quant) > Baseline for complex tasks

These hypotheses were tested by comparing results from K1 and K4.

H4. There will not be any significant difference between results obtained from qualitative, quantitative and combined intervention for a simple task. But for a complex task, results obtained from the combined intervention will be significantly higher than those from quantitative intervention, which in turn will be significantly higher than those obtained after qualitative intervention, i.e.

H4.1 (Qual, Quant) = Quant = Qual for simple tasks

H4.2 (Qual, Quant) > Quant > Qual for complex tasks

These hypotheses were tested by comparing results from K2, K3 and K4.

In addition to the key data mentioned above, the other tests were useful to determine if taking pre-test(s) had any impact on participants' performance. These are identified as L1 to L4 below.

L1. Group A results from Test 2: provide data on the performance of participants on simple and complex tasks with qualitative system dynamics training. However, before Test 2, Group A underwent a pre-test (Test 1).

L2. Group A results from Test 3: provide data on the performance of participants on simple and complex tasks with a combination of qualitative and quantitative

[5] (Qual, Quant) = Results obtained after a combination of qualitative and quantitative system dynamics intervention

system dynamics training. However, before Test 3, Group A underwent two pre-tests (Test 1 and Test 2).

L3. Group B results from Test 3: provide data on the performance of participants on simple and complex tasks with a combination of qualitative and quantitative system dynamics after training. However, before Test 3, Group B underwent a pre-test (Test 2).

Based of the data described above, the following hypotheses may be tested:

1. Effect of taking pre-test on performance in Qual-test. These were tested by comparing results from L1 and K2.
2. Effect of taking pre-test on performance in Quant-test. These were tested by comparing results from L2 and L3.
3. Effect of taking pre-test and Qual-test on performance in Quant-test. These were tested by comparing results from L2 and K4.
4. Effect of taking Qual-test on performance in Quant-test. These were tested by comparing results from L3 and K4.

4.3.5 Tests

There were three tests in all (identified as pre-test, Qual-test and Quant-test in Figure 4.1). Each test consisted of two tasks that are described in the next sub-section. The tasks in each test were based on a cover story (case studies and tests are provided in the Appendix). Three different but comparable cover stories were used (explained in detail below). This was necessary to remove the 'learning effect' that might contribute towards an improvement in performance.

Group A was administered Test 1 for the pre-test. All cohorts (A and B) were administered Test 2 for the Qual Test. Similarly, all cohorts (A, B, C and D) were administered Test 3 for the Quant Test. These tests were not randomly allocated to the cohorts because there was a possibility that participants would chat to one another in the coffee breaks and lunch break. This could have resulted in questions and/or answers being leaked.

For the simple task, participants were required to identify the time period when the value of stock reaches its maximum. To do this, the inflows and outflows to the stock were provided. Even though the cover stories used in the three tasks were different, the tasks were very similar to each other. In all the three tests, participants were required to identify the time period when the value of stock reaches its maximum. The system in all the three cases involved the analysis of one stock. The other factors involved were a single inflow that increased the value of the stock and a single outflow that decreased the value of the stock.

For the complex task, participants were required to draw the behaviour over time of a key factor. The background information to arrive at this graph was provided using a cover story. As discussed above, there were three equally complex cases. Prominent factors that contributed towards the complexity of the cover story/task were same in all three cases. For instance, the output in all the three cases comprised two main segments. The number of stocks and other variables required to construct a computer model to achieve the correct graph were the same as well. Further, all the three problems could be represented via one reinforcing feedback loop and one balancing feedback loop. These factors made the three complex tasks comparable. The tasks are described in the following sub-section.

4.3.6 Tasks

Simple Task

Task 1 was a simple question on stocks and flows (see Appendix for details). The task and the method of analysing responses from this task have been described in detail in Chapter 3.

Complex Task

The second task was also based on stocks and flows, but in addition, contained factors that contributed to dynamic complexity, such as feedback loops (Sterman 2002a). One of the tasks is described in Chapter 3. The remaining tasks are briefly described below.

Task used for Test 1

The case study concerned consumer adoption of a new product. Specifically, it detailed how a mobile phone manufacturer targeted potential customers in a town. The problem was that even though the company had been in similar circumstances before, it did not have knowledge of the amount of phones it might be able to sell in one year. This had led to late deliveries in the past which resulted in customers being unsatisfied. Information about the behaviour of sales during 12 months prior would have helped the company in improving their forecasts for production of phones. The specific details of the factors that influence sales, such as the number of potential customers, effect of word of mouth and the effect of advertising were provided. The task for participants was to draw a behaviour over time graph of sales over a period of one year on the space provided to them. In addition to this, participants also had to list all the factors that influence sales and how these factors related to each other. The correct behaviour over time was obtained by simulating the model as described in

Paich and Sterman (1993) and is shown below (Figure 4.2). Participants were required to indicate their confidence on a nine-point Likert scale.

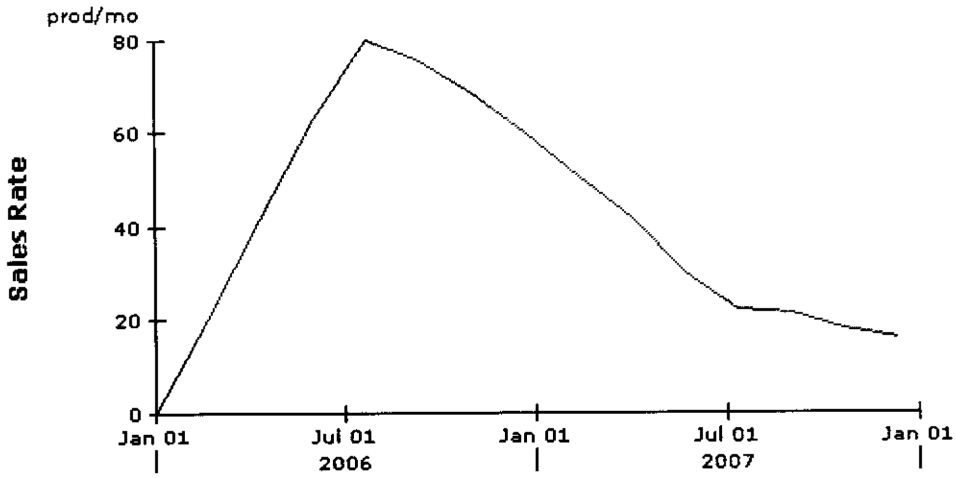


Figure 4.2: Simulated result of complex task in Test 1

Task used for Test 2

The case study concerned population boom in a growing economy. Specifically, it detailed how a country town was growing in terms of industry and population due to the discovery of natural resources in the area. According to the case, the mayor of the city would like to understand the behaviour of the growth of population over the next 50 years so that he could plan facilities such as schools and hospitals in advance. Knowledge of the way population might behave was important for him as during his previous stints, the mayor did not effectively plan for sufficient schools for children which led to unhappy residents. The specific details of the factors that influence total population, such as the number of young and adult population at the time, the effect of immigration and emigration and average birth/death rates were provided. The task for participants was to draw a behaviour over time graph of population over a period of fifty years on the space provided to them. In addition to this, participants also had to list all the factors that influence population and how these factors related to each other. The correct behaviour over time was obtained by simulating the model as described in Roberts et al. (1981) and is shown below (Figure 4.3).

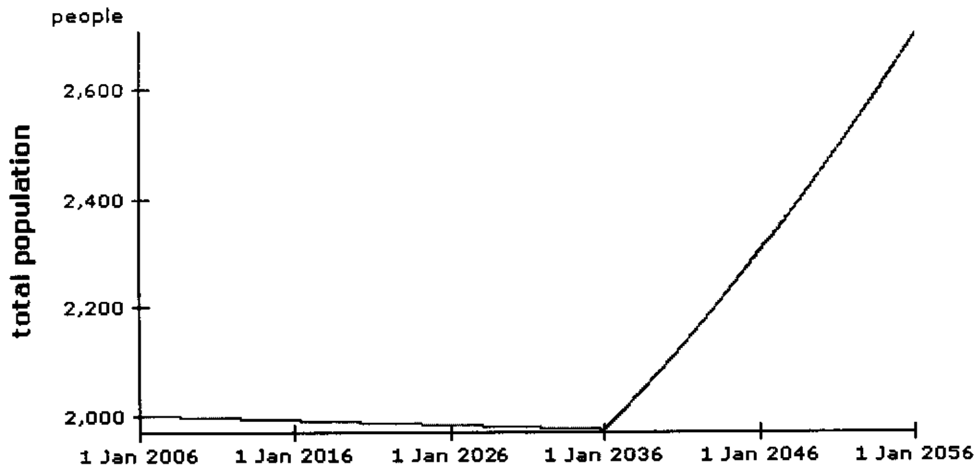


Figure 4.3: Simulated result of complex task in Test 2

Task used for Test 3

The complex task used in Test 3 was the same as that used in the previous experiment. This has been discussed in detail in the previous chapter (Chapter 3).

4.3.7 Method of analysis

The data collected from the complex tasks were analysed from various perspectives. The overall correctness of the graph was first assessed to ascertain participants' performance for this task. The graphs were then compared with the ideal result to calculate forecasting accuracy. Performance and forecasting were measured in the same way as that described in Chapter 3. It may be recalled that the correctness of the graph was assessed by comparing the shape of the graph with that of the ideal response. Each of the three output graphs consisted of two main segments. The output of Test 1 consisted of a linear growth in sales that peaks and is subsequently followed by a linear decline. In Test 2, population decreased linearly and then rose exponentially. In the last test (Test 3) workforce decreased linearly and then followed a positive goal-seeking behaviour. Secondly, the accuracy of the graphs was measured by calculating the MAPE at three points in time.

The method of analysing participants' 'understanding' and 'confidence' is discussed below.

Understanding of feedback

The main purpose of a system dynamics study is to infer the behaviour over time of a key variable from the feedback loop structure. Both, the feedback loop structure and the behaviour over time graph were separately analysed. This aspect of the analysis is new for this experiment and was not done for the previous experiment.

To gain an understanding of whether participants understood feedback loop structure, they were asked to list cause and effect relationships between all elements that had an influence on the variable in question. This was done to gain insight into the problem solving approach and measure changes in mental models.

The system in each of the case studies used in the present study could be represented by two feedback loops—one reinforcing and other balancing. Knowing about these was a logical step towards inferring the behaviour over time of the key variable. Hence, participants' responses were checked to see if they included the correct feedback loops that represent the dynamics of the system. As there is no single way to measure dynamic complexity, guidelines from Maani and Maharaj (2004) were adapted to be applied in the context of this study. All answers were coded in one of five categories of decreasing complexity namely responses containing:

(i) Two relevant feedback loops

(ii) One relevant feedback loop only

(iii) Cause and effect relationships but not culminating in a feedback loop

(iv) Only a list of variables

(v) No answer

Participants in the first two categories were classified as those who exhibited 'feedback thinking' and the others were classified as those who exhibited 'linear thinking'. The feedback loops could have been described either as text, by using a causal loop diagram or through a stock and flow diagram. However, these text and diagrams could be interpreted in different ways by different persons. Hence to assess the reliability of the coding, all responses were coded by a second coder. Once the coding scheme was agreed upon, the two coders coded a few randomly selected responses to make sure that they understood the coding scheme correctly. Both coders then individually categorised all responses into one of the five categories. The inter-rater reliability (Kendall's Tau) of such categories was .90 for pre-test, .95 for Qual-test and .97 for the Quant-test.

Confidence

Participants' confidence on the task was assessed using a 9-point Likert scale (see Appendix). Participants' responses on the scale were grouped under three categories.

1. 'Those who found the task easy'—participants that selected options 1 to 4 (Extremely Easy, Very Easy, Moderately Easy and Slightly Easy) were assigned to this category.
2. 'Those who found the task difficult'—participants that selected options 6 to 9 (Slightly Difficult, Moderately Difficult, Very difficult and Extremely Difficult) were assigned to this category.
3. 'Neither easy nor difficult'—participants who selected option 5 (Neither Easy nor Difficult) were assigned to this category.

4.3.8 Procedure

The following steps describe the process by which the experiment was conducted on the first day.

1. 60 (Groups A, B and C, see Figure 4.1) out of the total 80 participants arrived at 9 a.m. outside the experiment hall. The remaining 20 participants (Group D) came only on the second day.
2. Participants registered for the experiments and were randomly allocated an experiment identifier (e.g. A1, A2 etc.).
3. Group A participants then took the pre-test. They were administered Test1 and were provided one hour to answer the two tasks discussed previously. The test was supervised by two people. While the Group A participants completed Test 1, participants from groups B and C waited outside the computer lab.
4. After a short break, the experiment re-commenced with a short introduction to the day. All 60 participants (Groups A, B and C) participated in the study from this stage onwards.
5. Participants then underwent training on qualitative system dynamics tools. These were in the form of interactive lectures and included many practice questions. The main concepts covered during the lesson were the systems view, interrelationship between elements of a system, causality, causal loop diagrams, system archetypes, stocks and flows and behaviour over time graphs. These tools were collectively used to analyse various examples. Both individual and group activities were performed. A detailed outline of the concepts that were covered during qualitative training is provided in the Appendix.
6. At the end of the intervention, Groups A and B (total 40 participants) underwent Test 2. Test 2 was administered in the same way as Test 1. At this time participants from group C (20 in number) waited outside and were relieved after a short debriefing.
7. Following the test, Groups A and B were relieved after debrief. This marked the end of the first day.

The second day of the experiment was held one week after the first stage. This stage of the experiment involved the same 60 participants (Groups A, B and C) from the previous week as well as 20 new participants (Group D, see Figure 4.1). Hence a total of 80 participants participated at this stage.

The following steps describe the process by which the experiment was conducted on the second day.

1. All 80 participants (Groups A, B, C and D) arrived at 9 am and registered for the experiment.
2. The day commenced with a brief introduction to the experiment. This was particularly useful for the 20 participants who were new to the experiment.
3. All groups attended training in quantitative system dynamics. As in the previous week's session, this involved teaching and learning in an interactive setting with many practice examples that were solved both individually as well as in a group. Participants were familiarised with system dynamics software (Powersim™)[6] that enabled them to model scenarios. A detailed outline of the concepts that were covered during quantitative training is provided in the Appendix.
4. The intervention was followed by Test 3. All participants took the test. This time participants used Powersim as well as the written exam format to answer the questions.
5. Test 3 marked the end of the experiment's teaching/learning and testing phases.
6. Participants were then remunerated for their time as detailed before. Additional incentives were paid the following week.

[6] Powersim and Powersim™ have been used interchangeably in the thesis to refer to the system dynamics simulation software Powersim Studio that is a product of Powersim Software AS.

4.4 Results

4.4.1 Introduction

In the following section, results are presented first for simple task and then for the complex task. Following these subsections, the tests for learning effect and between those enrolled in PhD and Masters are presented.

4.4.2 Simple task

Table 4.1 presents an overview of the results from the three tests. Five key comparisons were made to test the hypotheses for the simple task. For each comparison, performance, strategy to solve the task and confidence are presented.

Table 4.1: Results for simple task

Group	1. Pre-Test		2. Qualitative Test		3. Quantitative Test	
A	Performance	35%	Performance	70%	Performance	85%
	Visual strategy	10%	Visual strategy	35%	Visual strategy	80%
	Confidence	85%	Confidence	95%	Confidence	95%
	Ref: A1 - Baseline (K1)		**Ref: A2**		**Ref: A3**	
B	NA		Performance	75%	Performance	95%
			Visual strategy	35%	Visual strategy	80%
			Confidence	90%	Confidence	100%
			Ref: B2 - Qualitative Only (K2)		**Ref: B3**	
C	NA		NA		Performance	90%
					Visual strategy	80%
					Confidence	95%
					Ref: C3 - Qualitative & Quantitative (K4)	
D	NA		NA		Performance	70%
					Visual strategy	45%
					Confidence	85%
					Ref: D3 – Quantitative Only (K3)	

Segments of the table that provide key data are shaded, and marked as K1 – K4.

Legend

Performance	Percentage of participants identifying maximum value of stock (simple)	**Confidence**	Percentage of participants that found the task 'easy'
Visual strategy	Percentage of participants using visual analysis		

4.4.2.1 Comparison: Efficacy of qualitative system dynamics as compared to baseline performance

Reference to main results table (Table 4.1): B2 vs. A1

Test for Hypothesis: H1.1

Performance: Prior studies have noted the importance of qualitative system dynamics in solving simple stock/flow tasks. In the current study too, qualitative system dynamics did make significant improvement in the performance of participants. Performance of participants without any system dynamics intervention was quite poor. Only 35% were able to correctly identify the time period at which the stock reaches its maximum. However, the qualitative intervention significantly improved participants' performance. The number of participants who could identify the correct result increased by 40% ($\chi^2 = 6.47$, df = 1, P=0.01). For example, after the qualitative intervention, one participant wrote *"The employee number equals to the remaining number plus the arrivals and minus the departures"* and supported this with a complete calculation to compute the time period at which the stock was largest. These results led to the acceptance of Hypothesis 1.1.

Incorrect responses: As seen in Table 4.2, in the pre-test, 20% of the participants used the correct logic to solve the problem but failed to achieve the correct result. These either used the computational method (15%) or the visual method (5%). 45% of the total participants could not solve this task at all. Their responses reflected fundamental errors in their ability to understand stocks and flows. For instance, 10% could not solve the task and simply subtracted the inflow from outflow for each time period and compared net flows for each of them. A typical incorrect response was *"I* *compared the number of new projects for each quarter, and choose the maximum that is 92 for October-December 2001"*. One-fifth of the participants identified the time period with the largest outflow as the response. Participants did not make these errors after the qualitative intervention. In both the tests, however, participants confounded the correct answer with the time period representing largest (inflow – outflow).

Table 4.2: Participants' strategy to solve the simple task

Reference to Figure 4.1: B2 vs. A1	Qualitative	Baseline
Correct Responses	**75%**	**35%**
1. Use visual strategy	35%	10%
2. Use computational strategy	40%	25%
3. No justification	0%	0%
Incorrect Responses	**25%**	**65%**
4. Use visual strategy correctly but arrive at the wrong answer	0%	5%
5. Use computational strategy correctly but arrive at the wrong answer	0%	15%
6. Confound with largest inflow	0%	0%
7. Confound with largest outflow	0%	20%
8. Confound with largest (outflow + inflow)	0%	15%
9. Confound with largest (outflow – inflow)	0%	0%
10. Confound with largest (inflow - outflow)	25%	10%
11. Add all inflows and all outflows separately and compare	0%	0%
12. Blank answer	0%	0%
Total	**100 %**	**100 %**
N=20		

Strategy: Even though the correct response could be identified by simple visual analysis, only 10% of the participants who arrived at the correct response adopted this method in the pre-test. However, in the Qual-test, 35% of the participants who identified the correct answer used the visual way of identifying the correct time period. However, this increase was not significant ($\chi^2 = 3.59$, df = 1, P=0.06).

Confidence: Though overall performance in the pre-test was quite poor and so was understanding of the concepts, surprisingly, the majority of the participants claimed they found this task easy to solve. In the Qual-test as well, the majority of the participants found the task easy. However, at this stage, the high confidence was substantiated with higher performance as well. The Chi-square analysis does not reveal any significant differences between the confidence reported in the two tests ($\chi^2 = 2.50$, df = 1, P=0.11)

4.4.2.2 Comparison: Efficacy of quantitative system dynamics as compared to baseline performance

Reference to main results table (Table 4.1): D3 vs. A1

Test for Hypothesis: H2.1

Performance: As mentioned before in the previous comparison, performance of participants without any system dynamics intervention is quite poor in the simple task. As compared to only 35% in the pre-test, participants who underwent quantitative system dynamics did much better to correctly identify the time period at which the stock reaches its maximum. The Chi-square results show a significant improvement in performance ($\chi^2 = 4.91$, df = 1, P=0.03). It was hypothesised that quantitative intervention will significantly improve participants' ability to solve simple stock/flow tasks when compared to baseline performance. The results indicate that it does. Therefore hypothesis H1.2 is accepted.

Incorrect responses: One third of participants who underwent quantitative intervention and one-fifth of the participants who did not undergo any intervention used the right method of arriving at the results but failed to identify the correct answer (Table 4.3). As previously described, during the pre-test, participants confounded the correct answer with that of the largest outflow, largest (outflow + inflow) and largest (inflow - outflow). However, after the quantitative intervention, participants did not make any of these errors.

Table 4.3: Participants' strategy to solve the simple task

Reference to Figure 4.1: D3 vs. A1	Quantitative	Baseline
Correct Responses	**70%**	**35%**
1. Use visual strategy	45%	10%
2. Use computational strategy	25%	25%
3. No justification	0%	0%
Incorrect Responses	**30%**	**65%**
4. Use visual strategy correctly but arrive at the wrong answer	10%	5%
5. Use computational strategy correctly but arrive at the wrong answer	20%	15%
6. Confound with largest inflow	0%	0%
7. Confound with largest outflow	0%	20%
8. Confound with largest (outflow + inflow)	0%	15%
9. Confound with largest (outflow – inflow)	0%	0%
10. Confound with largest (inflow - outflow)	0%	10%
11. Add all inflows and all outflows separately and compare	0%	0%
12. Blank answer	0%	0%
Total	**100 %**	**100 %**
N=20		

Strategy: In terms of understanding, 45% of the participants used the visual way to arrive at the answer after the quantitative intervention, as compared to only 10% that used this method prior to the intervention. The improvement in the use of the visual method was however not significant ($\chi^2 = 2.39$, df = 1, P<0.2).

Confidence: Confidence level of participants was the same at this stage compared to that after the pre-test. The chi square analysis does not show any significant difference between the two tests ($\chi^2 = 0.00$, df = 1, P=1)

4.4.2.3 Comparison: Efficacy of combined qualitative and quantitative system dynamics compared to baseline performance

Reference to main results table (Table 4.1): C3 vs. A1

Test for Hypothesis: H3.1

Performance: The results obtained from participants who underwent the combination intervention indicate a significant improvement in performance as compared to the pre-test (χ^2 = 12.9, df = 1, P=0.0003). Ninety percent of the participants got the correct result as compared to only 35% that were previously able to correctly identify the time period when the stock reaches its maximum. These results show that people have difficulty in discerning between stocks and flows and calculating the net value of a stock over a given time period but this ability greatly improves after they learn system dynamics' qualitative and quantitative techniques. Based on the fact that individual interventions would result in a significant improvement in performance in the simple task, it was hypothesised that the combination intervention too would produce similar results when compared to baseline performance. The current study found that the combination intervention is reasonably effective in improving performance. Hypothesis H3.1 is therefore accepted.

Incorrect responses: The number of participants who used the right strategy but resulted in a wrong answer decreased after the combined intervention. After the combined intervention, only one participant who used the right method could not arrive at the correct answer. As opposed to results of the pre-test where 45% participants who confused the right answer with other options, only 5% of total participants seemed to be confused. Detailed results are reported in the table below (Table 4.4).

Table 4.4: Participants' strategy to solve the simple task

Reference to Figure 4.1: C3 vs. A1	Qualitative+ Quantitative	Baseline
Correct Responses	**90%**	**35%**
1. Use visual strategy	80%	10%
2. Use computational strategy	10%	25%
3. No justification	0%	0%
Incorrect Responses	**10%**	**65%**
4. Use visual strategy correctly but arrive at the wrong answer	0%	5%
5. Use computational strategy correctly but arrive at the wrong answer	5%	15%
6. Confound with largest inflow	0%	0%
7. Confound with largest outflow	0%	20%
8. Confound with largest (outflow + inflow)	0%	15%
9. Confound with largest (outflow – inflow)	0%	0%
10. Confound with largest (inflow - outflow)	5%	10%
11. Add all inflows and all outflows separately and compare	0%	0%
12. Blank answer	0%	0%
Total	**100 %**	**100 %**
N=20		

Strategy: The combination intervention also helped participants to identify the correct answer using the visual approach with as many as 80% of the participants using this strategy (as compared to 10% in pre-test). For instance, one participant wrote: *"because after 1908, out-migration exceeded in-migration"*. The Chi-square results show a significant shift towards the visual method (χ^2 = 19.8, df = 1, P<0.0001).

Confidence: Confidence in solving this task was extremely high with 95% of the participants indicating the task as easy to solve. However, there are no significant differences from the confidence reported in the pre-test. (χ^2 = 1.11, df = 1, P=0.30)

4.4.2.4 Comparison: Efficacy of quantitative system dynamics as compared to qualitative system dynamics

Reference to main results table (Table 4.1): D3 vs. B2

Test for hypothesis: H4.1

Even though both quantitative and qualitative intervention produced significant improvement in performance when compared to the pre-test, there were no statistically significant differences between the two with respect to performance, use of visual method and confidence (performance: χ^2 = 0.13, df = 1, P=0.72; understanding: χ^2 = 0.91, df = 1, P=0.34; confidence: χ^2 = 0.23, df = 1, P=0.63). A third of participants who underwent quantitative intervention used the right strategy but failed to arrive at the correct result (Table 4.5).

Table 4.5: Participants' Strategy to Solve the Simple Task

Reference to Figure 4.1: D3 vs. B2	Quantitative	Qualitative
Correct Responses	**70%**	**75%**
1. Use visual strategy	45%	35%
2. Use computational strategy	25%	40%
3. No justification	0%	0%
Incorrect Responses	**30%**	**25%**
4. Use visual strategy correctly but arrive at the wrong answer	10%	0%
5. Use computational strategy correctly but arrive at the wrong answer	20%	0%
6. Confound with largest inflow	0%	0%
7. Confound with largest outflow	0%	0%
8. Confound with largest (outflow + inflow)	0%	0%
9. Confound with largest (outflow – inflow)	0%	0%
10. Confound with largest (inflow - outflow)	0%	25%
11. Add all inflows and all outflows separately and compare	0%	0%
12. Blank answer	0%	0%
Total	**100 %**	**100 %**
N=20		

None of the participants who underwent qualitative intervention faltered in this category. On the other hand, none of the participants who underwent quantitative intervention got confused with the common sources of errors. A quarter of those who underwent qualitative intervention confounded the correct response with the largest (inflow - outflow).

4.4.2.5 Comparison: Efficacy of combined qualitative and quantitative system dynamics as compared to qualitative system dynamics only and quantitative system dynamics only

Reference to main results table (Table 4.1): C3 vs. B2 and B3

Test for hypothesis: H4.1

It was hypothesised that participants who undergo the combined intervention would not perform significantly better than those who undergo qualitative only or quantitative only. The results obtained indicate that this is true. The chi square analysis between the combined intervention and qualitative intervention shows insignificant differences (χ^2 = 1.56, df = 1, P=0.21). Similarly, the comparison with quantitative intervention also fails to show significant improvements (χ^2 = 2.50, df = 1, P=0.11). Hypothesis 4.1 is hence accepted. 90% of the participants were able to identify the correct time period after the combined intervention. This was a 15% increase in performance from the results obtained from qualitative intervention alone and a 20% increase from results obtained after a quantitative intervention alone (Table 4.6). As can be seen from the table, there was a 20% increase in the number of participants who got confused with largest (inflow – outflow). However, Chi square analysis of these statistics reveal that this difference is not significant (χ^2 = 1.71, df = 1, P=0.18).

Table 4.6: Participants' Strategy to Solve the Simple Task

Reference to Figure 4.1: C3 vs. D3 vs. B2	Qualitative+ Quantitative	Quantitative	Qualitative
Correct Responses	**90%**	**70%**	**75%**
1. Use visual strategy	80%	45%	35%
2. Use computational strategy	10%	25%	40%
3. No justification	0%	0%	0%
Incorrect Responses	**10%**	**30%**	**25%**
4. Use visual strategy correctly but arrive at the wrong answer	0%	10%	0%
5. Use computational strategy correctly but arrive at the wrong answer	5%	20%	0%
6. Confound with largest inflow	0%	0%	0%
7. Confound with largest outflow	0%	0%	0%
8. Confound with largest (outflow + inflow)	0%	0%	0%
9. Confound with largest (outflow – inflow)	0%	0%	0%
10. Confound with largest (inflow - outflow)	5%	0%	25%
11. Add all inflows and all outflows separately and compare	0%	0%	0%
12. Blank answer	0%	0%	0%
Total	**100 %**	**100 %**	**100 %**

4.4.3 Complex task

Table 4.7 presents an overview of the results from the three tests. Five key comparisons were made to test the hypotheses for the complex task. For each comparison, performance, forecasting accuracy, feedback thinking and confidence are presented.

Table 4.7: Results for the complex task

Group	1. Pre-Test		2. Qualitative Test		3. Quantitative Test	
A	Performance	0%	Performance	35%	Performance	70%
	Forecasting Accuracy	57.7%	Forecasting Accuracy	69.2%	Forecasting Accuracy	11.6%
	Feedback Thinking	5%	Feedback Thinking	95%	Feedback Thinking	85%
	Confidence	10%	Confidence	30%	Confidence	75%
	Ref: A1 - Baseline (K1)		**Ref: A2**		**Ref: A3**	
B	NA		Performance	15%	Performance	65%
			Forecasting Accuracy	66.0%	Forecasting Accuracy	11.6%
			Feedback Thinking	70%	Feedback Thinking	75%
			Confidence	25%	Confidence	75%
			Ref: B2 - Qualitative Only (K2)		**Ref: B3**	
C	NA		NA		Performance	65%
					Forecasting Accuracy	8.8%
					Feedback Thinking	75%
					Confidence	75%
					Ref: C3 - Qualitative & Quantitative (K4)	
D	NA		NA		Performance	50%
					Forecasting Accuracy	11.3%
					Feedback Thinking	60%
					Confidence	50%
					Ref: D3 – Quantitative Only (K3)	

Segments of the table that provide key data are shaded, and marked as K1 – K4.

Performance	Percentage of participants inferring the correct behaviour (complex)	**Confidence**	Percentage of participants that found the task 'easy'
Understanding	Percentage of participants identifying feedback loops	**Forecasting Accuracy**	Mean Absolute Percentage Error (MAPE)

4.4.3.1 Comparison: Efficacy of qualitative system dynamics as compared to baseline performance

Reference to main results table (Table 4.7): B2 vs. A1

Test for hypothesis: H1.2

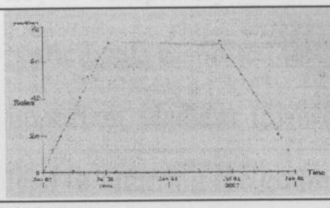
Performance: As shown in Table 4.8, in the pre-test, none of the participants could project 'sales' correctly before any systems intervention. Incorrect graphs included plateau shaped, linearly increasing graphs, s-shaped and goal-seeking curves. Five percent of the total participants did not attempt to answer the question.

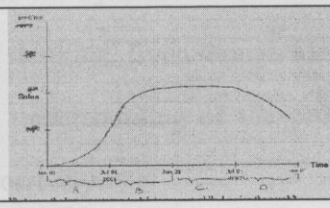
Table 4.8: Correctness

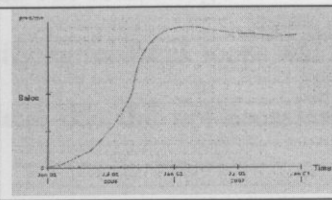
Reference to Figure 4.1: B2 vs. A1	**Qualitative**	**Baseline**
Correct (Both segments correct)	**15%**	**0%**
Incorrect	**85%**	**100%**
1. Only one segment correct	25%	10%
2. No segment correct	60%	85%
3. Blank answer	0%	5%
N=20		

Further statistical analyses of incorrect responses reveal that there were no significant differences between participants who drew one segment correctly ($\chi^2 = 0.69$, df = 1, P=0.41), those who drew none of the segments correctly ($\chi^2 = 2.01$, df = 1, P=0.16) and between those who did not draw any graph ($\chi^2 = 1.03$, df = 1, P=0.31).

Figure 4.4 shows typical responses from this stage. The results obtained from the qualitative system dynamics intervention were almost as poor as the pre-test (χ^2 = 3.24, df = 1, P=0.07). Only 15% of the participants could draw the correct behaviour of 'total population'. The behaviour was simulated as linear followed by a linear drop in population followed by exponential growth. Like the results from pre-test, here too, majority of responses bore no similarity with the simulated graph. Based on the findings of the previous experiment, it was hypothesised that there would not be any significant changes in performance after a qualitative system dynamics intervention. The results of this experiment indicate that this indeed was the case. Hence, hypothesis H1.2 is accepted.


4.4 (a)


4.4 (b)


4.4 (c)

Figures 4.4: Typical responses from pre-test

Forecasting Accuracy: Both the pre-test as well as the Qual-test showed an extremely high average MAPE (Table 4.9). In both the tests participants' performance was quite low. Incorrect responses meant that the accuracy of their graphs when compared to the ideal output was quite low.

Table 4.9: Forecast accuracy

Reference to Figure 4.1: B2 vs. A1	**Qualitative**	**Baseline**
Accuracy (MAPE)	**66.0%**	**57.7%**
1. T1	37.8%	70.2%
2. T2	80.2%	58.7%
3. T3	79.9%	44.4%
N=20		

Feedback Thinking: Measures of participants' understanding reflected on their naïve mental models before any system dynamics intervention. This was measured by looking at participants' causal loop diagrams and text. For the pre-test, participants' ability to think in terms of feedback loops was extremely poor. Only 5% of the total participants could correctly identify the two feedback loops operating in the system (Table 4.10). More than half the participants merely listed the variables. Another 40% exhibited cause and effect relationships only. In the qualitative test, there was a significant increase in the level of dynamic thinking ability in participants. As many as 70% of the participants recognized the key feedback loops operating in the system (χ^2 = 18.03, df = 1, P=0.0001). Out of these 40% could identify both the loops—a key to completely understand the dynamics of the system. The remaining 30% of the participants exhibited linear thinking. Even though identifying feedback loops was an important element towards understanding of the problem, this did not necessarily lead to inferring the graph correctly.

Table 4.10: Feedback thinking

Reference to Figure 4.1: B2 vs. A1	**Qualitative**	**Baseline**
Feedback Thinking	**70%**	**5%**
1. Two feedback loops	40%	5%
2. One feedback loop only	30%	0%
Linear Thinking	**30%**	**95%**
3. Cause and effect relationships only	20%	40%
4. List of variables only	10%	55%
5. Blank answer	0%	0%
Total	**100 %**	**100 %**
N=20		

Confidence: Unable to infer the correct behaviour of the key factor and overwhelmed by the complexity, the results from the confidence scale were not surprising. Only 10% of the participants found the task easy to solve in the pre-test. Unlike in the simple task, participants' reported a much lower confidence in the complex task. 20% more participants felt that the task was easy after the qualitative intervention. However, the Chi-square analysis of this change does not yield significant results (χ^2 = 1.56, df = 1, P=0.21).

4.4.3.2 Comparison: Efficacy of quantitative system dynamics as compared to baseline performance

Reference to main results table (Table 4.7): D3 vs. A1

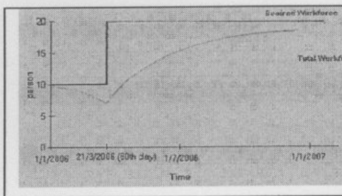
Test for hypothesis: H2.2

Performance: After the pre-test, none of the participants correctly projected the sales on the graph. However, after the quantitative intervention, half of these participants were able to draw the correct behaviour (Table 4.11). Chi-square results confirm that participants who underwent quantitative system dynamics did much better than those with no system dynamics skills ($\chi^2 = 13.3$, df = 1, P=0.0003).

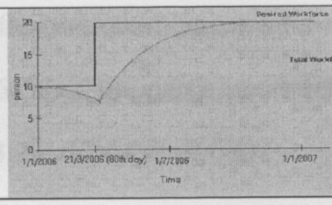
Table 4.11: Correctness

Reference to Figure 4.1: D3 vs. A1	**Quantitative**	**Baseline**
Correct (Both segments correct)	**50%**	**0%**
Incorrect	**50%**	**100%**
1. Only one segment correct	5%	10%
2. No segment correct	40%	85%
3. Blank answer	5%	5%
N=20		

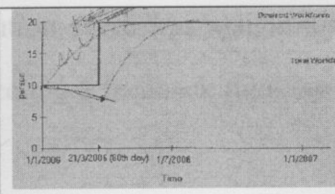
Figure 4.5 shows typical responses from the Quant-test. It was expected that the quantitative intervention would produce a significant improvement in performance when compared to the baseline performance. The findings show that performance significantly improved, leading to the acceptance of hypothesis H2.2.



4.5 (a)



4.5 (b)



4.5 (c)

Figures 4.5: Typical responses from quant-test

Forecasting Accuracy: Forecast accuracy significantly improved after the quantitative intervention when compared with that reported in the pre-test (MAPE dropped from 57.7% to 11.3%). Detailed results are shown in Table 4.12.

Table 4.12: Forecast Accuracy

Reference to Figure 4.1: D3 vs. A1	Quantitative	Baseline
Accuracy (MAPE)	**11.3%**	**57.7%**
1. T1	15.4%	70.2%
2. T2	14.1%	58.7%
3. T3	4.5%	44.4%
N=20		

Feedback Thinking: As previously stated, only 5% of the total participants correctly identified the two feedback loops operating in the system in the pre-test.

Table 4.13: Feedback Thinking

Reference to Figure 4.1: D3 vs. A1	Quantitative	Baseline
Feedback Thinking	**60%**	**5%**
1. Two feedback loops	50%	5%
2. One feedback loop only	10%	0%
Linear Thinking	**40%**	**95%**
3. Cause and effect relationships only	30%	40%
4. List of variables only	10%	55%
5. Blank answer	0%	0%
Total	**100 %**	**100 %**
N=20		

As shown in Table 4.13, after the quantitative intervention, there was a significant change in this figure with 60% of the participants reflecting a feedback thinking in their mental models ($\chi^2 = 13.79$, df = 1, P=0.0002).

Confidence: As compared to only 10% that had previously found the task easy to solve, the results from the Likert scale show that 50% found the task easy after the quantitative intervention ($\chi^2 = 7.61$, df = 1, P=0.06). These results were a true reflection of their overall performance as well.

4.4.3.3 Comparison: Efficacy of combined qualitative and quantitative system dynamics compared to baseline performance

Reference to main results table (Table 4.7): C3 vs. A1

Test for Hypothesis: H3.2

Performance: As expected, understanding and performance by participants who underwent the combination intervention was significantly higher than pre-test results (Table 4.14).

Table 4.14: Correctness

Reference to Figure 4.1: A1 vs. C3	**Qualitative+ Quantitative**	**Baseline**
Correct (Both segments correct)	**65%**	**0%**
Incorrect	**35%**	**100%**
1. Only one segment correct	5%	10%
2. No segment correct	30%	85%
3. Blank answer	0%	5%
N=20		

Most participants were able to identify the important aspects of the behaviour of 'total workforce'. These participants were able to draw the correct behaviour, represented inflow and outflow correctly and drew a goal-seeking curve (Table 4.14). The combination intervention produced statistically significant differences when compared with baseline performance (χ^2 = 19.26, df =1, P<0.0001) in this task. This leads to the acceptance of Hypothesis H3.2.

Forecasting Accuracy: Forecasting accuracy increased significantly as compared to baseline performance. The value of the MAPE fell to 7.8 from 64.1. This suggests that the combination intervention made a marked difference in the ability of participants to infer the behaviour over time (Table 4.15).

Table 4.15: Forecast Accuracy

Reference to Figure 4.1: D3 vs. A1	**Qualitative+ Quantitative**	**Baseline**
Accuracy (MAPE)	**8.7%**	**57.7%**
1. T1	11.8%	70.2%
2. T2	10.3%	58.7%
3. T3	4.2%	44.4%
N=20		

Feedback Thinking: Participants' mental models also reflected the improvement in performance (Table 4.16). The level of feedback thinking increased significantly, from 5% to 75% (χ^2 = 20.42, df = 1, P=0.0001). The majority of the participants justified their graphs using feedback loops operating in the system. Participants drew the loops either in the answer sheets or by using Powersim software.

Table 4.16: Feedback Thinking

Reference to Figure 4.1: C3 vs. A1	Qualitative+ Quantitative	Baseline
Feedback Thinking	**75%**	**5%**
1. Two feedback loops	70%	5%
2. One feedback loop only	5%	0%
Linear Thinking	**25%**	**95%**
3. Cause and effect relationships only	20%	40%
4. List of variables only	5%	55%
5. Blank answer	0%	0%
Total	**100 %**	**100 %**
N=20		

Confidence: An improved performance and greater understanding of the system also led to a higher confidence in participants' responses. Three-fourths of the total participants felt that this task was now easier to solve than before. Earlier, after the pre-test, only 10% of the participants had found the task easy to solve. The Chi-square tests reveal that there was a significant improvement in confidence from the combined intervention (χ^2 = 17.29, df = 1, P=0.0001)

4.4.3.4 Comparison: Efficacy of quantitative system dynamics as compared to qualitative system dynamics

Reference to main results table (Table 4.7): D3 vs. B2

Test for hypothesis: H4.2

Performance: In the complex task, quantitative intervention was expected to result in better performance than performance after qualitative intervention. As shown in

Table 4.17, quantitative intervention resulted in significantly better performance when compared to qualitative intervention ($\chi^2 = 5.58$, df = 1, P=0.02).

Table 4.17: Correctness

Reference to Figure 4.1: D3 vs. B2	**Quantitative**	**Qualitative**
Correct (Both segments correct)	**50%**	**15%**
Incorrect	**50%**	**85%**
1. Only one segment correct	5%	25%
2. No segment correct	40%	60%
3. Blank answer	5%	0%
N=20		

Forecasting: As seen in Table 4.18, the improvement in performance however led to significantly better forecast accuracy in the Quant-test (15.8 in Quant-test and 67.8 in Qual-test).

Table 4.18: Forecast accuracy

Reference to Figure 4.1: D3 vs. B2	**Quantitative**	**Qualitative**
Accuracy (MAPE)	**11.3%**	**66.0%**
1. T1	15.4%	37.8%
2. T2	14.1%	80.2%
3. T3	4.5%	79.9%
N=20		

Feedback Thinking: As seen in Table 4.19, there was no significant increase in the amount of feedback thinking after the two interventions ($\chi^2 = 0.44$, df = 1, P=0.51).

Table 4.19: Feedback thinking

Reference to Figure 4.1: D3 vs. B2	Quantitative	Qualitative
Feedback Thinking	**60%**	**70%**
1. Two feedback loops	50%	40%
2. One feedback loop only	10%	30%
Linear Thinking	**40%**	**30%**
3. Cause and effect relationships only	30%	20%
4. List of variables only	10%	10%
5. Blank answer	0%	0%
Total	**100 %**	**100 %**
N=20		

Confidence: There were no significant differences between the confidence reported in Qual and Quant-tests ($\chi^2 = 2.67$, df = 1, P=0.10).

4.4.3.5 Comparison: efficacy of combined qualitative and quantitative system dynamics as compared to qualitative system dynamics only and quantitative system dynamics only

Reference to main results table (Table 4.7): C3 vs. B2 and D3

Test for hypothesis: H4.2

Performance: For a complex task, a combination system dynamics intervention led to 65% correct responses as compared to only 15% obtained after the qualitative intervention, and 50% obtained after the quantitative intervention. Detailed results are shown in Table 4.20. The difference between the performance of the combined intervention and the qualitative intervention is significant (χ^2 = 10.42, df = 1, P=0.001) whereas it is not for the quantitative intervention (χ^2 = 0.92, df = 1, P=0.34). Hypothesis H4.2 is thus accepted. Table 4.21 depicts results of forecast accuracy in the three tests.

Table 4.20: Correctness

Reference to Figure 4.1: D3 vs. B2 vs. D3	Qualitative+ Quantitative	Quantitative	Qualitative
Correct (Both segments correct)	**65%**	**50%**	**15%**
Incorrect	**35%**	**50%**	**85%**
1. Only one segment correct	5%	5%	25%
2. No segment correct	30%	40%	60%
3. Blank answer	0%	5%	0%
N=20			

Table 4.21: Forecast Accuracy

Reference to Figure 4.1: D3 vs. B2 vs. D3	Qualitative+ Quantitative	Quantitative	Qualitative
Accuracy (MAPE)	**8.7%**	**11.3%**	**66.0%**
1. T1	11.8%	15.4%	37.8%
2. T2	10.3%	14.1%	80.2%
3. T3	4.2%	4.5%	79.9%
N=20			

Feedback Thinking: As shown in the table (Table 4.22), there were no significant differences in the level of feedback thinking reported in the three tests between the three interventions (χ^2 = 0.13, df = 1, P=0.72 for Qualitative + Quantitative vs. Qualitative intervention and χ^2 = 1.03, df = 1, P=0.31 for Qualitative + Quantitative vs. Quantitative).

Table 4.22: Feedback Thinking

Reference to Figure 4.1: D3 vs. B2	Qualitative+ Quantitative	Quantitative	Qualitative
Feedback Thinking	**75%**	**60%**	**70%**
1. Two feedback loops	70%	50%	40%
2. One feedback loop only	5%	10%	30%
Linear Thinking	**25%**	**40%**	**30%**
3. Cause and effect relationships only	20%	30%	20%
4. List of variables only	5%	10%	10%
5. Blank answer	0%	0%	0%
Total	**100 %**	**100 %**	**100 %**
N=20			

Confidence: The difference between the confidence reported from the combined intervention and the qualitative only is significant (χ^2 = 10.00, df = 1, P=0.0002) whereas it is not for quantitative only (χ^2 = 2.67, df = 1, P=0.10).

A summary of results for the four hypotheses that were tested are shown in Table 4.23.

Table 4.23: Results of hypotheses

Question	Expectation	Result	Hypotheses
Efficacy of **qualitative system dynamics** as compared to **baseline** performance	Simple: QualSD* > Baseline	Simple: QualSD > Baseline	H1.1: Accepted
	Complex: QualSD = Baseline	Complex: QualSD = Baseline	H1.2: Accepted
Efficacy of **quantitative system dynamics** as compared to **baseline** performance	Simple: QuantSD > Baseline	Simple: QuantSD > Baseline	H2.1: Accepted
	Complex: QuantSD > Baseline	Complex: QuantSD > Baseline	H2.2: Accepted
Efficacy of **combined qualitative and quantitative system dynamics** compared to **baseline** performance	Simple: (Qual, Quant)SD > Baseline	Simple: (Qual, Quant)SD > Baseline	H3.1: Accepted
	Complex: (Qual, Quant)SD > Baseline	Complex: (Qual, Quant)SD > Baseline	H3.2: Accepted
Relative efficacy of **combined qualitative and quantitative system dynamics, quantitative system dynamics** and **qualitative system dynamics**	Simple: (Qual, Quant)SD = QuantSD = QualSD	Simple: (Qual, Quant)SD = QuantSD = QualSD	H4.1: Accepted
	Complex: (Qual, Quant)SD > QuantSD > QualSD	Complex: (Qual, Quant)SD = QuantSD > QualSD	H4.2: Rejected

* SD – System Dynamics

4.4.4 Learning effect

As previously stated, a limitation of the first experiment was that the same cohort was used for measuring changes in baseline, qualitative intervention and combined intervention. Further, the use of the same cover story and tasks in all the three tests might have lead to familiarity with the tasks thereby resulting in improved performance.

Table 4.24: Comparing results to examine the effect of pre-tests

Comparison	Subcategory	Chi-square results
L1. Effect of taking pre-test on performance in Qual-test (A2 vs. B2)	***Simple***	
	Performance	(χ^2 = 0.13, df = 1, P<1), Not significant
	Understanding	(χ^2 = 0.13, df = 1, P<1), Not significant
	Confidence	(χ^2 = 0.36, df = 1, P<1), Not significant
	Complex Task	
	Performance	(χ^2 = 1.13, df = 1, P<1), Not significant
	Understanding	(χ^2 = 4.33, df = 1, P<0.05), Significant
	Confidence	(χ^2 = 1.11, df = 1, P<1), Not significant
L2. Effect of taking pre-test on performance in Quant-test (A3 vs. B3)	***Simple***	
	Performance	(χ^2 = 1.11, df = 1, P<1), Not significant
	Understanding	(χ^2 = 0.07, df = 1, P<1), Not significant
	Confidence	(χ^2 = 1.11, df = 1, P<1), Not significant
	Complex Task	
	Performance	(χ^2 = 4.33, df = 1, P<0.05), Significant
	Understanding	(χ^2 = 0.06, df = 1, P<1), Not significant
	Confidence	(χ^2 = 0, df = 1, P<1), Not significant
L3. Effect of taking pre-test and Qual-test on performance in Quant-test (A3 vs. C3)	***Simple***	
	Performance	(χ^2 = 0.23, df = 1, P<1), Not significant
	Understanding	(χ^2 = 0. 31, df = 1, P<1), Not significant
	Confidence	(χ^2 = 0, df = 1, P<1), Not significant
	Complex Task	
	Performance	(χ^2 = 0.11, df = 1, P<1), Not significant
	Understanding	(χ^2 = 0.63, df = 1, P<1), Not significant
	Confidence	(χ^2 = 0, df = 1, P<1), Not significant
L4. Effect of taking Qual-test on performance in Quant-test (B3 vs. C3)	***Simple***	
	Performance	(χ^2 = 0.36, df = 1, P<1), Not significant
	Understanding	(χ^2 = 0.68, df = 1, P<1), Not significant
	Confidence	(χ^2 = 1.03, df = 1, P<1), Not significant
	Complex Task	
	Performance	(χ^2 = 0, df = 1, P<1), Not significant
	Understanding	(χ^2 = 0, df = 1, P<1), Not significant
	Confidence	(χ^2 = 0, df = 1, P<1), Not significant

The section on experimental design (Design) shows how the current experiment was designed deliberately to remove this learning effect. To confirm this, three comparisons were made. In each of these three comparisons, one of the cohorts underwent an additional test (pre-test) before the test for which the results were being compared (Table 4.24). For instance, both cohorts A and B underwent qualitative intervention and their performance was measured using the Qual-test (Figure 4.1). Cohort A underwent a pre-test however cohort B did not. A comparison of Qual-test results of cohort A and B would reveal if cohort A was advantaged or disadvantaged by taking the pre-test. Chi-square tests were conducted to check if the results had any statistically significant differences. In each of the three comparisons, Chi-square tests are conducted on both simple and complex tasks. Detailed results are shown in the table (Table 4.24).

As can be seen from Table 4.24, for participants who took the Qual-test after undergoing the qualitative intervention, there is no significant difference in the performance, understanding and confidence for a simple task. Similarly, in the case of the complex task, except for their understanding of feedback loops, participants' performance and understanding show no significant difference. This implies that there was very little difference between participants who underwent the pre-test before the Qual-test as compared to the participants who did not.

The next step was to see whether there is any learning effect for participants who undertook the pre-test and the Qual-test prior to the Quant-test as compared to ones who did not undertake the pre-test but took the Qual-test. From Table 4.24 it can be seen that, similar to our inference in the case of the Qual-test, there is no significant effect of the pre-test on the performance of participants that undergo it prior to the Quant-test. In all cases, the Chi-square tests show insignificant differences, except in the case where one group undergoes only the Qual-test and another undergoes the pre-test and the Qual-test, in which case performance in the complex task is significantly different. Again, this implies that there was very little difference between participants who underwent both the pre-test and Qual-test as compared to participants who did not undergo the pre-test.

The third comparison was between the group that undertakes Quant-test without undertaking any test prior to that and the group that undertakes Quant-test but has undertaken pre-test and Qual-test previously. In this case, none of the comparisons showed any significant difference, hence dismissing the possibility of any learning that may have been induced by taking earlier tests.

The fourth comparison was between the two groups that did not undergo any pre-test but between group (B) that took the Qual-test and Quant-test and group (C) that took only the Quant-test. In this case as well, since there are no statistically significant differences, it can be assumed that the Qual-test did not affect the learning behaviour of participants.

The results presented above confirm that this experiment was successful in eliminating any learning effects, thereby removing the possibility of any bias. The results obtained in this experiment also reinforce those obtained from the first experiment.

4.4.5 Comparison between participants enrolled in PhD degree versus those enrolled in masters degree

As shown in the table (Table 4.25) below, there was no significant difference in performance between participants enrolled in PhD and Masters. A possible explanation of this could be that all participants had the basic knowledge to understand the system dynamics training and that any differences could have been due to technical competencies. As mentioned previously, all participants had at least a year of full-time work experience and would not have had difficulty in understanding the tasks.

Table 4.25: Difference between Performances of Participants Enrolled in PhDs and Masters

Comparison Group	Test	Chi-square results
1. Group A	Pre-test	***Simple:*** (χ^2 = 1.32, df = 1, P<1), Not significant
		Complex: None could answer, Not significant
2. Group A	Qual-test	***Simple:*** (χ^2 = 1.94, df = 1, P<0.2), Not significant
		Complex: (χ^2 = 0.59, df = 1, P<1), Not significant
3. Group B	Qual-test	***Simple:*** (χ^2 = 4.44, df = 1, P<0.05), Significant
		Complex: (χ^2 = 3.33, df = 1, P<0.1), Significant
4. Group A	Quant-test	***Simple:*** (χ^2 = 2.35, df = 1, P<0.2), Not significant
		Complex: (χ^2 = 0.38, df = 1, P<1), Not significant
5. Group B	Quant-test	***Simple:*** (χ^2 = 0.70, df = 1, P<1), Not significant
		Complex: (χ^2 = 2.97, df = 1, P<0.1), Not significant
6. Group C	Quant-test	***Simple:*** (χ^2 = 0.09, df = 1, P<1), Not significant
		Complex: (χ^2 = 0.59, df = 1, P<1), Not significant
7. Group D	Quant-test	***Simple:*** (χ^2 = 0.05, df = 1, P<1), Not significant
		Complex: (χ^2 = 0.83, df = 1, P<1), Not significant

4.5 Discussion

Results from the study provide us with useful insights regarding the role of qualitative and quantitative tools in the system dynamics methodology. To provide for a rigorous evaluation of task complexity, the tests were administered for a simple as well as a complex task. The current study has contributed to the literature and the practice of system dynamics by evaluating the individual and relative performances of qualitative and quantitative tools, thus presenting answers to a long standing debate regarding the relative effectiveness of these tools (Richardson 1976/1986; Homer and Oliva 2001, Coyle 2000, Wolstenholme 1999, Forrester 1991). Results of this study confirm the hypotheses that qualitative tools alone are not only effective but are also sufficient to solve simple problems; but complex problems require the use of quantitative tools. Specifically, the results suggest that qualitative tools contribute towards the understanding of simple stock/flow systems. However, they fail to influence performance in complex systems. On the other hand, quantitative tools, being equally useful for simple tasks, are also effective in modelling complex tasks. The forthcoming sections are divided according to task complexity.

4.5.1 Simple task

One aspect of this study was to measure the usefulness of system dynamics tools: qualitative, quantitative and a combination of qualitative and quantitative, in comparison with a scenario where no system dynamics training is provided. Without any system dynamics training, participants showed poor performance for both simple and complex tasks. In fact, performance for a simple task was so poor that worse performance in case of the complex task was not surprising. Participants' ability without system dynamics training to solve a simple stock/flow task points towards their incompetence to tackle dynamic situations, which is surprising but confirms the results of Sweeny and Sterman (2000). These findings are not only surprising for us, but even for the 85% participants who were confident of solving this problem. Out of the 35% correct responses, as expected, the majority use the regular method of calculating the largest value of stock. Out of those who answered incorrectly, 20% do use the correct strategy but make a computation error, which suggests that a surprisingly high number of people make calculation errors while solving simple calculations.

Participants who underwent the qualitative intervention showed clear improvement of performance as compared to those without any system dynamics training (40%

increase in performance). This can be explained considering that participants benefited by learning qualitative system dynamics concepts, specifically about stocks/flows. Given that stocks and flows are fundamental to learning about dynamic systems, these results imply that qualitative system dynamics is certainly a useful tool for learning these, conforming to the results of Kainz and Ossimitz (2002). Even though participants showed substantial improvement in performance, results show that they still used the computational approach to arrive at the answer, instead of using the visual strategy that was taught during the qualitative intervention. This may have resulted from participants being more confident about the computational method, which provided a "tried-and-tested" conventional approach to solving the problem. In contrast, the previous study (Chapter 3) which comprised a similar design had a higher number of participants use the visual strategy. This discrepancy shows that although the qualitative intervention had a positive impact on performance, participants not only used the visual strategy but also improved their basic computational skills.

These results corroborate the findings of a great deal of the previous work in this field. For instance, the results reported in Kainz and Ossimitz (2002) and that of the previous Chapter bear resemblance to those obtained in this experiment. As seen in Figure 4.6 below, the results from Kainz and Ossimitz (2002) in both the tests bear some resemblance and are not statistically significant prior to or after the qualitative training ($\chi^2 = 2.75$, df = 1, P<0.1 for pre-test and $\chi^2 = 0.2$, df = 1, P<1 for Qual-test). Further, there are not any significant differences between the results reported in this Chapter and those reported in Chapter 3 ($\chi^2 = 1.10$, df = 1, P=0.30 for pre-test and $\chi^2 = 0.18$, df = 1, P=0.68 for Qual-test).

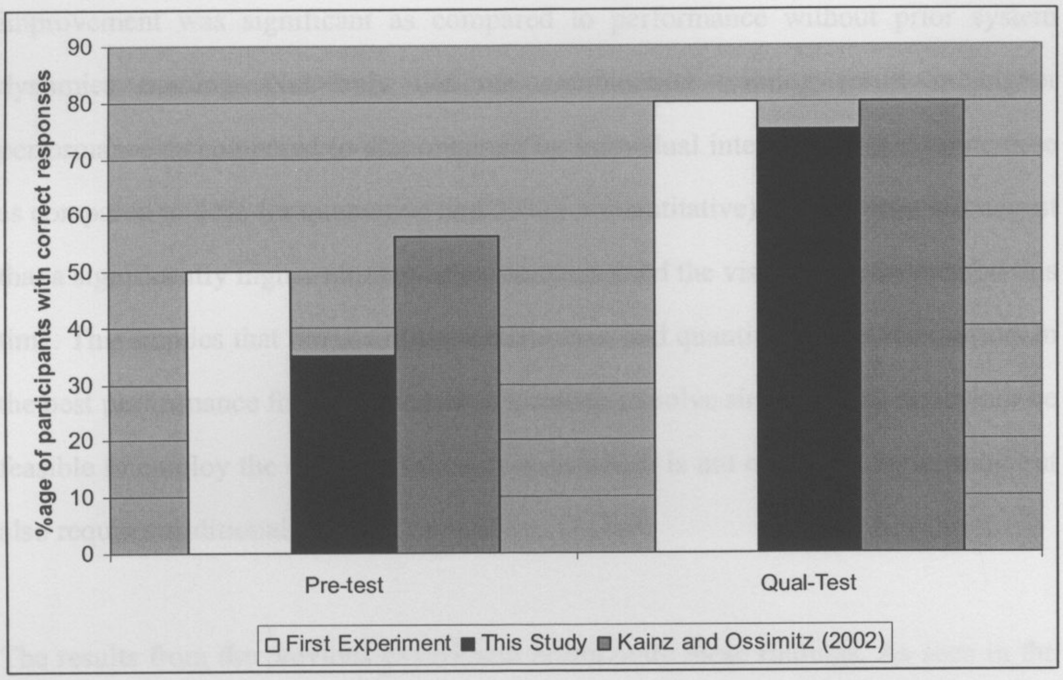


Figure 4.6: Pre-test and Qual-test comparisons between the current study, first experiment and Kainz and Ossimitz (2002)

In the case of participants' that underwent a quantitative intervention, results show significant improvement in performance as compared to participants without any system dynamics training (35% increase). In fact, the simple task was easily modelled using Powersim and the Powersim output could have led to identification of the point where inflow exceeds outflow—the answer to the question. But interestingly, the availability of Powersim did not lead to a significant increase of the use of the visual strategy. A possible explanation is that since the task was simple enough to be calculated using manual computation, participants' may not have bothered to make an extra effort to use Powersim. An analysis of incorrect errors suggests that the quantitative training led to a decrease in participants who made logical errors in solving the task.

Given that the qualitative and quantitative interventions individually resulted in improved performance, as expected, the use of qualitative and quantitative interventions together also resulted in similar performance improvement. This improvement was significant as compared to performance without prior system dynamics training. Not only did the combination training result in higher performance as compared to that obtained by individual interventions (55% increase as compared to 40% for qualitative and 35% for quantitative), the results also suggest that a significantly higher number of participants used the visual analysis method this time. This implies that the use of both qualitative and quantitative methods results in the best performance for a simple task. However, to solve simple tasks, it may not be feasible to employ the combination approach, which is not only time consuming but also requires additional skills.

The results from the previous experiment corroborate these findings. As seen in the figure below (Figure 4.7), there are no significant differences between the results reported in this study and the one reported in Chapter 3. It may be recalled that the cover story used in Chapter 3 was similar in complexity to those used in the current Chapter. All these case studies were two-stock problems with two feedback loops.

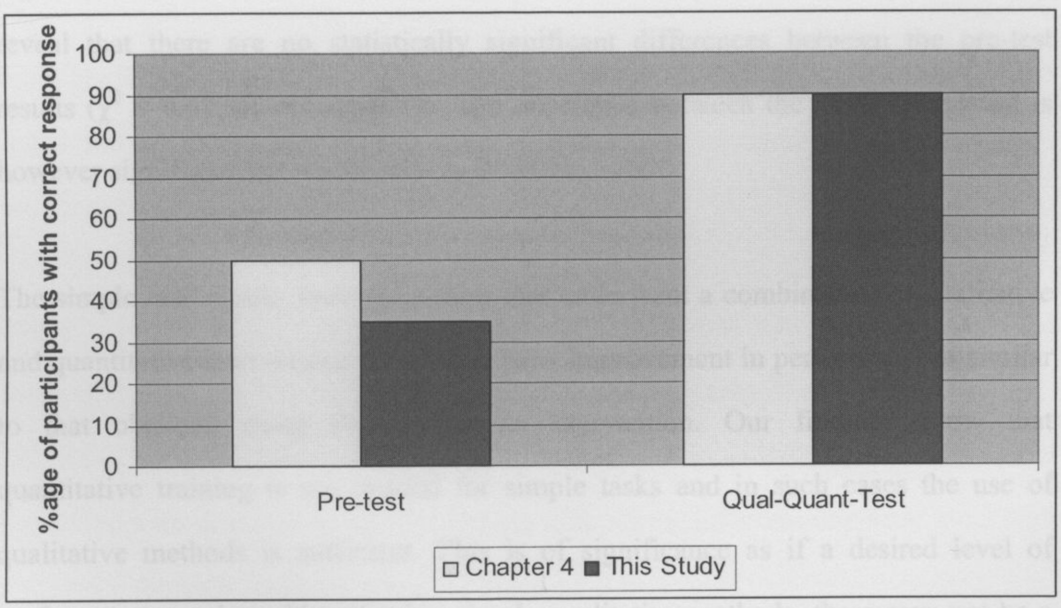


Figure 4.7: Comparison with results reported in Chapter 3

In a similar experiment, Pala and Vennix (2006) also tested performance using a combination intervention. A comparison with their results shows a similar pattern (Figure 4.8 below).

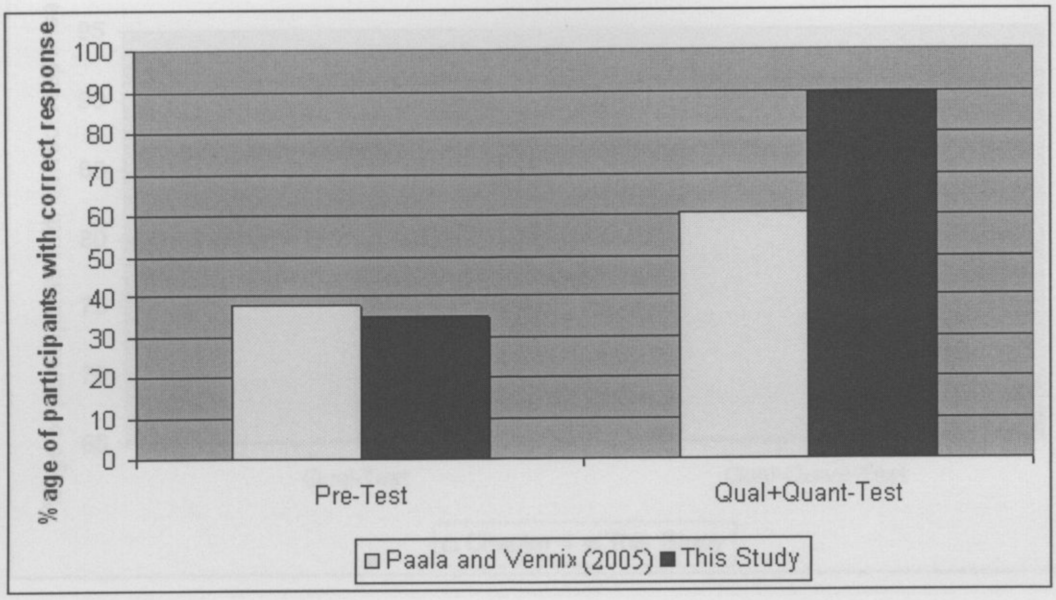


Figure 4.8: Comparison with Pala and Vennix (2005)

Chi-square test between the results of the current study and those of Pala and Vennix reveal that there are no statistically significant differences between the pre-test results (χ^2 = 0.07, df = 1, P=0.79). The difference between the Qual+Quant-test is however significant (χ^2 = 6.85, df = 1, P=0.01).

The simple task results from the cohort that underwent a combination of qualitative and quantitative intervention shows that their improvement in performance is similar to that obtained using the qualitative intervention. Our findings show that quantitative training is not needed for simple tasks and in such cases the use of qualitative methods is sufficient. This is of significance as if a desired level of performance can be achieved using simple qualitative methods, there may not be a need to use system dynamics modelling software, to save time and resources.

There are similarities between the results obtained in the previous experiment and those reported here. For instance, in both studies a combined intervention leads to the highest performance when compared to a qualitative only intervention (Figure 4.9).

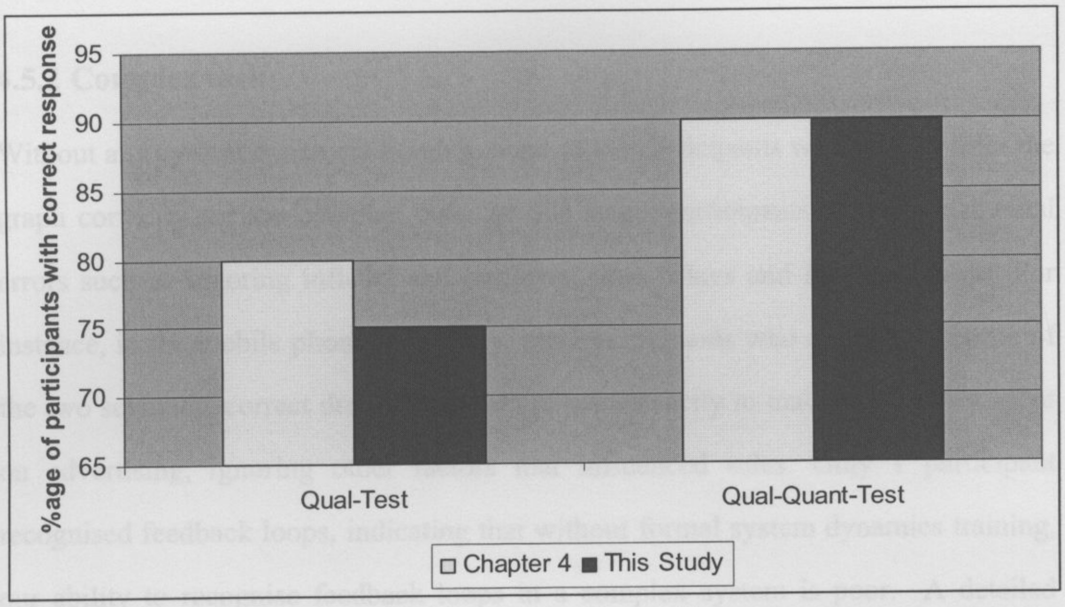


Figure 4.9: Qualitative + Quantitative and Qual comparison with results reported in Chapter 3

The combined intervention, however, led to a significant increase in the number of participants that used the visual strategy when compared to the qualitative interventions ($\chi^2 = 8.29$, df = 1, P=0.004) and also when compared to the quantitative intervention ($\chi^2 = 5.23$, df = 1, P=0.02).

There was no significant difference in the confidence between the combined intervention and that recorded after the individual interventions ($\chi^2 = 0.36$, df = 1, P=0.55—combined compared with quantitative; $\chi^2 = 1.11$ df = 1, P=0.29—combined compared with quantitative). Performances after all the interventions were quite high and hence the high confidence was justified.

It can be suggested that a combined intervention resulted in the highest performance and also resulted in a significant shift towards the visual approach to solve simple stock/flow tasks.

4.5.2 Complex task

Without any system dynamics training, none of the participants were able to infer the graph correctly for the complex task. At this stage, participants made fundamental errors such as ignoring inflows and outflows, time delays and feedback loops. For instance, in the mobile phone case study, most participants who did not get either of the two segments correct drew the pattern of sales exactly to match the money spent on advertising, ignoring other factors that influenced sales. Only 1 participant recognised feedback loops, indicating that without formal system dynamics training, our ability to recognise feedback loops in a complex system is poor. A detailed analysis of responses shows that participants' mental models are dominated by "linear thinking", instead of "feedback thinking", which clearly shows their inherent inability to understand feedback systems. As expected, without system dynamics training, participants' reported low confidence in tackling the complex task. Results from the cohort that underwent qualitative intervention show that these participants did not show much improvement when compared to ones without system dynamics training. However, what improved significantly was the level of feedback thinking. This may be attributed to participants' learning of causal loop diagramming during the qualitative intervention. Even though participants could recognise feedback loops, they failed at inferring the behaviour over time of the key variable, which was essential to achieving the desired result. An implication of this finding is that qualitative tools are not completely useless in complex tasks; they do contribute to the understanding of feedback systems.

Results show that participants who underwent the quantitative intervention showed major improvement in performance as compared to ones that did not receive any training. A detailed analysis reveals that not only were participants able to recognise feedback loops, they were also now successful in inferring behaviour over time of the key variable. This can be directly attributed to Powersim training, which involves system dynamics modelling of complex systems. It may be recalled that participants who underwent the quantitative intervention did not undergo any causal loop diagramming, which is a key component of qualitative training implying its needlessness in solving such complex tasks.

As previously indicated, the use of qualitative and quantitative techniques alone results in significant improvements for feedback thinking and/or performance for a complex task. A combination of these also improves performance and, as results indicate, participants were now not only able to recognise key feedback loops but also able to simulate the system. Combination of qualitative and quantitative tools results in significant improvement (increase of 65%) as compared to only 15% improvement in the case of qualitative and 50% improvement in the case of quantitative intervention, when compared to those who did not undergo any system dynamics training. This clearly indicates although the quantitative tools are useful in solving complex tasks, a combination of qualitative and quantitative tools results in even better performance; though not significantly better than the use of quantitative tools alone. An implication of this result is that for a complex task, the use of quantitative tools alone may be sufficient to solve the purpose.

As shown in Figure 4.10, these findings corroborate the results obtained in the previous experiment.

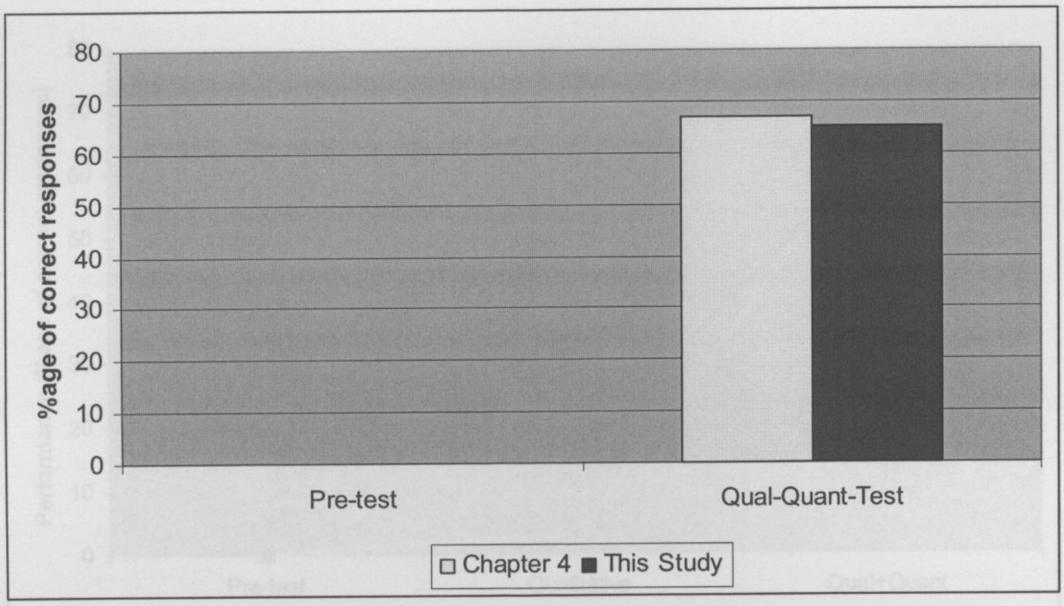


Figure 4.10: Qualitative + Quantitative-test results comparison with those obtained in previous experiment

The results reported in the present study are similar to the ones obtained in the first experiment (Figures 4.11 and 4.12). As discussed before, the complexity of the task used in the previous experiment and those used in the current experiment is similar. Therefore results from these studies can be compared with each other. Both studies tested the relative usefulness of qualitative intervention when compared with a combination of qualitative and quantitative intervention. In addition to the analyses mentioned above, in the current study, comparison with results obtained after quantitative intervention alone was also tested.

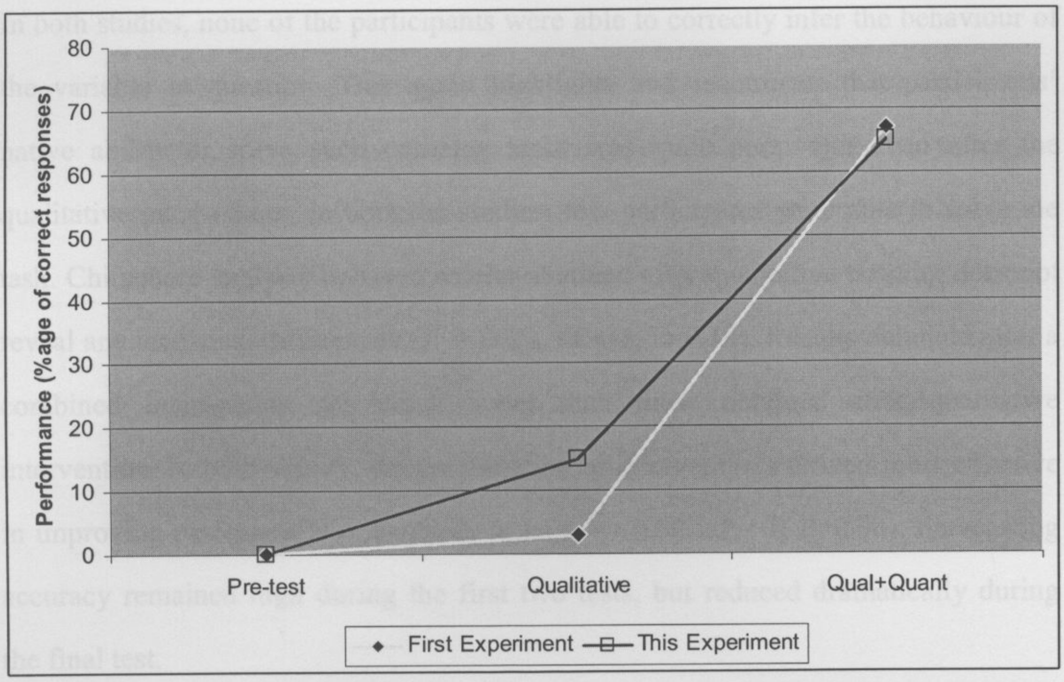


Figure 4.11: Comparison with experiment 1 (performance in complex task)

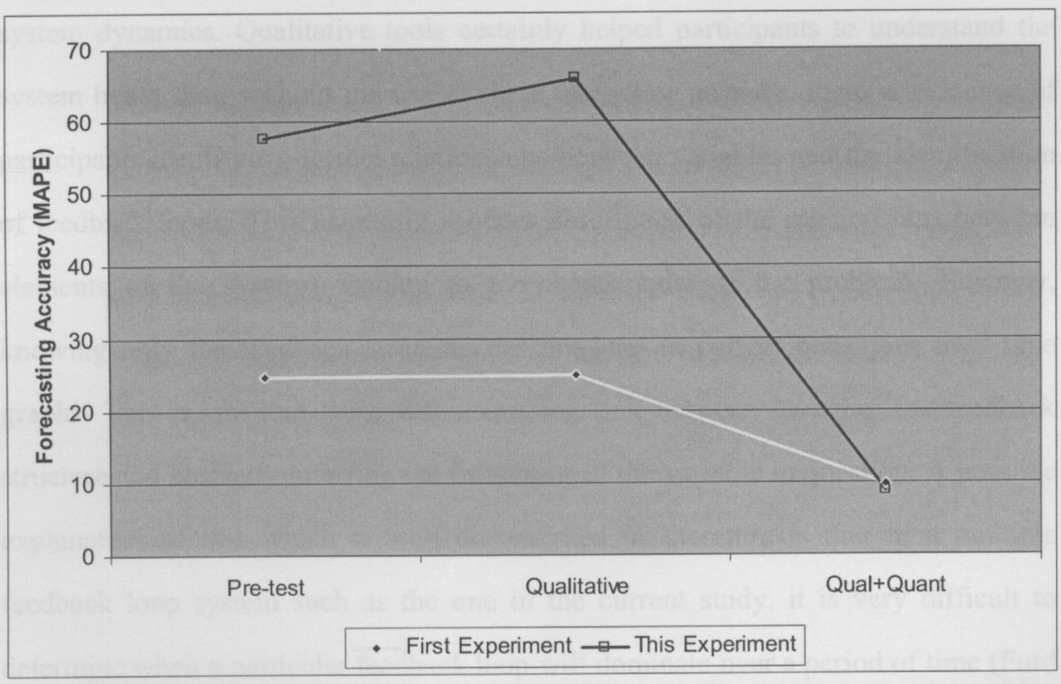


Figure 4.12: Comparison with experiment 1 (forecast accuracy in complex task)

In both studies, none of the participants were able to correctly infer the behaviour of the variable in question. This again highlights and reconfirms that participants' native ability to solve such complex tasks was quite poor. But even after the qualitative intervention, in both the studies, few participants were able to solve the task. Chi square analysis between results obtained after qualitative training does not reveal any statistical differences ($\chi^2 = 2022$, df = 1, P=0.14). Results obtained after a combined intervention are much better than those obtained after qualitative intervention. In both studies, the combination of interventions proved most effective in improving performance in complex tasks ($\chi^2 = 0.02$, df = 1, P=0.90). Forecasting accuracy remained high during the first two tests, but reduced dramatically during the final test.

These findings have important implications for the use of qualitative and quantitative system dynamics. Qualitative tools certainly helped participants to understand the system better than without the use of these tools. For instance, there is evidence of participants identifying correct relationships between variables and the identification of feedback loops. This helped in a better description of the relationships between elements of the system—leading to a systems view of the problem. However, knowing only the feedback structure did not lead to correct behaviour over time graphs. This meant that there was a missing link between knowing the feedback structure and correctly inferring the behaviour of the variable in question. A possible explanation of this which is well documented in literature is that in a multiple feedback loop system such as the one in the current study, it is very difficult to determine when a particular feedback loop will dominate over a period of time (Ford 1999). In general, it seems that though qualitative tools add to the understanding of a system, it is imperative to use quantitative tools when inferring behaviour over time in relatively complex tasks.

4.6 Conclusion

Previous work on the usefulness of system dynamics methodology suggests that our inherent understanding of situations involving components of dynamic complexity is quite poor and that qualitative or a combination of qualitative and quantitative system dynamics tools assist in overcoming this deficiency. However, none of these studies explicitly compared the relative efficacy of qualitative, quantitative and combined system dynamics tools for simple and complex tasks. The present study was designed to determine the influence of qualitative, quantitative and a combination of qualitative and quantitative system dynamics tools on performance, understanding and confidence of tasks that involve components of dynamic complexity. It was hypothesised that all three interventions would lead to an improved performance when compared to performance without these interventions. Further, it was suggested that the three interventions would not lead to significant differences in performance in the simple task. However, for the relatively complex task, participants who underwent the combined intervention and the quantitative intervention were hypothesised to perform better than those who underwent qualitative intervention only.

The most obvious finding to emerge from this study is that our ability to solve simple and complex tasks involving dynamic complexity is naturally quite poor. For a simple stock/flow task, this deficiency is significantly overcome by the three system dynamics interventions. Further, both the combined intervention and the quantitative intervention alone lead to significant improvement in performance in the complex task. Qualitative intervention alone does not significantly contribute towards correctly inferring behaviour in complex tasks. One of the more significant findings of this study emerges from a comparison of the three interventions amongst each other. This analysis revealed that the performances for simple tasks after all three types of intervention were comparable. However, a significantly higher number of participants who underwent the combined intervention used the visual strategy to arrive at the result. For the complex task, however, participants who underwent the combined intervention and those who underwent the quantitative intervention showed significantly higher performance, forecasting capability and confidence as compared to those who underwent the qualitative intervention. However, no significant differences emerged when participants' understanding of feedback loops was compared. Results from confidence test suggest that participant's feelings of confidence may have been somewhat independent of their performance in the task. A possible explanation of this is that such stock-flow situations occur quite frequently in daily life might have encouraged participants to feel confident of their responses.

In general, therefore, it seems that the three interventions do contribute to our understanding of tasks that involve components of dynamic complexity. More specifically, for simple stock/flow tasks, all interventions are equally useful for task performance. The results of this study support the idea that computer modelling is essential for relatively complex tasks such as inferring the behaviour over time of a variable. Taken together, these findings suggest that qualitative tools are sufficient for simple tasks but contribute only to a limited extent in complex tasks. Further, quantitative tools are mandatory to tackle more complex tasks.

Finally, a number of caveats need to be noted regarding the present study. Firstly, performance of only one complex system with two stocks was tested. In practice, however, a system might involve many more stocks, flows and time delays. This may substantially increase the complexity of the task. Secondly, the results reported here are dependent on the definition of performance, understanding and confidence as defined by the author. Thirdly, the measurement of performance in the complex task does not measure the precise accuracy of the graphs. For our purposes, the correctness of the graph simply depends upon the overall correctness of the pattern when compared with the simulated result. Finally, the current study only examined the influence of system dynamics tools soon after the interventions were administered. These results therefore may be influenced by 'recency bias' where people tend to focus primarily on what has happened recently. One suggested reason for the recency effect is that these items are still present in working memory when recall is solicited. Hence this study has not tested whether the interventions made long-term changes to the way participants think and deal with such tasks. In other words, did this intervention make *fundamental* changes to our mental models or not? This question forms the rationale for the next experiment in which a subset of these participants is tested on similar tasks after some time has elapsed.

Chapter 5
Experiment 3

5.1 Introduction

5.2 Background

5.3 Methodology

5.4 Results

5.5 Discussion

5.6 Conclusion

5.1 Introduction

"With few exceptions the relationship between the use of systems thinking and organizational performance remains the province of anecdote rather than rigorous follow up research." (Cavaleri and Sterman, 1997)

Even though many studies claim to have made changes to mental models, few have gone a step further to conduct longitudinal experiments to test the efficacy of the system dynamics training after some time has elapsed (Cavaleri and Sterman 1997). In the previous study (Chapter 4) as well, participants' performance for simple and complex tasks was measured immediately after the system dynamics intervention. The current study evaluates whether that intervention made "fundamental" changes to participants' mental models. Specifically, this chapter explores which components of the methodology were retained and which ones were forgotten five months after initial training. The current study also sought to examine the contribution of system dynamics software to performance and understanding. This information will help practitioners and trainers to help identify aspects of the system dynamics methodology that are easily forgotten and hence may need to be reinforced.

A rigorous longitudinal experiment was conducted to test if the original system dynamics training made fundamental changes to the way participants tackled simple and complex tasks. The following section discusses the need of such a study and analyses two studies conducted previously that address similar issues. The Methodology section details the experimental design and the procedure followed to conduct this study. Results are then presented and discussed and appropriate comparisons are made with the previous experiment. The contributions of the study and its limitations are discussed in the final section.

5.2 Background

The need for rigorous evaluative research is widely accepted. For example, Strake (1967) writes of educational innovations, "*...folklore is not a sufficient repository. In our data banks we should document the causes and effects, the congruence of intent and accomplishment, and the panorama of judgments of those concerned*" (p539). Moreover, it is well known that studies that involve the observation at a single point in time (cross-sectional studies) are deficient in accounting for a high percentage of variance (Heller et al. 1977). A longitudinal study may consider any possibility of events that might have occurred in the elapsed time period (Heller et al. 1977). Hence, conducting a longitudinal experiment becomes imperative in situations where the long-term effectiveness of an intervention needs to be tested. In the context of system dynamics the need for a rigorous longitudinal study arises from the fact that most evidence that exists to date in this area is anecdotal and lacks systematic evaluation (Huz et al. 1997). Further, Cavaleri and Sterman (1997) argue that rigorous longitudinal research is essential to build a strong foundation for the enhancement and optimal use of system dynamics methodologies. Therefore, assurance is needed that exposure to system dynamics training produces lasting changes in the way people think and analyse complex problems.

Generally, conducting longitudinal research has been problematic due to the cost associated and the relative paucity of adequate methodology to carry out such studies (Heller et al. 1977). The lack of commitment by clients to carry out longitudinal research in real applications has been an additional challenge for system dynamists (Huz et al. 1997).

In general, memories of business education are short lived. Anderson (2000) argues that decay, interference and absence of retrieval cues are the three prime factors that are responsible for forgetting what has been learnt in class. In this study, we assume that a time gap of five months would have resulted in decay, participants' involvement in other activities might have constituted interference and their abstinence from using system dynamics constituted the absence of retrieval cues. A logical expectation then would be that concepts that were taught five months ago would now have been forgotten.

Contrary to this belief, longitudinal studies in system dynamics have demonstrated that people do retain system dynamics skills. Two such studies are described in detail below. However, an important aspect should be noted. Participants in these studies were using system dynamics skills during the retention interval. Also, the first study was not conducted as a controlled experiment.

In a first-of-its-kind study, Cavaleri and Sterman (1997) evaluated the change in thinking, behaviour and performance of two groups of people (managers and non-managers) who were subjected to different levels of systems thinking intervention. Non-managers merely played a board game that helped in understanding inefficiencies of supply chains—the Beer game (Sterman 1992b). Managers also played the Beer game, but in addition to that, played another simulation game that was specific to the work they were involved in and also attended a seminar on systems thinking. The results of this retrospective study suggest that there were significant differences in the amount of systemic thinking in the behaviour of managers, but not in the case of non-managers. Both managers and non-managers rated the systems thinking training as 'moderately valuable'. There was no evidence that the system dynamics made any difference in performance during the six years following the initial training. Organizational performance was measured by using four standard measures that were specific to the insurance industry. The data collected did not suggest any pattern of improvement in any of these four measures.

Given the difficulty in conducting longitudinal experiments in industry, this study is a pioneering research which sets the stage for further research in this area. However, there are some serious limitations. Firstly, this study wasn't performed in a controlled setting and hence it is difficult to ascertain if the difference in systemic thinking and behaviour was due to system dynamics training alone or attributed to other factors such as intrinsic differences between the participants (managers and non-managers). Secondly, the majority of the data collected from participants was in the form of self-reporting questionnaires. As pointed out by the authors themselves, these are prone to 'demand effects'. In this situation managers would have reported positive impact of systems thinking training as they believed that this was what the researchers were looking for.

In another example, that focuses on the integration of mental health and vocational services, Huz et al. (1997) evaluated the impact of systems thinking interventions over a six-month period. They conducted a controlled experiment using a pre-test/post-test design in which they measured changes in participants' mental models to assess the impact of systems thinking interventions. Akin to Cavaleri and Sterman (1997), this study also measured participants' mental models and organizational performance. In addition to these two measures, the modelling team evaluated their own performance as well. These key measurements were operationalised using ten 'domains' to measure changes prior to and after the intervention. The change in participants' mental models was operationalized using five domains of measurement and analysis. These were: participants' perceptions of the intervention, shifts in participants' goal structure, shifts in participants' change strategies, alignment of participant mental models, and shifts in understanding of how the system functions. The first four were measured using pre-test/post-test surveys and the last was operationalized via follow-up interviews, formal meetings' minutes, and informal reflections by the modelling team. Participants' perceptions of the intervention were evaluated using the 'model building evaluation questionnaire'. Results suggested that participants perceived the intervention as productive. Specifically, building a formal model proved beneficial to their understanding of the system. To measure shifts in participants' goal structure, participants rated the level of importance they associated to 20 goal statements, before and after the intervention. The ratings were recorded on a five point Likert scale. Paired t-test analysis revealed that there was a significant shift in the importance of shared services and common intake between two departments. This was an important aspect as the fragmentation between the two departments was a concern for the efficient working of the client system. Shifts in participants' change strategies were measured in a similar way as that for the previous domain. Paired t-tests in this case show that there was significant shift in the importance of four change strategies. Alignment of participants' mental models was measured using two questionnaires. Results show that participants were more aligned in their perceptions of system goals. However, there were no significant changes in their alignment in their perceptions on strategies for change. Using data from archives, informal reflection and observation in meetings, the authors concluded that the formal model also facilitated participant understanding of system structure and behaviour. Organisational performance was defined as shifts in the network of agencies that support services' integration, changes in policies and changes in outcomes for clients. These were measured via follow-up interviews and analysis of project archives and administrative data. The modelling team's reflection of the process was analysed using minutes of sessions, archival analysis and informal reflections of the team members. Huz et al. (1997 p165) raise important research questions based on their follow-up, such as *'what sub-components of the intervention matter most?'* The current study seeks to answer this question to an extent.

5.3 Methodology

5.3.1 Introduction

The experiment was spread over a period of five months and was based on a pre-test/post-test design. Briefly, as described in the previous Chapter, participants underwent a pre-test, followed by system dynamics training and then were administered the first post-test. Then after a gap of five months, a subset of these participants underwent two more post-tests. The tasks, detailed procedure and basis of the experimental design are described in the following sub-sections.

5.3.2 Participants

It may be recalled that in the previous experiment (Chapter 4), sixty (groups A, B and C) out of the total eighty participants underwent the complete system dynamics training. The remaining twenty participants (Group D) underwent training on quantitative tools only. Group D participants were therefore not eligible to participate in the current study. It is also worth noting that Groups A, B and C had similar results after qualitative and quantitative training.

The test used in this study was the same as that used as the Pre-test of the previous experiment (Chapter 4). It may be recalled that in the previous experiment Group A participants underwent the same test used in the current study. Therefore, to avoid the possibility of a learning effect that might have occurred due to repeated exposure to the same cover story and tasks, Group A participants were not eligible to participate in this experiment. The remaining forty eligible participants[1] were invited to participate in this follow-up study. Thirty-one of these responded positively to the

[1] Eligible participants were those who had undergone complete system dynamics training, i.e. qualitative as well as quantitative phases, but were not part of Group A.

call. They were subsequently recruited for this experiment. Participants were paid AU$40 for their time.

5.3.3 Design and procedure

The experiment consisted of two main stages as depicted in Figure 5.1. In the first stage (reported in Chapter 4), participants first underwent a pre-test. This was followed by qualitative and quantitative system dynamics training. The intervention was immediately followed by a post-test (Post-test 1). This phase of the experiment is marked as 'original learning' in the figure. Then after a gap of five months, participants underwent two more tests (Post-test 2 and Post-test 3). This phase of the experiment is marked as 'retention testing' in Figure 5.1. Each test consisted of two tasks that are described in detail below.

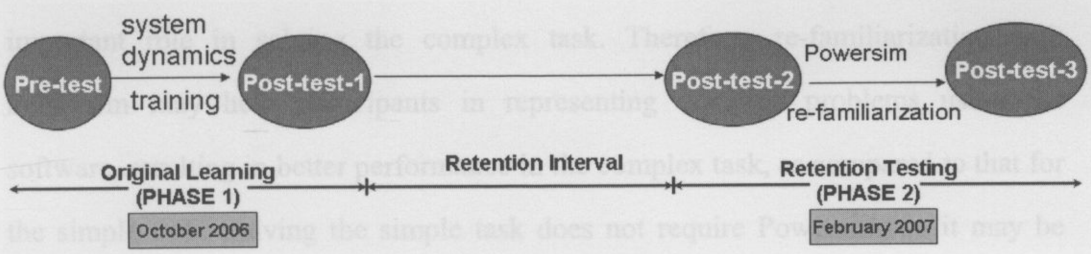


Figure 5.1: Design

5.3.4 Hypotheses

In the previous study it was found that system dynamics training substantially helped in improving performance for both simple and complex tasks. The aim was to evaluate participants' performance on comparable tasks. During the five month gap, the participants neither underwent any system dynamics training, nor were they involved in formal system dynamics modelling. It can thus be assumed that they would have forgotten some aspects of the training they were provided with five months ago. Hence it can be hypothesised that for both simple and complex tasks, participants' performance would have deteriorated after 5 months, i.e.

H1.1 SD (OL)[1] >[2] SD (R)[3] for simple task

H1.2 SD (OL) > SD (R) for complex task

Apart from forgetting the concepts, it is likely that participants would have also forgotten how to use the system dynamics software (Powersim). Hence, a decrease in performance shown by the retention test (SD(R)) could be attributed to either forgetting the basic system dynamics concepts or inability to successfully use the software. Therefore to measure the retention of system dynamics concepts and to eliminate the bias that may arise due to participants' lack of unfamiliarity with the software, participants were allowed time to practise with Powersim. From the results of the previous experiment (Chapter 4), it may be recalled that Powersim plays an important role in solving the complex task. Therefore, re-familiarization with Powersim may help participants in representing complex problems using the software, resulting in better performance in the complex task, as compared to that for the simple task. Solving the simple task does not require Powersim, so it may be hypothesised that performance in the simple task remains the same before and after Powersim re-familiarisation, i.e.

H2.1 SD (P-R)[4] = SD (R) for simple task

H2.2 SD (P-R) > SD (R) for complex task

---

[1] SD (OL): Original Learning: results obtained immediately after system dynamics training in October 2006.

[2] Significant difference in results.

[3] SD (R): Retention: results obtained after the end of retention interval in February 2007.

[4] SD (P-R): Retention with Powersim: results obtained after participants re-familiarize with Powersim.

5.3.5 Tasks and methods of analyses

The test used in this study was same as that used as the Pre-test of the previous experiment reported in the previous Chapter. The cover story of the task centered on sales of mobile phones in a small town. It may be recalled that the test consisted of two tasks—simple and complex. Task 1 was a simple question on stocks and flows in which participants were required to identify the time period at which the value of stock reached its maximum. Task 2 required participants to draw the behaviour of a key variable. The detailed description of the two tasks and the method of analyzing responses have been discussed in the previous chapters (Chapter 3 and Chapter 4). The inter-rater reliability (Kendall's Tau) for Post-test 2 was 0.95 and that for the Post-test 3 was 0.93.

5.3.6 Procedure

Original learning phase: October 2006 (as reported in chapter 4)

Refer to Phase 1 in Figure 5.1.

- *Pre-test*: Twenty participants underwent a pre-test. They completed the two tasks as described above. Participants were seated at a sufficient distance from each other and the tests were invigilated by two instructors.
- *System dynamics intervention*: Sixty participants, including the twenty participants who took the Pre-test underwent system dynamics training (categorized as Groups A, B and C in previous Chapter). The training was spread over two weeks and lasted for fifteen hours. It involved interactive sessions in which participants were taught concepts of qualitative and quantitative system dynamics.

- *Post-test 1*: Immediately after the intervention, all participants were administered the Post-test 1.

Retention interval: October 2006 to February 2007

- Participants did not undergo any additional training during this stage.

Retention Testing Phase: February 2007

Refer to Phase 2 in Figure 5.1.

- *Post-test 2*: Thirty-one out of the initial sixty participants who positively responded to the invitation to participate in this sequel study were administered Post-test 2.
- *Software practice*: Soon after the Post-test 2, participants spent an hour re-familiarising themselves with system dynamics software (Powersim) using online tutorials.
- *Post-test 3*: After the software familiarisation phase, participants underwent Post-test 3 under the same conditions as in the previous tests.
- *Self-reporting questionnaire*: Participants were given a questionnaire to fill after they completed Post-test 3. Through this questionnaire participants were required to report if they had been exposed to system dynamics during the retention interval. They also had to reflect on the impact of the initial system dynamics training on their thinking.

5.4 Results

5.4.1 Introduction

The results section is presented in two parts based on two questions that we seek to answer to test the above mentioned hypotheses: which aspects of system dynamics methodology are retained? Is retention affected by Powersim re-familiarisation?

The hypotheses offered above suggest that participants might have forgotten the majority of system dynamics concepts resulting in significantly lower performance in both simple and complex tasks. As Table 5.1 reveals, participants were able to maintain their improved performance for a simple task, however, their performance in a complex task was only as good as it was prior to any system dynamics training (Table 5.4). Re-familiarisation with Powersim does not result in significant improvement for either simple or complex tasks.

5.4.2 Simple Task

Table 5.1 presents an overview of the results from the two tests and compares these with the results obtained from the previous study (Chapter 4). Two key comparisons were made to test the hypotheses for the simple task. For each comparison, performance, strategy to solve the task and confidence are presented.

Table 5.1: Performance of participants for simple task

1. Pre-Test		**2. Post-Test 1**		**3. Post-Test 2**		**4. Post-Test 3**	
Performance	35.0%	Performance	90.0%	Performance	81.0%	Performance	87.0%
Visual Strategy	29.0%	Visual Strategy	89.0%	Visual Strategy	24.0%	Visual Strategy	26.0%
Confidence	85.0%	Confidence	97.0%	Confidence	87.0%	Confidence	87.0%
Ref: Baseline		**Ref: SD (OL)**		**Ref: SD (R)**		**Ref: SD (P-R)**	

Phase 1 **Phase 2**

Legend

Performance	Percentage of participants identifying maximum value of stock (simple)	**Confidence**	Percentage of participants that found the task 'easy'
Visual strategy	Percentage of participants using visual analysis		

5.4.2.1 Comparison: Which aspects of system dynamics methodology are retained

Reference to main results table (Table 5.1): Post-Test 1 vs. Post-Test 2

Test for hypothesis: H1.1

Performance: From the results of the previous chapter we know that system dynamics training in Phase 1 resulted in a significant improvement in performance for a simple task: 55% more participants were able to identify the time period at which the value of stock reaches its maximum. It was expected that a time gap of 5 months without any exposure to system dynamics would naturally dissipate participants' learning. However, interestingly, almost the same numbers of participants were still able to attempt the simple task successfully ($\chi 2 = 0.80$, df = 1, P<1). Eighty-one percent of the participants answered correctly and out of the remaining 19%, 6% used the right method of computing but faltered in the arithmetic and hence did not arrive at the correct answer. The 13% who got the wrong answer fell into the classic trap of identifying the time period at which the value of inflow-outflow is the largest.

Strategy: Even though the performance in Post-test 2 was at par with that of Post-test 1, there was no corresponding similarity in understanding (understanding fell to 24% from 89%, see Table 5.2). In Post-test 2, the level of sophistication fell and a substantially higher number of participants used the "hard way" to solve the problem, instead of visual analysis ($\chi 2 = 33.3$, df = 1, P<0.001). This may be attributed to the time gap participants were subjected to. The level of understanding was similar to the level participants had prior to system dynamics training. A detailed break-up of participants' strategy to solve the task is presented in the table below.

Table 5.2: Participants' strategy to solve the simple task

	SD (OL)	SD (R)
Correct Responses	**90 %**	**81%**
1. Use visual strategy	80.0 %	19.4 %
2. Use computational strategy	10.0 %	61.3 %
3. No justification	0 %	0 %
Incorrect Responses	**10.0%**	**19.0%**
4. Use visual strategy correctly but arrive at the wrong answer	3.3 %	0 %
5. Use computational strategy correctly but arrive at the wrong answer	5 .0%	6.5 %
6. Confound with largest inflow	0 %	3.2 %
7. Confound with largest outflow	0 %	0 %
8. Confound with largest (outflow + inflow)	0 %	9.7 %
9. Confound with largest (outflow – inflow)	0 %	0 %
10. Confound with largest (inflow - outflow)	1.7 %	0 %
11. Add all inflows and all outflows separately and compare	0 %	0 %
12. Blank answer	0 %	0 %
Total	**100 %**	**100 %**

Further analysis of correct responses reveal that there were statistically significant differences between responses in the original learning test and the retention test. Chi square tests show that a much higher number of participants used the visual strategy immediately after the intervention ($\chi 2$ = 28.69, df = 1, P<0.0001). As previously discussed, this strategy is a quicker way to arrive at the correct answer and was taught during the system dynamics interventions.

Confidence: Similar to Phase1, participants were quite confident of their performance ($\chi 2$ = 3.03, df = 1, P<0.10).

5.4.2.2 Comparison: Is retention affected by Powersim re-familiarisation

Reference to main results table (Table 5.1): Post-Test 2 vs. Post-Test 3

Test for Hypothesis: H2.1

Performance: Even though participants use the software to model complex stock-flow problems, the simple task, per se, does not require Powersim to arrive at the correct answer (Phase 1). Re-familiarisation with Powersim did not result in a substantial increase in participants giving the correct response compared to those who did not re-familiarise themselves with the software ($\chi 2 = 0.48$, df = 1, P<1).

Strategy: As seen in Table 5.3, the level of understanding, as expected, remained almost unchanged post Powersim tutorial ($\chi 2 = 0.03$, df = 1, P<1).

Table 5.3: Participants' strategy to solve the simple task

	SD (R)	**SD (PR)**
Correct Responses	**81.0%**	**87.0%**
1. Use visual strategy	19.4%	22.6%
2. Use computational strategy	61.3%	64.5%
3. No justification	0%	0%
Incorrect Responses	**19.0%**	**13.0%**
4. Use visual strategy correctly but arrive at the wrong answer	0%	0%
5. Use computational strategy correctly but arrive at the wrong answer	6.5%	0%
6. Confound with largest inflow	3.3%	0%
7. Confound with largest outflow	0%	0%
8. Confound with largest (outflow + inflow)	9.7%	9.7%
9. Confound with largest (outflow – inflow)	0%	0%
10. Confound with largest (inflow - outflow)	0%	0%
11. Add all inflows and all outflows separately and compare	0%	0%
12. Blank answer	0%	3.2%
Total	**100%**	**100%**

Confidence: As expected, participants remained equally confident before and after Powersim familiarisation ($\chi 2 = 0$, df = 1, $P<1$), thus indicating that in simple tasks Powersim did not play a role in boosting confidence.

5.4.3 Complex Task

Table 5.4 presents an overview of the results from the two tests and compares these with the results obtained from the previous study (Chapter 4). Two key comparisons were made to test the hypotheses for the complex task. For each comparison, performance, strategy to solve the task and confidence are presented.

Table 5.4: Performance of Participants for Complex Task

1. Pre-Test		**2. Post-Test 1**		**3. Post-Test 2**		**4. Post-Test 3**	
Performance	0%	Performance	68.0%	Performance	3.0%	Performance	16.0%
Feedback Thinking	5.0%	Feedback Thinking	78.0%	Feedback Thinking	10.0%	Feedback Thinking	26.0%
Confidence	10.0%	Confidence	75.0%	Confidence	6.0%	Confidence	19.0%
Forecasting Accuracy	58.0%	Forecasting Accuracy	10.6%	Forecasting Accuracy	47.0%	Forecasting Accuracy	39.0%
Ref: BASELINE		**Ref: SD (OL)**		**Ref: SD (R)**		**Ref: SD (P-R)**	

← **Phase 1** → ← **Phase 2** →

Legend

Performance	Percentage of participants inferring the correct behaviour (complex)	**Confidence**	Percentage of participants that found the task 'easy'
Understanding	Percentage of participants identifying feedback loops	**Forecasting Accuracy**	Mean Absolute Percentage Error (MAPE)

5.4.3.1 Comparison: Which aspects of system dynamics methodology are retained?

Reference to main results table (Table 5.4): Post-Test 1 vs. Post-Test2

Test for Hypothesis: H1.2

Performance: After the 5-month time gap, participants' performance fell back to nearly the level when they had no system dynamics training. As shown in Table 5.5, it fell from 68% in Phase 1 to 3% ($\chi2 = 34.9$, $df = 1$, $P<0.001$). This clearly shows that participants did not retain system dynamics concepts to an extent that would help them deal with this task.

Table 5.5: Correctness

	SD(OL)	SD(R)
Correct (Both segments correct)	**68.3 %**	**3.3 %**
Incorrect	**31.7 %**	**96.7 %**
1. Only one segment correct	5 %	19.4 %
2. No segment correct	25 %	51.6 %
3. Blank answer	1.7	25.8 %

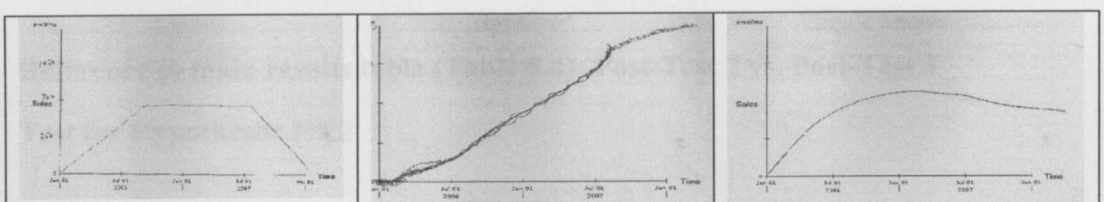
A detailed analysis of incorrect responses revealed that participants drew the graph of the key factor instead of the actual variable. This occurred in both situations (Post-test 1 and Post-test 2). For instance, in Post-test 2, the majority of those who drew both the segments incorrectly drew the behaviour of advertising (plateau-shaped) as opposed to the correct one (overshoot and collapse). In Post-test 1, this type of error was shown by 12% of the participants. Other incorrect responses in Post-test 2 included graphs that depicted different behaviours such as linear growth and linear growth followed by saturation. The results from Post-test 2 shows that participants failed to take *all* important variables and their relationships into account before drawing the graph. 26% of the participants did not attempt to draw the graph in Post-test 2 as opposed to only 2% in Post-test 1.

Forecasting Accuracy: The fact that the majority of the responses in Post-test 2 were incorrect led to a high MAPE in Post-test 2 (47%). However, participants' forecasts were much more accurate in Post-test 1 (11%). The results of forecast accuracy are shown in Table 5.6.

Table 5.6: Forecast Accuracy

	SD(OL)	SD(R)
Accuracy (MAPE)	**10.6%**	**47.0%**
1. T1	11.8%	56.9%
2. T2	12.1%	55.8%
3. T3	8.0%	28.4%

Some of the typical responses from Post-test 2 are shown below (Figure 5.2).



2(a) 2(b) 2(c)

Figure 5.2: Typical responses from Post-test 2

Understanding of feedback: As seen in Table 5.7, the percentage of participants who could identify feedback loops decreased to 10% (from 80% initially). This significant change ($\chi2 = 41.0$, df = 1, $P<0.001$) shows that representation of feedback structure was highly sophisticated after system dynamics training, but this learning was not retained over a period of time.

Table 5.7: Feedback Thinking

	SD(OL)	SD(R)
Feedback Thinking	**78.3 %**	**10 %**
1. Two feedback loops	70 %	5 %
2. One feedback loop only	8.3 %	5 %
Linear Thinking	**21.7 %**	**90 %**
3. Cause and effect relationships only	15 %	40 %
4. List of variables only	6.7 %	50 %
5. Blank answer	0 %	0 %
Total	**100 %**	**100 %**

Confidence: After the retention interval, participants' confidence fell by 69% to 6% in Post-test 2 ($\chi 2 = 38.5$, df = 1, $P<0.001$) showing that participants were aware of their lack of ability to solve the complex task.

5.4.3.2 Comparison: Is retention affected by Powersim re-familiarisation

Reference to main results table (Table 5.4): Post-Test 2 vs. Post-Test 3

Test for Hypothesis: H2.2

Performance: Training with Powersim tutorial led to an increased number of participants that attempted the task correctly (from 3% to 16%). However, this increase was not statistically significant ($\chi 2 = 2.95$, df = 1, $P<0.1$). Table 5.8 shows a detailed break-up of the responses.

Table 5.8: Correctness

	SD(R)	SD(PR)
Correct (Both segments correct)	**3.3 %**	**16.1 %**
Incorrect	**96.7%**	**83.9 %**
1. Only one segment correct	19.4 %	19.4 %
2. No segment correct	51.6 %	48.4 %
3. Blank answer	25.8 %	16.1 %

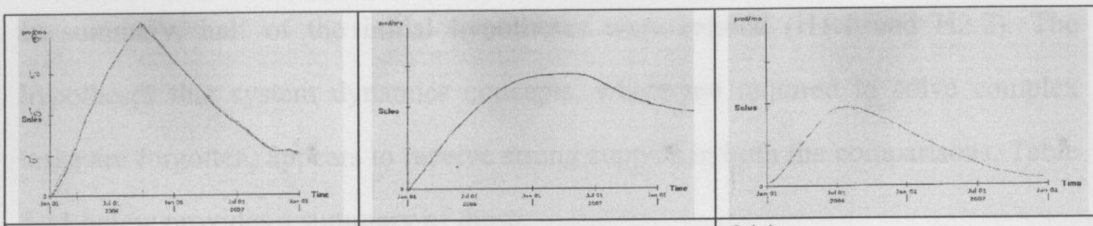
A detailed analysis of the incorrect graphs in Post-test 3 shows that in this test as well, those who could not draw both the segments correctly, also took only advertising into account, thereby drawing a plateau-shaped graph. A similar result was observed in Post-test 2. 16% of participants did not draw any graph in Post-test 3 as opposed to 26% who did not draw the graph in Post-test 2.

Forecasting accuracy: The fact that some of the participants were able to use Powersim to arrive at the correct graph improved the forecast accuracy from average MAPE of 47% to average MAPE of 39%. Detailed responses are shown in Table 5.9.

Table 5.9: Forecast Accuracy

	SD(R)	SD(PR)
Accuracy (MAPE)	**47.0%**	**39.0%**
1. T1	56.9%	47.5%
2. T2	55.8%	44.7%
3. T3	28.4%	24.8%

Some of the typical responses from Post-test 3 are shown below in Figure 5.3.



3(a)	3(b)	3(c)

Figure 5.3: Typical responses from Post-test 3

Understanding of feedback: Corresponding to the increase in performance, participants' understanding of feedback loops improved, although not significantly, as compared to when measured without re-familiarisation of Powersim ($\chi 2 = 3.04$, df $= 1$, $P<0.1$). As seen in Table 5.10, sixteen percent more participants were able to identify feedback loops that were essential in inferring the behaviour of the variable. This indicates that the interaction between the participant and the formal model led to some increase in the understanding of feedback loops, although not significantly.

Table 5.10: Feedback Thinking

	SD(R)	**SD(PR)**
Feedback Thinking	**10 %**	**26%**
1. Two feedback loops	5 %	16 %
2. One feedback loop only	5 %	9.6 %
Linear Thinking	**90 %**	**74 %**
3. Cause and effect relationships only	40 %	55 %
4. List of variables only	50 %	19 %
5. Blank answer	0 %	0 %
Total	**100 %**	**100 %**

Confidence: Confidence increased to 19% in Post-test 3, compared to only 7% in Post-test 2. An improved performance and understanding seemed to have boosted confidence in participants.

In summary, half of the initial hypotheses were refuted (H1.1 and H2.2). The hypotheses that system dynamics concepts, which are required to solve complex tasks are forgotten, appears to receive strong support in both the comparisons. Table 5.11 below provides a summary of these.

Table 5.11: Summary of Results of Hypotheses

Question	Expectation	Result	Hypotheses
What aspects of the system dynamics methodology were retained after 5 months?	Simple: SD(OL) > SD(R)	Simple: SD(OL) = SD(R)	H1.1—Rejected
	Complex: SD(OL) > SD(R)	Complex: SD(OL) > SD(R)	H1.2—Accepted
Is retention affected by Powersim re-familiarisation?	Simple: SD(P-R) = SD(R)	Simple: SD(P-R) = SD(R)	H2.1—Accepted
	Complex: SD(P-R) > SD(R)	Complex: SD(P-R) = SD(R)	H2.2—Rejected

5.4.4 Self-Reporting Questionnaire

Participants were asked if they used any system dynamics tools during the five-month retention interval (Appendix). The results show that, by and large, participants did not use these tools. For instance, one participant reported that he used the causal loop diagrams to analyse complicated situations; while another participant reported that he used it for thinking about situations like "*the consequences or the cause of my PhD achievement*". Twenty-three percent of the participants reported that they used causal loop diagramming. One participant used causal loop diagramming to understand the relation between events in auditing. These examples show that some participants made use of causal loop diagramming to solve diverse issues, which indicates some transferability of these skills. Almost none of the participants used formal modelling during the retention interval. Only one participant of the participants reported any exposure to Powersim or any similar system dynamics software.

Overall, 55% of the participants reported that the system dynamics intervention changed their approach to deal with problems. Nineteen percent of the participants thought that the intervention changed their style of thinking about the problem. This eventually led to better decision-making. While some reported that the intervention helped them in general, to "*think logically*" and "*stimulate creative thinking*", others reported that the intervention had helped them to think "*clearer*" resulting in better decisions. Another 19% reported that the intervention helped them focus more on the relationships between factors involved in solving a problem. For instance one participant wrote "*it is a very good exercise to work out the relation between various variables involved in problem solving*" while another reported that "*it provides a good platform to structure problems and makes it easier to identify cause and effect*". Another participant wrote that he "*adopted a systemic approach to problem solving especially where multiple causes, factors and variables are involved*". Six percent of the participants found the intervention specifically useful for problem solving, but did not detail as to how the intervention specifically helped them in solving problems. For instance one participant reported that the intervention helped him at "*looking at the problem more rationally*", while another wrote that the intervention helped in "*dealing with problems more effectively*". Ten percent of the participants reported that the intervention had been useful, but did not specifically mention the aspects they thought had changed as a result of the intervention. Thirty-two percent of the participants reported that the intervention had made no changes to their style of thinking or helped them in any way. The remaining (13%) participants did not answer this question.

5.5 Discussion

Previous work has suggested that a one-off system dynamics intervention followed by the use of system dynamics-based simulators made long-term changes to our mental models. Specifically, our systemic thinking abilities improved over time due to these interventions. However, prior work did not evaluate the extent to which sub-components of the system dynamics training were retained, and by those who were not involved in use of such tools after the initial training. The current study tested the retention of system dynamics skills after a five-month period for a simple and complex task. The findings suggest that participants remembered how to solve the simple task, but were unsuccessful in solving the complex task.

It was expected that participants would have forgotten even simple stock/flow concepts after the five-month retention gap. On the contrary, results suggest that participants performed as well as before. It may be recalled that the understanding of stocks and flows is fundamental to the understanding of dynamic systems. Results from the previous studies (Chapter 3 and Chapter 4) suggested that baseline performance was sub-optimal and the lack of this skill can successfully be overcome through system dynamics training. The results from the current study are therefore quite encouraging. People encounter such one-stock systems with no feedback in their daily life. But they also fall in the trap of making typical errors while calculating the time period where the value of the stock is highest. The one-off system dynamics training seems to rectify these errors for the long term. The findings also suggest that most participants do not use the time-saving, visual strategy. This may be because either participants had forgotten this method or they wanted to use the tried and tested computational method. The former is unlikely as the visual method merely involves noticing the time period in the table where the outflow exceeds the inflow. It therefore seems that participants tend to employ the traditional method even if they are knowledgeable about a better technique.

The re-familiarisation with Powersim did not significantly improve the performance on the simple task. As discussed in the previous chapters, the simple task does not necessarily require the use of Powersim. The use of Powersim could have encouraged participants to use the visual strategy. However, this was not the case. Results suggest that the Powersim practice did not boost the number of participants who used the visual strategy. These results indicate that participants probably did not use Powersim for the simple task even though they could. The practice with the software also had no effect on participants' feeling of confidence. These findings confirm the results from previous chapters that system dynamics software does not play a significant role in understanding simple stock/flow tasks.

The results from the complex task suggest that most participants were not able to solve the task after the five-month retention interval. There are many reasons for this dismal performance. An overarching and evident reason is that the absence of exposure to these tools during the retention phase caused these results to deteriorate. The analyses also indicate towards specific concepts that were retained and those that were forgotten. In particular, participants' ability to identify stocks and flows was retained. This is also clear from the results of the simple task. On the other hand, the level of feedback thinking in participants' mental models went back to a level similar to that recorded prior to the system dynamics intervention. Just like in the Pre-test, participants tended to associate the behaviour over time of the variable in question with that of another key input. Though the performance after the retention interval was in no way near that measured immediately after the intervention, it was slightly better than the results recorded for the baseline condition. Specifically, whereas in the Pre-test, 85% of the participants could not get even one segment of the output graph correct, this figure dropped to 52% in the retention test. In general, these results show that though participants learnt to recognize key feedback loops during the original intervention, this skill did not develop for the long-term. Another factor that might have contributed to this performance is the inability to use the simulation software. The majority of participants constructed incomplete Powersim models during the retention test. This could be because of either not being able to conceptualize the model clearly to be simulated using Powersim or being able to do the former but having forgotten the use of the software itself.

As it turned out, there was no significant improvement in performance when participants were allowed to practise Powersim. However, 13% more participants were able to get the correct graph. Also, 16% more participants were able to recognize feedback loops. This indicates that even a short Powersim re-familiarisation might be helpful in reinforcing key concepts needed to solve the complex task.

Another point worth discussing is the huge drop in confidence during the retention testing phase. Participants felt much less confident about solving the complex task at the retention testing phase as compared to immediately after the original intervention. A possible explanation of this can be that participants realised that they had not answered the task correctly. In the previous study too (Chapter 4), participants seemed realistic about their confidence in the complex task. The Powersim re-familiarisation led to some increase in this figure (from 6% to 19%). This too was in sync with their performance. This suggests that participants' performance in complex task was a reasonable indicator of their confidence. However, as noted previously (in Chapter 3 and Chapter 4), this wasn't true for the simple task, where participants had suffered from overconfidence.

Like confidence, forecasting accuracy also first plummeted during the retention testing phase and then improved after Powersim re-familiarisation. A possible justification of this trend can be provided by analysing the correct responses at each stage. Correct graphs were more accurate than the incorrect ones. The highest performance was during the original learning stage, which fell drastically during the retention test and then improved for the last test.

Similar to the results reported in Cavaleri and Sterman (1997), in the current study too, system dynamics training was partially successful in the retention of system dynamics skills.

Table 5.12: Comparison of Participants from Cavaleri and Sterman (1997) and the Current Study

	Cavaleri and Sterman (1997)—(Managers)	**Cavaleri and Sterman (1997)—(Non-Managers)**	**This study**
Participants' experience	Managerial	Non-managerial	Non- managerial
Content of training-explicit discussion of the problem to be solved	Yes	No	No
Explicit usage of ST tools post initial training	Yes	No	No
Exposure to systems thinking	High (Beer game + MFS specific to the problem + ST tools seminar)	Low (Beer game)	High (ST tools seminar + formal modelling seminar and practice)

However, it should be noted that there are two key differences between their study and the one reported here. First, whereas the experiment reported here was conducted in a controlled setting, the former was not. Second, participants in Cavaleri and Sterman (1997) were involved in the use of system dynamics tools during the retention interval as well. Hence, a perfect comparison between results from the current study and that of Cavaleri and Sterman (1997) may not be possible. Nevertheless, some interesting comparisons can still be made. In Cavaleri and Sterman (1997), the two cohorts (managers and non-managers) were exposed to different levels of systems thinking. As shown in Table 5.12, non-managers in Cavaleri and Sterman (1997) and participants in this study are comparable in three characteristics—usage of systems thinking tools post training, whether the training incorporated the actual problem to be solved, and their level of experience. However, in this study, participants were exposed to much more systems thinking than non-managers in Cavaleri and Sterman (1997). It was equivalent to that provided to managers in Cavaleri and Sterman (1997). Interestingly, in both the studies, even though 'performance' showed no improvement in the follow-up phase, participants felt that the initial training had been useful. As stated previously, performance of participants in this study plummeted for the complex task. In the context of Cavaleri and Sterman (1997) also, organisational performance decreased after the training. Contrary to the dismal performance, the self-reporting questionnaires in both studies tell a different story. As a result of the training, in Cavaleri and Sterman (1997), managers felt that they were more 'aware' of the changes to their own thinking and thinking about company's needs and how decisions were made in the company. They also reported positive changes in the way they interacted with others and the way they did their job. Non-managers were unsure whether the training helped them in any of the above. Both managers and non-managers reported that the training helped them to be aware of how the company was managed and that overall the training was valuable. In this study too, the majority of the participants felt that the training was valuable and that it had changed their approach of problem solving. Hence, both studies show that, in general, systems thinking/system dynamics training is viewed positively by participants. This study confirms some of the findings of Cavaleri and Sterman (1997), but this time within a rigorous experimental setting.

A similar study by Huz et al. (1997) was conducted in a controlled setting, unlike Cavaleri and Sterman (1997). The length of intervention to participants in Huz et al. (1997) was similar to that provided in this study. However, a major difference between this study and that conducted by Huz et al. (1997) is that, in Huz et al. (1997) the intervention was designed for a specific domain where participants were involved in the building of a system dynamics model with the support of experts. Also, participants used their systems thinking skills after the intervention. Both these characteristics were absent in the present study. Shifts in participants' mental models in the follow-up phase were evaluated using self-reporting questionnaires in Huz et al. (1997) and hence these are compared with results obtained from the self-reporting questionnaire in this study. In both studies the intervention was perceived positively by the participants. Specifically, in Huz et al. (1997), participants reported that their 'goal structure' and 'change strategies' had significantly changed and that they were more aligned in their perception of systems goals. In the present study, participants, in general, found that the intervention had helped them in thinking rationally about a problem. This shows that systems thinking interventions are perceived as not only useful when used in specific problem domains, but the skills learnt during the intervention are perceived as transferable to other contexts.

Implications for teaching and learning of system dynamics skills

The findings suggest that qualitative system dynamics skills were retained by participants but quantitative system dynamics tools were not. There could be various reasons for this. Qualitative tools are relatively easier to learn. This is because causal loop diagrams and system archetypes, extend concepts of cause and effect and the principle of accumulation with which participants are familiar. One does not require special software for drawing a causal loop diagram. Some participants applied these diagrams to examples outside the scope of the interventions. Another advantage of causal loop diagram is that it does not require expertise to build such informal models. On the other hand, quantitative system dynamics tools, like any other software tool, take a while to be completely mastered. For instance, a Powersim model cannot be simulated unless the model is completely void of inconsistencies. Even re-familiarisation with Powersim was not of much help. Even though qualitative tools were retained during the five months, their applications to situations that involve dynamic complexity are limited. The more useful quantitative tools were not retained. There are two possible options to tackle this loss. In situations where the system dynamics intervention is designed for a specific purpose such as those described in Huz et al (1997) and Cavaleri and Sterman (1997), the knowledge gained from the intervention could be converted into a simulation game. This management flight simulator can then be used to understand the dynamics of the issue even after the intervention is over. Of course this would involve at least one expert system dynamics modeller to design the simulator. Another option is to refresh memories of participants using follow-up training sessions. However, in a typical higher education system, there is no straightforward way by which such refresher courses might be offered.

5.6 Conclusion

This chapter has argued the necessity, and discussed the lack of follow-up studies, to test the efficacy of system dynamics training. The purpose of the present study was to determine if components of system dynamics training are retained after some time has elapsed. To do so, an experimental condition was set where participants underwent two Post-tests, five months after they had undergone system dynamics training. Their performance and understanding was recorded using comparable tests—before and after practice with system dynamics software.

The study has found that generally, in the absence of retrieval cues, tools learnt during system dynamics training are forgotten. Specifically, participants neither remembered qualitative nor quantitative tools to an extent that enabled them to solve complex tasks. Only a few participants could recognise feedback structures of a system or represent these in the simulation software. This inability led to participants being unable to infer correct behaviour over time of the variable in question. However, not all concepts were completely forgotten even after five months. Participants were able to discern between stocks and flows of the system and were also able to compute the time period when a stock reaches its maximum. This enabled them to perform well in the simple task.

Participants reported changes in their style of thinking as a result of the initial training. The majority of them claimed that they thought more logically than before when solving a problem. In the absence of further training and any system dynamics applications, some of the participants used qualitative tools in routine work. The use of formal tools was limited. Qualitative tools, like causal loop diagrams and associated system archetypes are based on concepts of cause and effect which most participants are familiar with. Extending these concepts to form simple causal loop diagrams can be useful to describe complex problems. Moreover, drawing a causal loop diagram does not require special software. However, building error-free formal models requires practice with the software, which most participants might have been unwilling to undertake.

The results of this study indicate that one-off system dynamics training does not fundamentally improve peoples' performance and understanding in complex tasks. The training does improve some understanding of dynamic complexity for the long-term—such as how to deal with simple stock/flow tasks.

It is important to make note of the fact that in this case, participants did not use the training in a system dynamics based application during the five-month period, which may have reinforced the concepts resulting in better performance and understanding. Further work needs to be done to establish the role of usage of system dynamics during the retention interval on performance and understanding. Also, the design of the previous experiment was such that all participants that underwent qualitative training eventually underwent quantitative training as well. Hence it was not possible to investigate the retention in participants who were initially trained *only* in qualitative system dynamics. Further research is required in this area as well.

Chapter 6
Conclusions, Implications and Future Work

6.1 Introduction

The main objective of this thesis was to experimentally test the relative efficacy of the two phases of the system dynamics methodology, both individually and when applied together (combination system dynamics). The level of decision aid (no aid, qualitative, quantitative or both) and task complexity (simple or complex) were used as independent variables to evaluate performance and understanding. The thesis not only explored the immediate effect of system dynamics interventions but also their long-term effectiveness.

We started our investigation with an experiment that tested the relative efficacy of qualitative system dynamics, combination system dynamics and baseline performance. The next study evaluated the role of qualitative, quantitative and combination system dynamics. As before, task performance after these tools was measured against baseline condition and against each other. The final study investigated the retention of system dynamics interventions.

This chapter looks back at the current research to highlight its contributions and how these were achieved and also to look ahead to suggest avenues for further research. The chapter also discusses implications for the practise of system dynamics and limitations of the current thesis.

6.2 Conclusions about the research problem

This section discusses overall conclusions of the thesis. Specific conclusions related to each hypothesis are detailed in the next section. Within the context of the nature of tasks discussed in the thesis, this research has found that system dynamics tools either applied individually or in combination with each other are in general useful in solving both simple and complex tasks, i.e. those who undergo system dynamics interventions perform better than those who do not. The extent of the effectiveness of different system dynamics tools however varies.

Specifically, qualitative tools alone improve our understanding of stocks and flows and feedback loops. These are fundamental concepts required to understand dynamic complexity. Our findings indicate that availability of qualitative system dynamics tools ensures significantly higher performance in simple tasks when compared with the baseline condition, and equivalent performance when compared with quantitative only and combination system dynamics tools.

Qualitative tools however do not enable the forecasting of behaviour of a variable over time in a relatively complex system. In most situations this requires simulating the system using computer software. Quantitative system dynamics tools lend their computer modelling capabilities that enable such forecasting. This leads to significantly higher performance when compared to performance after qualitative intervention alone. For simple stock/flow tasks too, the knowledge of quantitative system dynamics results in significantly higher performance than in the baseline condition.

The complete system dynamics process, i.e. the combination of qualitative and quantitative system dynamics, resulted in highest performance for both simple and complex tasks. However, performance with the combined intervention was not significantly higher than qualitative intervention alone or quantitative intervention alone for the simple task. This shows that for the simple task, quantitative system dynamics tools may not be necessary. On the other hand performance after combined intervention was not significantly higher than quantitative intervention alone for the complex task. This implies that in such cases qualitative intervention does not have a significant role. In the long-term, concepts learnt during qualitative intervention are retained even in the absence of retrieval cues. Unlike qualitative intervention, a one-off quantitative intervention does not ensure long-term retention of these concepts.

The highlights of the thesis are:

- The use of a rigorous experimental approach to measure the efficacy of system dynamics interventions.
- Relying on 'model-building', rather than mere exploration of a pre-made model to test system dynamics efficacy. This approach tests the effect of the process rather than the effect of playing a simulation game.
- Testing efficacy of individual components of the system dynamics process (qualitative and quantitative), as well as their combination against baseline condition, i.e. performance without the aid of these tools.
- Testing the relative efficacy of qualitative system dynamics, quantitative system dynamics and combined system dynamics.
- The evaluation of the long-term efficacy of the system dynamics method using a longitudinal experiment.
- Using task complexity as the independent variable.

6.3 Conclusions about hypotheses

Findings of each hypothesis are summarised below. These findings are also explained in the context of previous experimental research. Detailed comparisons are made within relevant chapters (Chapters 3, 4 and 5).

6.3.1 Qualitative versus baseline

Hypothesis H1 stated that participants who attended qualitative system dynamics training will perform better on simple tasks but not on complex tasks when compared to participants who did not undergo any system dynamics training, i.e.,

H1.1 Qual[1] >[2] Baseline[3] for simple tasks

H1.2 Qual = Baseline for complex tasks

Based on the results reported in Chapter 3 and Chapter 4, Hypotheses H1.1 and H1.2 were accepted. Both experiments showed that the qualitative intervention was able to bring about significant improvement in performance in simple task, but not for the complex task.

These results reveal two important findings. First, they reveal that performance without system dynamics is sub-optimal. This applies to both simple and complex tasks. Results show that people do not understand the concept of accumulation and fail to distinguish between stocks and flows. These results are in line with those of Sweeney and Sterman (2000), Cronin and Gonzalez (2007) and Cronin et al. (2007). The second revelation pertains to the effectiveness of qualitative interventions. Some studies have tested the efficacy of qualitative system dynamics interventions on the bathtub tasks. In a seminal study, Kainz and Ossimitz (2002) demonstrated that training in qualitative stocks and flows could significantly improve performance. Our results demonstrate the same.

[1] Qual: Results obtained after qualitative SD intervention

[2] Significant difference in results. e.g. Quant > Qual implies that the performance under quantitative SD will be superior to performance under qualitative SD

[3] Baseline: Performance of participants prior to any intervention

6.3.2 Quantitative versus baseline

Hypothesis H2 stated that participants who attended quantitative system dynamics training will perform better on both simple and complex tasks when compared to participants who did not undergo any system dynamics training i.e.,

H2.1 Quant[4] > Baseline for simple tasks

H2.2 Quant > Baseline for complex tasks

Hypotheses 2.1 and 2.2 were accepted. Results reported in Chapter 4 confirm that those who underwent quantitative system dynamics performed significantly better than those who did not undergo any intervention.

6.3.3 Qualitative and quantitative versus baseline

Hypothesis H3 stated that participants who attended both qualitative and quantitative system dynamics training will perform better on both simple and complex tasks when compared to participants who did not undergo any system dynamics training, i.e.,

H3.1 (Qual, Quant)[5] > Baseline for simple tasks

H3.2 (Qual, Quant) > Baseline for complex tasks

Hypotheses H3.1 and H3.2 were found to be true and hence accepted. Our findings from the experiment reported in Chapter 5 indicated that those who underwent the combination intervention performed significantly better than those without system dynamics decision aid.

[4] Quant: Results obtained after quantitative SD intervention

[5] (Qual, Quant): Results obtained after a combination of qualitative and quantitative SD intervention

The majority of experimental studies that have tested the efficacy of system dynamics interventions in the past have administered the combined intervention. As argued in Chapter 2, the combination approach is widely accepted as *the* system dynamics methodology. A study by Pala and Vennix (2005) reveals that participants' performance significantly improved in the simple bathtub task. These results are on the same lines as that of the current research. However, Pala and Vennix (2005) and the current study disagree on results of the complex task. Pala and Vennix (2005) reported that the combined intervention failed to improve participants' performance in relatively complex tasks. In fact performance significantly deteriorated in one of these tasks. The authors argue that this occurred due to participants' inability to understand time delays and due to the overall complexity of the tasks. The current study used a different complex task, though the aim was the same—to draw behaviour of a key variable over time. In the current study it was found that the combination intervention led to the significantly higher performance in the complex task. Another study that combined qualitative and quantitative intervention using debriefing, causal loop diagrams and management flight simulator found that the intervention produced significant changes in the degree of feedback thinking when compared to baseline condition (Doyle et al. 1998). In the current study also, the number of participants who could identify feedback loops increased significantly after the combined intervention. Our study also revealed that it is the qualitative intervention that largely contributes to understanding of feedback.

6.3.4 Qualitative versus quantitative

Hypothesis H4 stated that there will not be any significant difference between results obtained from qualitative, quantitative and combined intervention for a simple task. But for a complex task, results obtained from the combined intervention will be significantly higher than those from quantitative intervention, which in turn will be significantly higher than those obtained after qualitative intervention, i.e.

H4.1 (Qual, Quant) = Quant = Qual for simple tasks

H4.2 (Qual, Quant) > Quant > Qual for complex tasks

H4.1 was accepted. Results from Chapters 3 and 4 suggest that for the simple task, none of the interventions were significantly better than the other two. It was evident that participants' understanding of stock and flow problems is initially quite poor. As these situations are commonly encountered, participants' confidence in such tasks even without system dynamics is high. As both qualitative and quantitative interventions deal with stocks and flows, it appears that neither of these is sufficient to improve stock/flow understanding.

H4.2 was rejected. It was initially asserted that the combination intervention would lead to significantly better performance when compared with quantitative intervention. Our findings from Chapters 3 and 4 indicate that this was not the case. Results revealed the following relation:

(Qual, Quant) = Quant > Qual for complex tasks.

These results are in support of the anecdotal evidence that suggests that quantitative tools alone are better than qualitative tools individually or in combination of quantitative tools (e.g. Warren 2004).

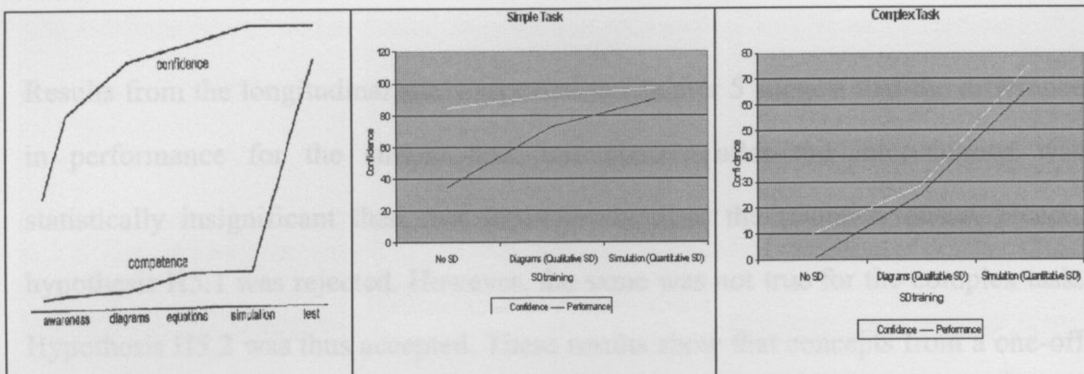
Qualitative tools have some clear benefits over quantitative tools with respect to (i) the time taken to master these tools (ii) the time taken in the application of these tools and hence the amount of money required for such applications and (iii) the need of a computer and special software. As qualitative tools do not require computer modelling, they may be taught to an audience that may find it difficult to master the art of modelling. Teaching quantitative system dynamics also requires a substantial amount of time, which may not be available. Similarly, applying quantitative system dynamics requires time as well as money to be available for the modelling team. It also requires the presence of expert modellers. Further, qualitative tools are not computer or software dependent as they do not require a computer and special system dynamics software to run simulations.

For reasons stated above, qualitative tools appear to be more relevant in situations where significant insight could be gained by the application of such tools. Examples of such situations being: understanding of a system through description of the relationships between system components, explanation of problems using system archetypes, using stocks and flows to explain dynamics of the system and constructing a dynamics hypothesis from the feedback loop diagrams.

Quantitative tools on the other hand are advised when the situation demands accuracy in forecasting the behaviour of the system. Like qualitative tools, quantitative tools also assist in development of systemic view, identification of stocks and flows and identifying feedback loops.

When time and money are not significant factors in problem solving, then the combination of qualitative and quantitative tools could produce best results. The understanding gained through the qualitative phase acts as a precursor to the quantitative modelling.

Anecdotal evidence suggests that in general, system dynamics improves one's confidence in solving tasks. Peterson and Eberlein (1994) suggest that confidence largely increases during the qualitative phase and not in the quantitative phase. They argue that competence follows the opposite trend (Figure 6.1(a)).



6.1 (a) Peterson and Eberlein (1994)	6.1 (b) Simple Task—Second Experiment	6.1 (c) Complex Task—Second Experiment

Figure 6.1: Performance versus confidence

These claims are somewhat in line with the finding of the current investigation. The trend of competence suggested by Peterson and Eberlein (1994) is similar to the one obtained for a complex task. In both cases, the larger contribution is from the quantitative phase. The trend for confidence suggested by Peterson and Eberlein (1994) does not match well with either the simple or complex task from our findings. Our findings suggest that participants are initially over-confident for the simple task but not for the complex task. After the interventions, though the level of confidence is more realistic than before. This is true especially after the quantitative intervention (Figures 6.1(b) and 6.1(c)). This aspect of convergence of competence and performance as found by the current research is shared by Peterson and Eberlein (1994) as well (Figure 6.1 (a)).

6.3.5 Original learning versus retention learning

Hypothesis H5 states that for both simple and complex tasks, participants' performance would have deteriorated after 5 months, i.e.

H5.1 SD (OL)[6] >[7] SD (R)[8] for simple task

H5.2 SD (OL) > SD (R) for complex task

Results from the longitudinal study reported in Chapter 5 suggest that the difference in performance for the simple task immediately after the interventions was statistically insignificant than that immediately after the retention phase. Hence, hypothesis H5.1 was rejected. However, the same was not true for the complex task. Hypothesis H5.2 was thus accepted. These results show that concepts from a one-off system dynamics intervention are retained only to a small extent.

System dynamics interventions are used to train consultants in industry as well as train students in universities. It may be, in some cases, that this training is one-off and trainees do not undergo further training. In many of these cases, those who get trained do not end up using system dynamics in the short term after the intervention. This would particularly be in the case of students who enrol in a system dynamics course at university. This research has found that, trainees retain and also apply qualitative system dynamics tools after the training. Hence, fundamental system dynamics concepts are learnt for the long-term. This is an encouraging result for system dynamics trainers. Though many argue that quantification is an essential component of system dynamics methodology, unfortunately, results reveal that this skill is forgotten in the absence of further practice.

Even though the majority of participants could not retain system dynamics concepts to enable them to solve the complex problem, they still perceived the interventions as useful. This is an encouraging finding for the system dynamics community. It is not

[6] SD (OL): Original Learning: results obtained immediately after system dynamics training in October 2006.

[7] Significant difference in results.

[8] SD (R): Retention: results obtained after the end of retention interval in February 2007.

surprising that some of the concepts originally learnt were forgotten. This led to a decrease in confidence and task performance. Irrespective of this loss, participants had positive feelings towards the interventions. Positive feelings of system dynamics interventions have been reported in previous research (Cavaleri and Sterman 1997 and Huz et al. 1997).

6.3.6 Retention learning versus Powersim re-familiarisation

Hypothesis H6 suggested that performance in the simple task remains the same before and after Powersim re-familiarisation, but improves for the complex task, i.e.

H6.1 SD (P-R)[9] = SD (R) for simple task

H6.2 SD (P-R) > SD (R) for complex task

As hypothesised, performance after Powersim re-familiarisation was not significantly different than that observed before the re-familiarisation. This led to the acceptance of hypothesis H6.1. It was thought that Powersim re-familiarisation would lead to a significant improvement in results for the complex task. This was not the case as indicated by the statistical tests. Hypothesis H6.2 was therefore rejected.

We earlier mentioned that barring a few examples, there has been paucity in experimental studies that test the efficacy of system dynamics methods. Sweeney and Sterman (2000) claim:

> *"The use of systems thinking and system dynamics is increasing dramatically, yet there is little evidence, or even systematic research, to support educators' and consultants' faith in its efficacy".* (p249)

[9] SD (P-R): Retention with Powersim: results obtained after participants re-familiarize with Powersim.

Cavaleri and Sterman (1997) assert:

> *'Although various intervention techniques that fall under the rubric 'systems thinking' have become quite popular, little is known about their efficacy in enhancing organizational effectiveness or productivity. With few exceptions the relationship between the use of systems thinking and organizational performance remains the province of anecdote rather than rigorous follow up research.'* (p171)

Similar assertions have been made by others, such as Doyle (1997), Maani and Maharaj (2004), and Pala and Vennix (2005). The experiments in the current research address the gap mentioned above.

Furthermore, the controversy surrounding the relative usefulness of qualitative and quantitative tools may be recalled from Chapter 2 (Literature Review). Questions pertaining to 'qualitative versus quantitative' system dynamics have been a highly debated topic that has highlighted crucial research questions such as:

> *"The field must address the relationships between qualitative mapping and quantitative modelling—in short, when to map and when to model"* Richardson (1999) p8

And,

> *"How can we* measure *the value added by the extra effort of simulation or the value lost by not simulating?"* Coyle (2001) p359

The current research addresses these questions by comparing the relative efficacies of qualitative mapping, quantitative modelling and a combination of the two.

6.4 Contributions transferable to other studies in system dynamics

Some components developed for the current research could be used by other experimental studies in system dynamics. These are discussed below.

6.4.1 Instrument to measure system dynamics effectiveness

The current research provides three complex tasks that can be used as instruments to measure system dynamics interventions. Each of these tasks is a two-stock task that involves feedback loops. These tasks add to the bathtub tasks (Sweeney and Sterman 2000), and provide a great way to test the effectiveness of system dynamics interventions.

6.4.2 Experimental design

The design of the second experiment is rare in the system dynamics community. It can be used to test the effectiveness of a methodology that can be split into two stages. The design enables the measurement of the complete methodology and also the measurement of individual phases of the method. The pre-test/ post-test design allows for the measurement of task performance before any intervention is administered as well as after the intervention is administered. The design enables the comparison of two individual stages of the method individually and combined together with baseline condition and with each other.

6.4.3 Methodology

Experimental studies that use rigorous methodology to measure system dynamics interventions are virtually absent (Doyle et al. 1998). The few studies that follow a scientific method include those that rely on debriefing and outcome feedback (e.g. Doyle et al. 1998). The current study provides an example of a study that uses rigorous methods such as having a pre-test/ post-test, measuring changes in individuals, using a sample size that is large enough, eliminating any learning effect and measuring short-term as well as long-term changes.

6.4.4 Dynamic decision making

The experiment could be extended to a dynamic decision making setting as well. Many system dynamics based management flight simulators already exist. These have been discussed in detail in Chapter 2. It may be recalled that in MFS-based studies, a participant typically does not model the scenario using causal loop diagrams or stock/flow models. These models are pre-supplied and the participant merely explores these models to improve understanding and make decisions. The experimental design proposed in the current thesis can be adapted by MFS-based studies that intend to test system dynamics' efficacy to dynamic decision making.

6.5 Limitations

The limitations of experimental design and methodology have been previously discussed Chapters 3, 4 and 5. Here, we summarise some of the more general limitations of this research.

Experimental research methods have been criticised for the lack of external validity. However, the benefits of this approach outweigh this drawback. The laboratory experimental setup provides an opportunity to control variables other than the ones in question. Only in an experimental setting can one be sure that it was really the treatment that caused the difference between the performance of the participants in the control and experimental groups.

All three experiments used students as participants. It may be argued that since the experiment was conducted in a university setting, these results cannot be widely generalized. However, it should be noted that understanding the case studies and the tasks did not require expertise or detailed knowledge of a particular domain. Hence, not having substantial industry exposure does not in any way undermine the findings. Participants were provided with substantial monitory incentive to motivate them to take the task seriously. It should also be noted that in the past, a majority of such experimental studies have relied on tertiary students (Rouwette et al. 2004) and that postgraduates students are a good substitute for managers (Remus 1996). It should also be noted that all participants had at least one degree and some prior industry exposure.

Though many system dynamics interventions occur in group settings, the aim of the current research was to explore the effect of system dynamics intervention on individuals. Typically in group model building, an expert builds a system dynamics model with the help of a group of clients. However, group interventions suffer drawbacks such as the discussions being dominated by some individuals that might lead to non-participation of others, participants complying with the group and also experimental bias (Doyle et al. 1996).

6.6 Implications for further research

The current investigation has opened up many areas that need further exploration. Each of these will add to more knowledge on the efficacy of system dynamics tools.

6.6.1 Variation of task complexity

It may be recalled that the simple task involves a one-stock system whereas the complex task involves two stocks. Although the complexity of the 'complex task' used in the current research is greater than that of the 'simple task', it is still less complex than most of the real world problems. A possible extension of the current research would be increase in the level of complexity to test the effectiveness of the system dynamics interventions. Task complexity could be varied by varying the number of stocks in the system and by varying the degree of feedback. Hsiao and Richardson (1999) summarise studies that use task complexity as an independent variable and the way in which task complexity is varied. Although system dynamics claims to improve understanding of dynamic complexity, the effectiveness of system dynamics interventions on other forms of complexity (see Wood 1986) could also be investigated.

6.6.2 Length of interventions

The increase in task complexity might require an increase in the length of system dynamics interventions. This is because a more complex situation might include more stocks, flows and other variables that contribute to the complexity of the task. Understanding the task and then representing the problem in system dynamics software would then require a higher level of understanding. This would require a more detailed tutorial on system dynamics as well as substantial exposure to system dynamics modelling software.

6.6.3 Background of participants

Another dimension along which the current research can be extended is by choosing managers as participants. People with significant experience with business systems might be able to relate more to the methodology and tasks. For example, managers who make resource allocation, forecasting and inventory decisions might form an ideal group to experiment with. However, managers might not make different decisions to those made by the participants in this research.

6.6.4 Providing system dynamics practice during retention period (longitudinal study)

The longitudinal investigation measured the retention of system dynamics intervention in the absence of further system dynamics training or the application/practice of system dynamics concepts. This opens up two areas to be further investigated. The current study reveals that quantitative tools are easily forgotten. Hence, first area of investigation could be related to measuring performance after a refresher of quantitative system dynamics tools. Second, participants could be provided with system dynamics practice exercises during the retention phase. This might help them retain system dynamics concepts.

6.6.5 Comparison with allied approaches

Another avenue to pursue could be the comparison of allied system approaches with system dynamics. Some of these allied approaches promote common methods and tools for system analysis. The experimental design used in this research could be used to compare these methods. For instance qualitative, system dynamics could be compared with the soft system methodology (Checkland 1981). The individual approaches could be then compared with a method that is a combination of the two approaches as proposed by some researchers (Lane and Oliva 1998). System dynamics methodology can also be compared with critical systems thinking (Jackson 1991) and with management cybernetics (Beer 1959). Through such comparisons researchers may be able to identify situations where (i) one method is superior to the other in specific circumstances, and (ii) if any two methods can complement each other.

6.7 Overall conclusion

The current research was undertaken to (i) fundamentally test the efficacy of system dynamics, (ii) compare the relative efficacy of individual system dynamics tools and (iii) test the long-term effectiveness of system dynamics. Through three rigorous experiments, this research has shown that: (i) system dynamics is a useful methodology to analyse complex systems, (ii) qualitative tools are sufficient alone for simple tasks, and quantification is required for complex tasks but (iii) in the long-term, this work found evidence to suggest that people may experience difficulties in retaining the ability to use quantitative system dynamics.

All these findings add to our knowledge of the short-term and long-term efficacy of system dynamics tools. In doing so, the current research has answered several questions that have plagued the system dynamics community for years.

Bibliography

Abdel-Hamid, T. (1989). "The Dynamics of Software Project Staffing: A System Dynamics Based Simulation Approach." IEEE Transactions on Software Engineering **15**(2): 109-119.

Abdel-Hamid, T. and S. E. Madnick (1991). Software Project Dynamics: An integrated Approach Upper Saddle River, NJ, Prentice-Hall, Inc.

Abdel-Hamid, T. K. (2002). "Modeling the Dynamics of Human Energy Regulation and its Implications for Obesity Treatment." System Dynamics Review **18**(4): 431-471.

Ackoff, R. (1971). "Towards a System of Systems Concepts." Management Science **17**(11): 661-671.

Ackoff, R. L. (1994). "Systems Thinking and Thinking Systems." System Dynamics Review **10**(2/3): 175-188.

Alessi, S. M. (2000). "Designing Educational Support in System-Dynamics-Based Interactive Leraning Environments." Simulation and Gaming **31**(2): 178-196.

Anderson, E. G., Jr., D. J. Morrice, et al. (2005). "The "Physics" of Capacity and Backlog Management in Service and Custom Manufacturing Supply Chains." System Dynamics Review **21**(3): 217-247.

Anderson, J. R. (2000). Learning and memory. New York, John Wiley.

Arquitt, S., X. Honggang, et al. (2005). "A System Dynamics Analysis of Boom and Bust in the Shrimp Aquaculture Industry." System Dynamics Review **21**(4): 305-324.

Arquitt, S. and R. Johnstone (2004). "A Scoping and Consensus Building Model of a Toxic Blue-Green Algae Bloom." System Dynamics Review **20**(2): 179-198.

Axelrod, R. (1976). Structure of Decision: The Cognitive Maps of Political Elites. Princeton, New Jersey, Princeton University Press

Bajaj, A (2004). "The Effect of the Number of Concepts on the Readability of Schemas: An Empirical Study with Data Models." Requirements Engineering Journal **9**: 261 - 270

Bakken, B. E. (1993). Learning and Transfer of Understanding in Dynamic Decision

Environments, M. I. T.

Barlas, Y. and I. Bayraktutar (1992). An Interactive Simulation Game for Software Project Management (Softsim). Proceedings of the 1992 International System Dynamics Conference of the System Dynamics Society, Utrecht, the Netherlands, The System Dynamics Society.

Barros , M. O., C. M. L. Werner, et al. (2002). Evaluating the Use of System Dynamics Models in Software Project Management International System Dynamics Conference, Palermo, Italy, System Dynamics Society.

Batra, D. and Wishart N.A. (2004). "Comparing a Rule-based Approach with a Pattern-based Approach at Different Levels of Complexity of Conceptual Data Modelling Tasks. The Effectiveness of a Representation" International Journal of Human-Computer Studies **61**(4): 397-419.

Bayer, S. and D. Gann (2006). "Balancing work: bidding strategies and workload dynamics in a project-based professional service organization." System Dynamics Review **22**(3): 185-211.

Beer, S. (1959). Cybernetics and Management. London, The English Universities Press.

Beer, S. (1984). "The Viable System Model: Its Provenance, Development, Methodology and Pathology." The Journal of the Operational Research Society **35**(1): 7-25.

Bertalanffy, L. v. (1950). "An Outline of General System Theory." The British Journal for the Philosophy of Science **1**(2): 134-165.

Bianchi, C. and E. Bivona (2002). "Opportunities and Pitfalls Related to E-Commerce Strategies in Small-Medium Firms : A System Dynamics Approach." System Dynamics Review **18**(3): 403-429.

Bianchi, C. e. (2002). "Special Issue : Systems Thinking and System Dynamics in Small-Medium Enterprises." System Dynamics Review **18**(3): 311-430.

Brehmer, B. (1992). "Dynamic decision making: human control of complex systems." Acta

Psychologica **81**: 211-241.

Brekke, K. A. and E. Moxnes (2003). "Do numerical simulation and optimization results improve management? Experimental evidence." Journal of Economic Behavior & Organization **50**(1): 117–131.

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

Cavaleri, S. A., M. Raphael, et al. (2002). Evaluating the Performance Efficacy of Systems Thinking Tools. Proceedings of the 20th International Conference of the System Dynamics Society, Palermo, Italy, The System Dynamics Society.

Cavana, R. Y., D. Boyd, et al. (2004). A Systems Thinking Study of the New Zealand Army Electronic Technician Trade Group. 22nd International Conference of the System Dynamics Society, Oxford, England, The System Dynamics Society.

Cavana, R. Y. and L. V. Clifford (2006). "Demonstrating the utility of system dynamics for public policy analysis in New Zealand: the case of excise tax policy on tobacco." System Dyamics Review **22**(4): 321-348.

Cavana, R. Y., P. K. Davies, et al. (1999). "Drivers of Quality in Health Services : Different Worldviews of Clinicians and Policy Managers Revealed." System Dynamics Review **15**(3): 341.

Checkland, P. B. (1981). Systems Thinking. Systems Practice. Chichester, Wiley.

Cooke, D. L. (2003). "A System Dynamics Analysis of the Westray Mine Disaster." System Dynamics Review **19**(2): 139-166.

Cooke, D. L. and T. R. Rohleder (2006). "Learning from incidents: from normal accidents to high reliability." System Dynamics Review **22**(3): 213-239.

Cooper, K. G. (1980a). "Naval Ship Production: A Claim Settled and a Framework Built." Interfaces **10**(6).

Cooper, K. G. (1980b). Strategic Analysis for Program Management: A Computer-Based Aid for the Senior Management of Large-Scale Design and Construction Programs, Pugh-Roberts Associates, Inc.

Coyle, J. M., D. R. Exelby, et al. (1999). Development Of A Suite Of Operation and Support Cost Analysis Models (OSCAM) for the Defence Community. 17th International Conference of the System Dynamics Society and 5th Australian & New Zealand Systems Conference, Wellington, New Zealand, The System Dynamics Society.

Coyle, R. G. (2000). "Qualitative and quantitative modelling in system dynamics: some research questions." System Dynamics Review **16**(3): 225–244.

Coyle, R. G. (2001). "Rejoinder to Homer and Oliva." System Dynamics Review **17**(4): 357-363.

Craik, K. (1943). The Nature of Explanation. London, New York, Cambridge University Press.

Cronin, M., A., G. Cleotide, et al. (2007) "Why Don't Well-Educated Adults Understand Accumulation? A Challenge to Researchers, Educators, and Citizens " **Volume**, DOI:

Cronin, M., A. and C. Gonzalez (2007). "Understanding the building blocks of dynamic systems." System Dyamics Review **23**(1): 1-17.

Croson, R. and K. Donohue (2006). "Behavioral Causes of the Bullwhip Effect and the Observed Value of Inventory Information." Management Science **52**(3): 323-336.

Dangerfield, B. C. and C. A. Roberts (1995). "Projecting Dynamic Behavior in the Absence of a Model: An Experiment." System Dynamics Review **11**(2): 157-172.

Davidsen, P. I. (2000). "Issues in the Design and Use of System Dynamics-Based Imteractive Learning Envrionments." Simulation & Gaming **31**(2): 170-177.

Diehl, E. and J. D. Sterman (1995). "Effects of Feedback Complexity on Dynamic Decision Making." Organizational Behavior and Human Decision Processes **62**(2): 198-215.

Dorner, D. (1980). "On the Difficulties People have in Dealing with Complexity." Simulation & Games **11**(1): 87-106.

Dörner, D. (1996). The Logic of Failure. New York, Metropolitan Books/Henry Holt:.

Doyle, J. K. (1997). "The Cognitive Psychology of Systems Thinking." System Dynamics Review **13**(3): 253-265.

Doyle, J. K. and D. N. Ford (1998). "Mental Models Concepts for System Dynamics Research." System Dynamics Review **14**(1): 3-29.

Doyle, J. K. and D. N. Ford (1999). "Mental Models Conepts Revisited : Some Clarifications and a Reply to Lane." System Dynamics Review **15**(4): 411-415.

Doyle, J. K., M. J. Radzicki, et al. (1996). Measuring the Effect of Systems Thinking Interventions on Mental Models. Conference of the International System Dynamics Society, Boston, MA, International System Dynamics Society.

Doyle, J. K., M. J. Radzicki, et al. (1998). Measuring Changes in Mental Models fo Dynamic Systems: An Exploratory Study. 16th International Conference of the System Dynamics Society, Quebec '98, Quebec City, Canada, System Dynamics Society.

Dudley, R. G. (2004). "Modeling the Effects of a Log Export Ban in Indonesia." System Dynamics Review **20**(2): 99-116.

Espejo, R. (1994). "What is Systemic Thinking?" System Dynamics Review **10**(2/3): 199-212.

Faust, L. J., R. Jackson, et al. (2004). "Models for Management of Wildlife Populations : Lessons from Spectacled Bears in Zoos and Grizzly Bears in Yellowstone." System Dynamics Review **20**(2): 163-178.

Fiddaman, T. S. (2002). "Exploring Policy Options with a Behavioral Climate - Economy Model." System Dynamics Review **18**(2): 243-267.

Fischhoff, B., P. Slovic, et al. (1977). "Knowing with Certainty: The Appropriateness of Extreme Confidence." Journal of Experimental Psychology **3**(4): 552-564.

Fisher, D., M. (2003). Student performance on the bathtub and cash flow problems. Proceedings of the 21st International Conference of the System Dynamics Society, New York, System Dynamics Society: Albany, NY.

Ford, A. (1990). "Estimating the Impact of Efficiency Standards on the Uncertainty of the

Northwest Electric System." Operations Research **38**(4): 580-597.

Ford, D. N. (1999). "A Behavioral Approach to Feedback Loop Dominance Analysis." System Dynamics Review **15**(1): 3-36.

Forrester, J. W. (1961). Industrial Dynamics. Cambridge MA, Productivity Press.

Forrester, J. W. (1969). Urban dynamics. Boston, MIT Press.

Forrester, J. W. (1971). "Counterintuitive Behavior of Social Systems." Technology Review **73**(3): 52-68.

Forrester, J. W. (1991). System Dynamics and the Lessons of 35 Years. A Systems-based Approach to Policymaking. K. De Greene, B. . Boston Kluwer Academic Publishers

Forrester, J. W. (1995). System Dynamics and Learner-Centered Learning in K-12 Education.

Forrester, J. W. and P. M. Senge (1980). Tests for Building Confidence in System Dynamics Models. System Dynamics. A. A. Legasto, Jr. and e. al. New York, North-Holland. **14:** 209-228.

Friedman, S. (2006). "Is counter-productive policy creating serious consequences? The case of highway maintenance." System Dyamics Review **22**(4): 371 - 394.

Gary, M. S. and R. E. Wood (2005). Mental Models, Decision Making and Performance in Complex Tasks. International System Dynamics Conference, Boston, System Dynamics Society.

Georgantzas, N. C. (1990). Cognitive Biases, Modelling and Performance: An Experimental Analysis. International System Dynamics Conference, System Dynamics Society.

Georgantzas, N. C. (2003). "Tourism Dynamics : Cyprus' Hotel Value Chain and Profitability." System Dynamics Review **19**(3): 175-212.

Gonçalves, P. M., J. H. Hines, et al. (2005). "The Impact of Endogenous Demand on Push-Pull Production Systems." System Dynamics Review **21**(3): 187-216.

Goodwin, P. (2002). "Integrating management judgment and statistical methods to improve short-term forecasts." Omega **30**(2): 127-135.

Goodwin, P. (2005). "How to integrate judgment with statistical forecasts." The International Journal of Applied Forecasting **1**: 8-12.

Graham, A. K. and C. A. Ariza (2003). "Dynamic, Hard and Strategic Questions : Using Optimization to Answer a Marketing Resource Allocation Question." System Dynamics Review **19**(1): 27-46.

Grösser, S. (2005). Does Experience or an Education in System Dynamics Help People to Solve Simple, Dynamic Problems? - A Laboratory Experiment. Proceedings of the 23rd International Conference of the System Dynamics Society, Boston, The System Dynamics Society.

Hall, R., I. (1976). "A System Pathology of an Organization: The Rise and Fall of the Old Saturday Evening Post." Administrative Science Quarterly **21**(2): 185-211.

Hecht, G (2004). "Systems Thinking Skills, Mental Representations and Identification of Unintended Consequences." PhD Thesis, University of Illinois at Urbana-Champaign

Heider, F. (1958). The psychology of interpersonal relations. New York, Wiley.

Heinbokel, J. and J. Potash (2003). Bathtub Dynamics at Vermont Commons School. Proceedings of the 21st International Conference of the System Dynamics Society, New York, System Dynamics Society: Albany, NY.

Heller, F., A., P. J. D. Drenth, et al. (1977). "A Longitudinal Study in Participative Decision-Making " Human Relations **30**(7): 567-587

Hilmola, O.-P., P. Helo, et al. (2003). "The Value of Product Development Lead Time in Software Startup." System Dynamics Review **19**(1): 75-82.

Homer, J. and R. Oliva (2001). "Maps and Models in System Dynamics: a Response to Coyle." System Dynamics Review **17**(4): 347-355.

Homer, J. B., G. B. Hirsch, et al. (2004). "Models for Collaboration : How System Dynamics Helped a Community Organize Cost-Effective Care for Chronic Illness." System Dynamics Review **20**(3): 199-222.

Howick, S., F. Ackermann, et al. (2006). "Linking event thinking with structural thinking: methods to improve client value in projects." System Dynamics Review **22**(2): 113-140.

Hsiao, N. and G. P. Richardson (1999). In Search of Theories of Dynamic Decision Making: A Literature Review Conference of the International System Dynamics Society, Wellington, New Zealand, International System Dynamics Society.

Huz, S., D. F. Andersen, et al. (1997). "A Framework for Evaluating Systems Thinking Interventions: An Experimental Approach to Mental Health System Change." System Dynamics Review **13**(2): 149-169.

Ison, R. L. (2008). Systems thinking and Practice for Action Research. Handbook of Action Research (2nd ed.). P. Reason and H. Bradbury. London, Sage Publications.

Ison, R. L. and D. B. Russell (2000). Agricultural Extension and Rural Development: Breaking Out of Traditions. Cambridge, Cambridge University Press.

Jackson, M., C. (1991). "The origins and nature of critical systems thinking " Systemic Practice **4**(2): 131-149.

Jensen, E. and B. Brehmer (2003). "Understanding and Control of a Simple Dynamic System." System Dynamics Review **19**(2): 119-137.

Jones, A. P., D. A. Seville, et al. (2002). "Resource Sustainability in Commodity Systems : The Sawmill Industry in the Northern Forest." System Dynamics Review **18**(2): 171-204.

Kainz, D. and G. Ossimitz (2002). Can students learn stock-flow-thinking? An empirical investigation. International System Dynamics Conference, Palermo, Italy, System Dynamics Society.

Kauffman, D., L. (1980). Systems One: An Introduction to Systems Thinking. Minneapolis, Future Systems, Inc.

Kelley, H. H. (1967). Attribution theory in social psychology. Nebraska Symposium on Motivation D. Levine. Lincoln, University of Nebraska Press. **15:** 192-240.

Kelly, T. J. (1970). The Dynamics of R&D Project Management, M. I. T.

Keloharju, R. and E. F. Wolstenholme (1989). "A Case Study in System Dynamics Optimization." J. Operations Research Soc. **40**(March): 221-230.

Kirkwood, C., W. (1998). System Dynamics Methods : A Quick Introduction. Arizona Arizona State University.

Kleinmutz, D. N. (1993). "Information Processing and Misperceptions of the Implications of Feedback in Dynamic Decision Making." System Dynamics Review **9**(3): 223-237.

Kubanek, G. (2003). Bathtub Dynamics—Ottawa. Proceedings of the 21st International Conference of the System Dynamics Society, New York, System Dynamics Society: Albany, NY.

Lane, D. C. (2000). "Viewpoint; Diagramming Conventions in System Dynamics." the Journal of the Operational Research Society **51**(2): 241.

Lane, D. C. and R. Oliva (1994). The Greater Whole : Towards a Synthesis of System Dynamics and Soft Systems Methodology. City University Business School Working Paper.

Lane, D. C. and R. Oliva (1998). "The greater whole: Towards a synthesis of system dynamics and soft systems methodology." European Journal of Operational Research **107**(1): 214-235.

Lane, D. C. and R. Oliva (1998). "The greater whole: Towards a synthesis of system dynamics and soft systems methodology." European Journal of Operational Research **107**(1): 214-235.

Lawrence, M., P. Goodwin, et al. (2006). "Judgmental forecasting: A review of progress over the last 25 years." International Journal of Forecasting **22**(3): 493-518.

Lee, H., L. , V. Padmanabhan, et al. (1997b). "The Bullwhip Effect in Supply Chains." Sloan Management Review **Spring**: 93-102.

Lee, H. L., V. Padmanabhan, et al. (1997a). "Information Distonion in a Supply Chain: The Buliwhip Effect." Management Science **43**(4): 546-558.

Lee, M.-H., N. H. Choi, et al. (2005). "A Systems Thinking Approach to the New Administrative

Capital in Korea : Balanced Development or Not?" System Dynamics Review **21**(1): 69-85.

Lichtenstein, S. and B. Fischhoff (1977). "Do those who know more also know more about how much they know?: The calibration of probability judgments." Organizational Behavior and Human Performance **20**: 159-183.

Liddell, W. G. and J. H. Powell (2004). "Agreeing Access Policy in a General Medical Practice : A Case Study Using QPID." System Dynamics Review **20**(1): 49-73.

Lim, J. S. and M. O'Connor (1996). "Judgmental forecasting with time series and causal information." International Journal of Forecasting **12**(1): 139-153.

Luna-Reyes, L. F. and D. L. Andersen (2003). "Collecting and Analyzing Qualitative Data for System Dynamics : Methods and Models." System Dynamics Review **19**(4): 271-296.

Lyneis, J., M., K. G. Cooper, et al. (2001). "Strategic Management of Complex Projects: A Case Study using System Dynamics." System Dynamics Review **17**(3): 237-260.

Lyneis, J. and D. Lyneis (2003). Bathtub dynamics at WPI. Proceedings of the 21st International Conference of the System Dynamics Society, New York, System Dynamics Society: Albany, NY.

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

Maani, K. and V. Maharaj (2004). "Links Between Systems Thinking and Complex Decision Making." System Dynamics Review **20**(1): 21-48.

Maani, K. E. and R. Y. Cavana (2000). Systems Thinking and Modelling : Understanding Change and Complexity. New Zealand, Pearson Education.

Maani, K. E. and V. Maharaj (2004). "Links Between Systems Thinking and Complex Decision Making." System Dynamics Review **20**(1): 21-48.

Mackay, J., S. Barr, et al. (1992). "An Empirical Investigation of the Effects of Decision Aids on Problem-Solving Processes." Decision Sciences **23** (3): 648–672.

Maier, F. H. and A. Größler (2000). "What are we Talking About? -- A Taxonomy of Computer Simulations to Support Learning." System Dynamics Review **16**(2): 135-148.

Makridakis, S. and S. C. Wheelwright (1982). "Handbook of Forecasting: A Manager's Guide."

Martínez-Fernández, J. and M. A. E. Selma (2004). "The Dynamics of Water Scarcity on Irrigated Landscapes : Mazarron and Aguilas in South-Eastern Spain." System Dynamics Review **20**(2): 117-137.

Maxwell, T., A. (1995). Decisions: *Cognitive Styles, Mental Models, and Task Performance.* Albany, State University of New York at Albany. **Ph.D.**

McIntyre, B. (2005). "Further Insights into the "Perspectives on Rationality in System Dynamics" Research Problem." System Dynamics Review **21**(1): 87-88.

Meadows, D. L. (2002). "Dana Meadows : Asking Hard Questions, Speaking Simple Truths." System Dynamics Review **18**(2): 111-119.

Milling, P. M. (2002). "Understanding and Managing Innovation Processes." System Dynamics Review **18**(1): 73-86.

Mingers, J. and S. Taylor (1992). "The Use of Soft Systems Methodology in Practice." The Journal of the Operational Research Society **43**(4): 321-332.

Mojahedzadeh, M., D. F. Andersen, et al. (2003). "Using *Digest* to Implement the Pathway Participation Method for Detecting Influential System Structure." System Dynamics Review **20**(1): 1-20.

Mojahedzadeh, M., D. F. Andersen, et al. (2004). "Using *Digest* to Implement the Pathway Participation Method for Detecting Influential System Structure." System Dynamics Review **20**(1): 1-20.

Moore, D. S. and G. P. McCabe (1993). Introduction to the Practice of Statistics. New York, W.H. Freeman.

Morecroft, J. D. W. (1985). "The Feedback View of Business Policy and Strategy." System Dyamics Review **1**(1): 4-19.

Moxnes, E. (1998). "Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics." Management Science **44**(9): 1234-1248.

Moxnes, E. (1998). "Overexploitation of renewable resources: The role of misperceptions." Journal of Economic Behavior and Organization **37**(1): 107-127.

Moxnes, E. (2004). "Misperceptions of Basic Dynamics : The Case of Renewable Resource Management." System Dynamics Review **20**(2): 139-162.

Moxnes, E. (2005). "Policy Sensitivity Analysis : Simple Versus Complex Fishery Models." System Dynamics Review **21**(2): 123-145.

Moyaux, T., B. Chaib-draa, et al. (2007). "Information Sharing as a Coordination Mechanism for Reducing the Bullwhip Effect in a Supply Chain." IEEE Transactions on Systems, Man, and Cybernetics **37**(3): 396-409.

Oliva, R. and J. D. Sterman (2001). "Cutting Corners and Working Overtime: Quality Erosion in the Service Industry." Management Science **47**(7): 894-914.

Oliva, R., J. D. Sterman, et al. (2003). "Limits to Growth in the New Economy : Exploring the 'Get Big Fast' Strategy in E-Commerce." System Dynamics Review **19**(2): 83-117.

Ossimitz, G. (2002). Stock–flow-thinking and reading stock–flow-related graphs: an empirical investigation in dynamic thinking abilities. Proceedings of the 20th International Conference of the System Dynamics Society, Palermo, Italy, System Dynamics Society: Albany, NY.

Otto, P. A. and J. Struben (2004). "Gloucester Fishery : Insights from a Group Modeling Intervention." System Dynamics Review **20**(4): 287-312.

Paich, M. and J. D. Sterman (1993). "Boom, Bust, and Failures to Learn in Experimental Markets." Management Science **39**(12).

Pala, Ö. and J. A. M. Vennix (2005). "Effect of System Dynamics Education on Systems Thinking Inventory Task Performance." System Dynamics Review **21**(2): 147-172.

Park, M. and F. Pena-Mora (2003). "Dynamic Change Management for Construction :

Introducing the Change Cycle into Model-Based Project Management." System Dynamics Review **19**(3): 213-242.

Pavlov, O. V. and K. Saeed (2004). "A Resource-Based Analysis of Peer-to-Peer Technology." System Dynamics Review **20**(3): 237-262.

Pejic-Bach, M. (2003). "Surviving in an Environment of Financial Indiscipline : A Case Study from a Transition Country." System Dynamics Review **19**(1): 47-74.

Peterson, D. and R. Eberlein (1994). "Reality Check: A bridge between systems thinking and system dynamics." System Dynamics Review **10**(2-3): 159-174.

Peterson, D. W. and R. L. Eberlein (1994). "Reality Checks : A Bridge Between Systems Thinking and System Dynamics." System Dynamics Review **10**(2/3): 159-174.

Piattelli, M. L., M. A. Cuneo, et al. (2002). "The Control of Goods Transportation Growth by Modal Share Re-Planning : the Role of a Carbon Tax." System Dynamics Review **18**(1): 47-69.

Quaden, R. and A. Ticotxky (2003). Bathtub dynamics at Carlisle public schools. Proceedings of the 21st International Conference of the System Dynamics Society, New York, System Dynamics Society: Albany, NY.

Qudrat-Ullah, H. (2005). Improving Dynamic Decision Making through Debriefing: An Empirical Study. Proceedings of the 23rd International Conference of the System Dynamics Society, Boston, The System Dynamics Society.

Rajadhyaksha, G. and A. Dwivedi (2005). "Forecasting Practice: Decision Support System to Assist Judgmental Forecasting." Computing in Economics and Finance **203**.

Reichelt, K. and J. Lyneis (1999). "The dynamics of project performance: benchmarking the drivers of cost and schedule overrun." European Management Journal **17**(2): 135-150.

Remus, W. (1996). "Will Behavioral Research on Managerial Decision Making Generalize to Managers?" Managerial and Decision Economics **17**(1): 93-101.

Repenning, N., P. (2000). "A Dynamic Model of Resource Allocation in Multi-Project Research and Development Systems." System Dynamics Review **16**(3): 173-212.

Repenning, N. P. (2003). "Selling System Dynamics to (Other) Social Scientists." System Dynamics Review **19**(4): 303-327.

Richardson, G. P. (1986/1976). "Problems with Causal Loop Diagrams." System Dynamics Review **2**(2): 158-170.

Richardson, G. P. (1991). System Dynamics: Simulation for Policy Analysis from a Feedback Perspective. Qualitative Simulation Modeling and Analysis. P. A. Fishwick and P. A. Luker. New York, Springer Verlag.

Richardson, G. P. (1996). "Problems for the Future of System Dynamics." System Dynamics Review **12**(2).

Richardson, G. P. (1997). President's address to the System Dynamics Society. 15th International System Dynamics Conference, Istanbul Turkey, System Dynamics Society.

Richardson, G. P. (1997). "Problems in Causal Loop Diagrams Revisited." System Dynamics Review **13**(3): 247-252.

Richardson, G. P. (1999). "Reflections for the Future of System Dynamics." The Journal of the Operational Research Society **50**: 440-449.

Richardson, G. P. (2001). Mapping versus Modeling : THE Answer to the Debate. The 19th International Conference of the System Dynamics Society, Atlanta, Georgia, System Dynamics Society.

Richardson, G. P. and D. F. Andersen (1995). "Teamwork in Group Model Building." System Dynamics Review **11**(2): 113-137.

Richardson, G. P. and A. L. Pugh, III (1981). Introduction to System Dynamics Modeling with DYNAMO. Cambridge MA, Productivity Press.

Richardson, G. P., E. Wolstenholme, et al., Eds. (1994). Systems Thinkers, Systems Thinking. System Dynamics Review, System Dynamics Society.

Richmond, B. (1993). "Systems thinking: critical thinking skills for the 1990s and beyond "

System Dynamics Review 9(2): 113-133.

Richmond, B. (1997). ithink Version 5 User Manual. Hanover, New Hampshire, USA, High Performance Systems.

Roberts, E. B. (1964). "New Directions in Industrial Dynamics." Industrial Management Review **VI**(1): 5-14.

Roberts, N. H., D. F. Andersen, et al. (1981). Introduction to Computer Simulation. Waltham, MA, Pegasus Communications.

Roberts, N. H., D. F. Andersen, et al. (1983). Introduction to Computer Simulation: The System Dynamics Modeling Approach. Reading, MA, Addison-Wesley.

Rodrigues, A. G. and J. Bowers (1996). "System Dynamics in Project Management: A Comparative Analysis with Traditional Methods." System Dynamics Review **12**(2): 121-139.

Rodrigues, A. G. and T. M. Williams (1995). The Application of System Dynamics to Project Management: An Integrated Model with the Traditional Procedures, University of Strathclyde.

Roethlisberger, F., J. and W. Dickson, J. (1939). Management and the Worker. Cambridge, MA, Harvard University Press.

Rosenthal, R. (1966). Experimental Effects in Behavioral Research. New York, Appleton-Century-Crofts.

Rouwette, E. A. J. A., A. Größler, et al. (2004). "Exploring influencing factors on rationality: a literature review of dynamic decision-making studies in system dynamics." Systems Research and Behavioral Science **21**(4): 351.

Rouwette, E. A. J. A., J. A. M. Vennix, et al. (2002). "Group Model Building Effectiveness : A Review of Assessment Studies." System Dynamics Review **18**(1): 5-45.

S. Makridakis, A. (1982). "The accuracy of extrapolation (time series) methods: Results of a forecasting competition." Journal of Forecasting **1**(2): 111-153.

Saysel, A. K. and Y. Barlas (2006). "Model simplification and validation with indirect structure validity tests." System Dynamics Review **22**(3): 241-262.

Schmidt, M. J. and M. S. Gary (2002). "Combining System Dynamics and Conjoint Analysis for Strategic Decision Making with an Automotive High-Tech SME." System Dynamics Review **18**(3): 359-379.

Scholl, G. J. (1995). "Benchmarking the System Dynamics Community: Research Results." System Dynamics Review **11**(2): 139-155.

Senge, P. M. (1990). The Fifth Discipline: The Art and Practice of the Learning Organization, Doubleday/Currency.

Senge, P. M. (1996). The Fifth Discipline and the Infrastructures of a Learning Organization. Waltham, MA, Pegasus Communications.

Sharda, R., S. Barr, et al. (1988). "Decision Support System Effectiveness: A Review and an Empirical Test." Management Science **34**(2): 139-159.

Siders, A., A. Kaufman, et al. (2006). "Do Practice Effects on Wechsler's Performance Subtests Relate to Children's General Ability, Memory, Learning Ability, or Attention?" Applied Neuropsychology **13**(4): 242–250.

Stake, R., E. (1967). "The Countenance of Educational Evaluation." Teachers College Record **68**: 523–540.

Stata, R. (1989). "Organizational Learning - the Key to Management Innovation." Sloan Management Review **30**(Spring): 63-73.

Stave, K. A. (2002). "Using System Dynamics to Improve Public Participation in Environmental Decisions." System Dynamics Review **18**(2): 139-167.

Stephens, C. A., A. K. Graham, et al. (2005). "System Dynamics Modelling in the Legal Arena : Meeting the Challenges of Expert Witness Admissibility." System Dynamics Review **21**(2): 95-122.

Sterman, J. D. (1989). "Misperceptions of Feedback in Dynamic Decision Making "

Organizational Behaviour and Human Decision Processes **43**(3): 301-335.

Sterman, J. D. (1989). "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment " Management Science **35**(3): 321-339.

Sterman, J. D. (1992a) "System Dynamics Modelling for Project Management." **Volume**, DOI:

Sterman, J. D. (1992b). "Teaching Takes Off " OR/MS Today **35**(3): 40-44.

Sterman, J. D. (1994). "Learning in and about complex systems." System Dynamics Review **10**(2-3): 291-330.

Sterman, J. D. (2000). Business Dynamics: Systems Thinking and Modeling for a Complex World, Irwin McGraw-hill.

Sterman, J. D. (2001). "System Dynamics Modeling: Tools for Learning in a Complex World." California Management Review **43**(4).

Sterman, J. D. (2002). "Dana Meadows : Thinking Globally, Acting Locally." System Dynamics Review **18**(2): 101-107.

Sterman, J. D. (2002a). "All Models are Wrong : Reflections on Becoming a Systems Scientist." System Dynamics Review **18**(4): 501-531.

Sterman, J. D. and L. B. Sweeney (2002b). "Cloudy Skies : Assessing Public Understanding of Global Warming." System Dynamics Review **18**(2): 207-240.

Strake, R., E. (1967). "The Countenance of Educational Evaluation." Teachers College Record **68**(7): 523–540.

Strohhecker, J. (2005). "Scenarios and Simulations for Planning Dresdner Bank's E-Day." System Dynamics Review **21**(1): 5-32.

Sweeney, L. B. and J. D. Sterman (2000). "Bathtub dynamics: initial results of a systems thinking inventory." System Dynamics Review **16**(4): 249-286.

Sweeney, L. B. and J. D. Sterman (2007). "Understanding public complacency about climate change: adults' mental models of climate change violate conservation of matter " Climatic Change **80**(3-4): 213-238.

SystemDynamicsSociety. (2006). "What is System Dynamics." Retrieved 1 January, 2006, from http://systemdynamics.org/.

Taylor, T. and D. N. Ford (2006). "Tipping point failure and robustness in single development projects." System Dynamics Review **22**(1): 51-71.

Thun, J.-H. (2006). "Maintaining preventive maintenance and maintenance prevention: analysing the dynamic implications of Total Productive Maintenance." System Dynamics Review **22**(2): 163-179.

Tignor, W. W. (2007). System Dynamics: Publication Metrics (2000-2005). Proceedings of the 25th International Conference of the System Dynamics Society, Boston, The System Dynamics Society.

Trees, S., J. Doyle, et al. (1996). Using Cognitive Styles Typology to Explain Dynamic Decision Making in a Computer Simulated Game Environment. International System Dynamics Conference, Cambridge, Massachusetts, USA, System Dynamics Society.

Trost, C. S. (2002). "A Dynamic Model of Work Quality in a Government Oversight Organization." System Dynamics Review **18**(4): 473-495.

Vennix, J. A. M. (1996). Group Model Building : Facilitating Team Learning Using System Dynamics. Chichester, Wiley.

Vennix, J. A. M. (1999). "Group Model-Building : Tackling Messy Problems." System Dynamics Review **15**(4): 379-401.

Warren, K. (2004). "Why Has Feedback Systems Thinking Struggled to Influence Strategy and Policy Formulation? Suggestive Evidence, Explanations and Solutions." Systems Research and Behavioral Science **21**(4): 331-347.

Webby, R. and M. O'Connor (1994). "The effectiveness of decision support systems: the implications of task complexity and DSS sophistication." Journal of information technology **9**(1): 19-28.

Weber, M. and M. Schwaninger (2002). "Transforming an Agricultural Trade Organization : A

System-Dynamics-Based Intervention." System Dynamics Review **18**(3): 381-401.

Winch, G. W. (2000). System Dynamics: From Theory to Practice. 1st International Conference on Systems Thinking in Management, Geelong, Australia.

Wolstenholme, E. (1982). "System Dynamics in Perspective." The Journal of the Operational Research Society **33**: 547-556.

Wolstenholme, E. (1999). "Qualitative vs quantitative modelling: The evolving balance." The Journal of the Operational Research Society **50**(4): 422-428.

Wolstenholme, E. F. (1983). "System Dynamics: A System Methodology or a System Modeling Technique." Dynamica **9**(2): 84-90.

Wolstenholme, E. F. (1990). System Enquiry - A System Dynamics Approach, John Wiley.

Wolstenholme, E. F. (2003). "Towards the Definition and Use of a Core Set of Archetypal Structures in System Dynamics." System Dynamics Review **19**(1): 7-26.

Wolstenholme, E. F. (2004). "Using Generic System Archetypes." System Dynamics Review **20**(4): 341-356.

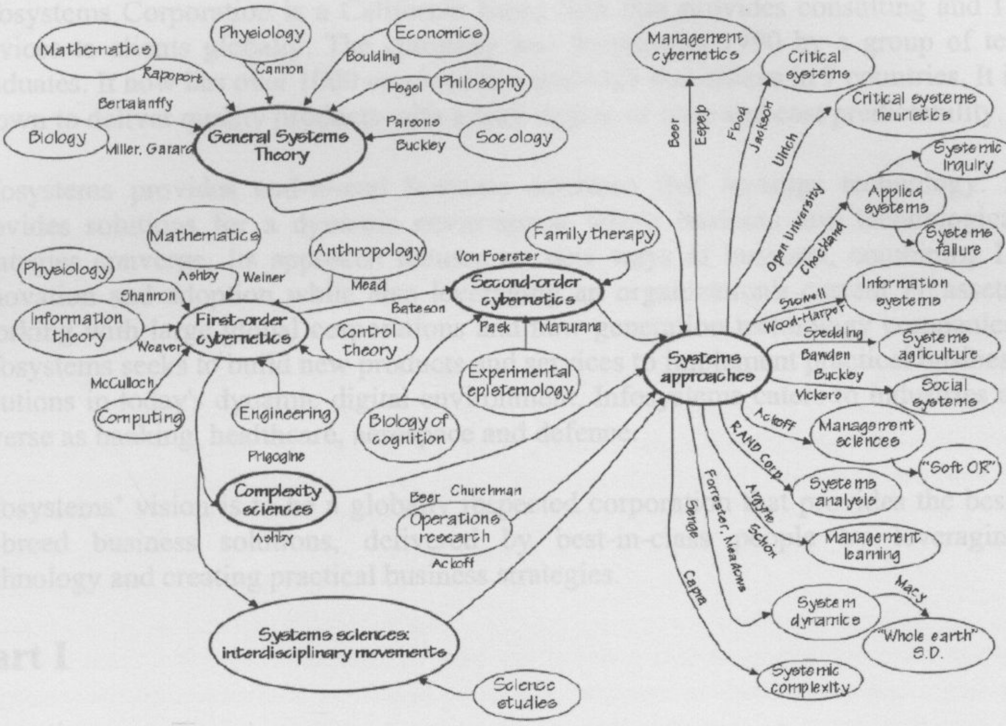
Wolstenholme, E. F. and R. G. Coyle (1983). "The Development of System Dynamics as a Methodology for System Description and Qualitative Analysis." Journal of the Operational Research Society (UK) **34**(7): 569-581.

Wood, R. E. (1986). "Task Complexity: Defintion of the Construct." Organizational Behavior and Human Decision Processes **37**: 60-82.

Zaraza, R. (2003). Bathtub dynamics in Portland at SyMFEST. Proceedings of the 21st International Conference of the System Dynamics Society, New York, System Dynamics Society: Albany, NY.

Appendix

2A: Influences that have shaped contemporary systems approaches (Ison 2008)



3A: The following is the case study that was provided to the participants for the experiment reported in Chapter 3.

Background

Infosystems Corporation is a California based firm that provides consulting and IT services to clients globally. The company was founded in 1990 by a group of ten graduates. It now has over 1000 employees worldwide and offices in 7 countries. It is known to deliver quality products with a high degree of time and cost predictability.

Infosystems provides end-to-end business solutions that leverage technology. It provides solutions for a dynamic environment where business and technological strategies converge. Its approach focuses on new ways of business, combining IT innovation and adoption while also leveraging an organization's current IT assets. Working with large global corporations and new generation technology companies, Infosystems seeks to build new products and services to implement practical business solutions in today's dynamic digital environment. Infosystems caters to industries as diverse as banking, healthcare, aerospace and defence.

Infosystems' vision is to be a globally respected corporation that provides the best-of-breed business solutions, delivered by best-in-class people by leveraging technology and creating practical business strategies.

Part I

Operations

Infosystems gets most of their clients through referrals. A contract serves as a mutually binding agreement that obligates Infosystems to provide the specified products/services and obligates the client to pay for them. The time frame, cost and quality requirements are the highlights of the deal. Late projects cause huge penalties and this is the major cause of worry most of the time.

As in any other organisation, completing projects on time, in the stipulated budget and conforming to high-quality standards is of prime importance to the management. Projects not completed on time incur huge penalties and adversely affect profits. For simplicity it can be thought that the amount of revenue generated is proportional to the projects completed on time. Hence there is enormous emphasis on completing all projects in time.

Problem Statement

The IT products and services market has been growing at a rapid rate since the past ten years. Surprisingly, Infosystems' revenue has not followed a similar trend. Revenues rose continuously between the years 1990 and1994. The number of new contracts kept pouring in. At this time Infosystems decided to increase its capacity by recruiting new employees in bulk. The company even decided to open two new offices. All was going well for Infosystems when all of a sudden the numbers of new contracts started decreasing and Infosystems' revenue plummeted during 1994-95. They hardly had any projects on hand! The projects they had already had not completed successfully. Either they were late and incurred penalties or they did not conform to quality standards. Revenues fell lower than before. Top management and the team of strategists at Infosystems devised a scheme. Fortunately, it worked. It took a year for Infosystems to get back their original market share. However, a similar fall occurred in the year 2000 as well. Infosystems was not sure what caused the downfall. And this time it was worse than the previous one. Top management employed the same scheme as before. It worked again.

The plan was simple. Top management related project failure to employee performance. They identified all projects that were running late and would potentially incur losses. Project managers of these projects were invited to a meeting where the scheme was announced. According to the scheme, project managers who could put an extra effort into the project and complete it in time would get a salary increment. The scheme also implied that those who were unsuccessful might even lose their job. Most project managers made their team work day and night and completed the projects in time. At the HR level, top management reacted to these potential losses by cancelling fresh recruitments. Additionally, there was mass retrenchment of employees that Infosystems thought were incompetent. The scheme worked. Revenues got better. These job cuts seemed justified when the projects in hand were less. However, the problem always returned after a while (see figure 1). A number of projects got delayed subsequently, which lead to losses. This adversely affected Infosystems' reputation.

It is now mid- 2005. The numbers of projects running late is increasing once again. Furthermore Infosystems is unable to grab new projects as well. The trend seems to be repeating – and is worse than ever. The overall demand for consulting and IT services continues to rise and other firms are enjoying growth in their revenue. Top management is getting panicky. They are not sure what's happening. They have run out of solutions. Should they apply the same strategy as before or try something else?

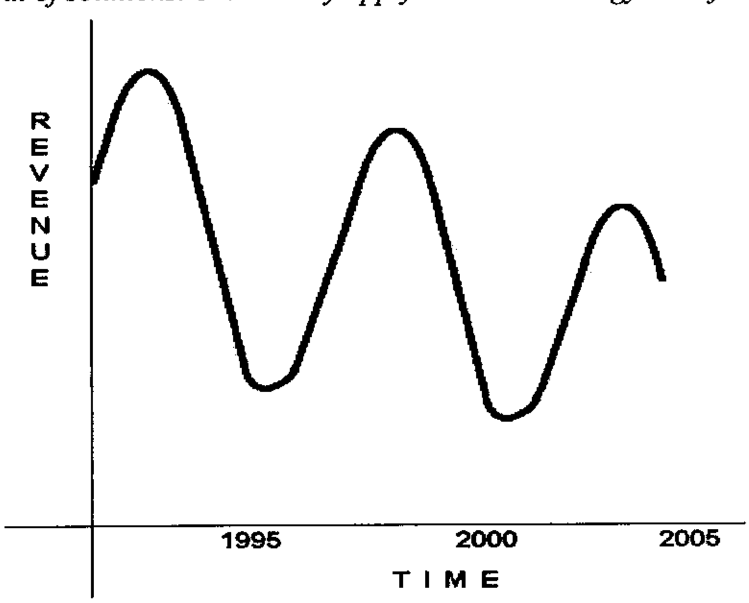


Figure1: Infosystems' revenue during the last ten years

Your Role

Infosystems' board of directors are perplexed by the situation. Internal strategists are unable to explain the cause of this phenomenon. In their report they explain that each department was separately examined for lapses. However each was doing the best they could. The strategists could not explain the periodic fall in revenues in their report. Given the situation, the board has decided to hire the services of an external consultant. You are a renowned business analyst. Some of the board members of Infosystems have heard your success stories. The board members are impressed by your record after two CEOs that you have worked for as an analyst provide suitable references. Infosystems offers you the job and you decide to take up the challenge.

Data Gathering

You begin your job by gathering information regarding strategies of the top management, policies and decisions that effect employees, market trend, history and other factors that you feel may have been a cause of the issue.

You personally speak to relevant employees to get their feedback. The following summarises employee perceptions after the initial phase of collecting information. This is what they had to say:

The freshly hired employees were happy during the mentoring phase. However at times they felt that there was too little time to get assimilated into the project and start working on it. The knowledge that they gained at university was very different than what they had to work on. There was undue pressure on them to complete the work assigned to them in the stipulated time. In addition, they had to produce work that conformed to high industry standards – a stipulation that was absent in the academic projects they undertook at university.

Experienced employees were by and large happy with the salaries and perks they received from Infosystems. Given their experience they could understand their job but complained that they were under pressure most of the time. Not completing the project tasks on time could spoil their reputation in the organisation and might be a hindrance in obtaining future promotions. Additional pressure mounted on them when the firm took up too many projects. Their continued requests for acquisition of new employees had fallen on deaf years. One of the team members said, "I remember that on the ABC project, the HR department could not provide us with fresh employees when we needed them, and then all of a sudden we got so many that we did not have time to train them". During final stages of many projects, employees had to work overtime and were stretched to the limit. They were offered incentives for the extra work.

The HR department is responsible for recruiting new employees from universities. Infosystems HR manager Andrew Baker explains that "at times there are periods when there are not many incoming projects". Many new recruits are retrenched during these "dry periods".

The strategy team is the "think tank" of Infosystems. It is responsible for making policies and decisions that drive the entire firm along with creating policies for recruitment, expansion, downsizing etc. The strategy team claims that its strategies and decisions have been instrumental in helping Infosystems maximise its profits in the past. Moreover, its basic policy remains constant: more the projects, more the revenue. All major decisions are based on this unwritten rule. Whenever the fall began, they cut down on all major resources to save money. Resources cuts have always been justified by the downward trend in revenue. However, the cause of cyclic troughs in the revenue was still an unsolved mystery.

Part II

After assessing the HR operations at Infosystems and having proposed a hypothesis of what might be going wrong, you now go into deeper detail to help them predict some important variables for an ongoing project. Your aim is to analyse how key variables behave over a period of one year as this might help you to understand the situation better. You are provided with the following information about the "workforce" and "staffing" factors:

For all practical purposes the total workforce working on a project is divided into "Experienced workforce" and "newly-hired workforce". According to the HR manager, these are the two significant levels (employee pools) for the project's dynamics. This division of workforce is done as the newly-hired project members pass through an orientation during which they are less than fully productive.

The project starts with a team of 10 experienced employees. There are no newly-hired employees at this stage. Hiring commences only when new tasks are discovered and the project manager requests for fresh employees. The number of experienced employees changes when new employees join this pool or when experienced employees leave the project.

Prospective employees undergo the recruitment process. All those who are hired are part of the newly hired workforce. The number of newly-hired employees changes when prospective employees join this pool or when newly-hired employees join the experienced pool after training. Since Infosystems likes to hire the best brains available, it has a long and intensive recruitment process. Hiring of new employees takes about 40 days; this is the hiring delay. The rate of hiring (persons hired per day) is defined as either 0 (when no hiring is required) or as "workforce gap"/ "hiring delay". Following the hiring process these employees are added to the newly-hired pool.

All newly-hired employees then undergo an "assimilation" phase while they are trained. The training has technical as well as social dimensions. The average assimilation time is of 80 days; this is the assimilation delay. The rate of assimilation is dependent upon the "newly-hired workforce" level. This rate can be defined as "newly-hired workforce"/ "assimilation delay". All assimilated employees are added to the experienced-workforce pool.

Experienced employees who are not productive are either asked to leave or are not promoted. Employees who are made stagnant generally find lucrative opportunities in other firms. The average duration that an employee spends on the project before quitting is 600 days; this is the average employment time. The rate at which employees quit is directly related to the number of employees in the experienced workforce level. The quit rate is defined as "Experienced Workforce"/ "Average Employment Time".

For the purpose of hiring new employees, the total number of employees working on the project needs to be known; this is known as "total workforce". It is the sum of the experienced workforce and the newly hired workforce. "Total workforce" is not an accumulation of employees as it is derived from two existing levels.

The "desired workforce" is determined by the project manager from time to time. For this project the project manager decides to increase the total workforce by 10 people after the first 80 days.

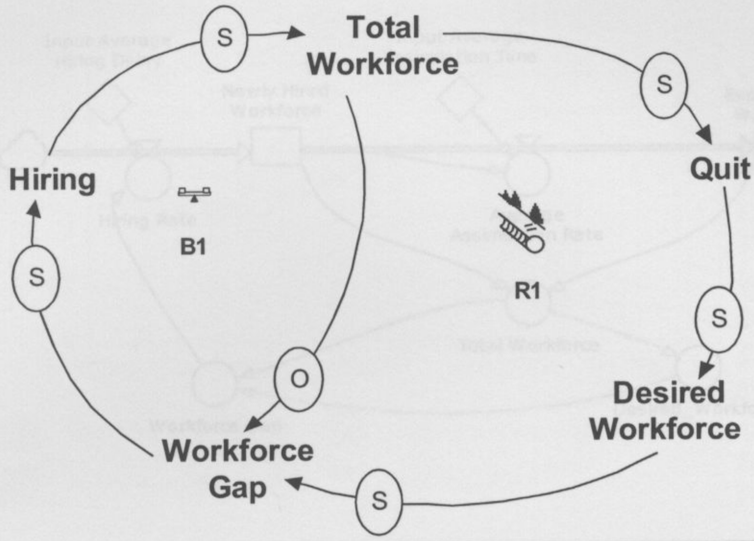
This difference between the "desired workforce" and the "total workforce" is known as the "workforce gap". This gap then influences the rate of hiring. As described before by the HR manager, the rate of hiring is determined by the workforce gap and the average hiring delay.

The project manager, from time to time, likes to know the fraction of experienced people on the team. This is known as the "experienced workforce fraction" and is derived by dividing the "experience workforce" by the "total workforce".

Assumptions

The HR department is not concerned with the people who are unsuccessful in the hiring phase. It is assumed that none of the fresh employees quit during the assimilation phase. The company is neither concerned about the past of a new employee nor about the future of an employee who has quit the job.

3B: Causal Loop Diagram



Explanation of Variables

Total Workforce
Total number of persons employed - experienced and inexperienced.
'Experienced Workforce'+'Newly Hired Workforce'
Original unit: "persons"

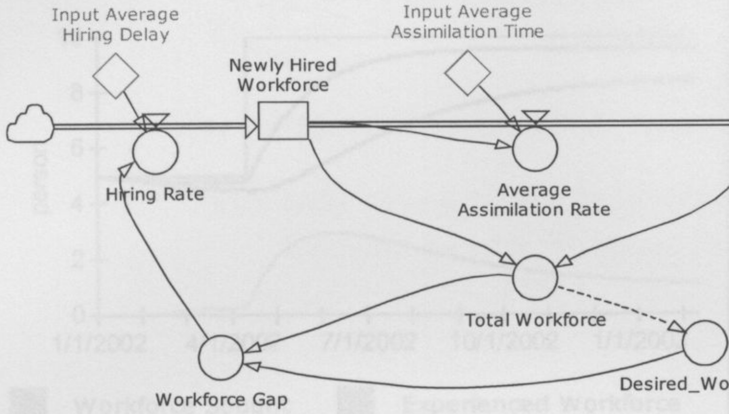
Workforce Gap
The shortfall or excess number of employees, or the difference between the number of employees required and that already employed.
'Workforce Sought'-'Total Workforce'
Original unit: "persons"

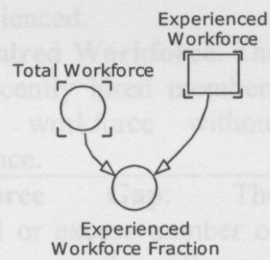
Quit Rate
The number of occasions within a given period that employees are quitting.
'Experienced Workforce'/'Input Average Employment Time'
Original unit: "persons/day"

Hiring
Hiring Rate - the rate at which new employees are hired, determined by the workforce gap.
MAX(0<<person/day>>,'Workforce Gap'/'Input Average Hiring Delay')
Original unit: "persons/day"

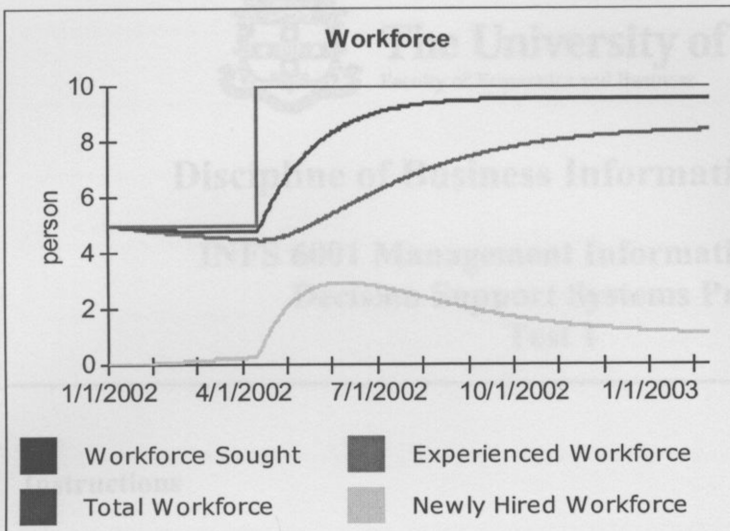
Desired Workforce

3C: Stock and Flow Diagram for Workforce





3D: Simulated Results

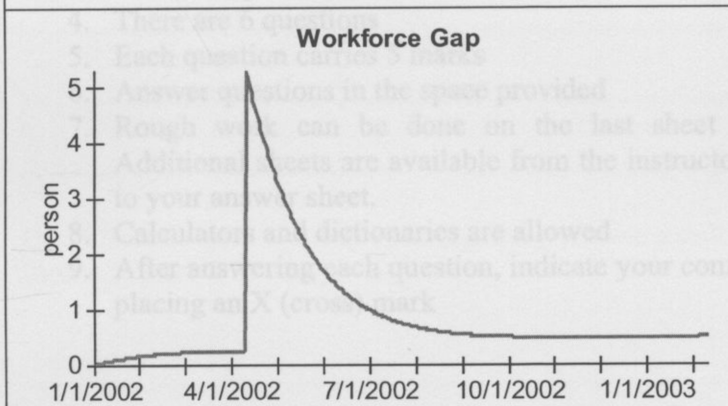


Workforce Sought: The equals either the target workforce or the total workforce ceiling, depending on which of the two is in shortfall.

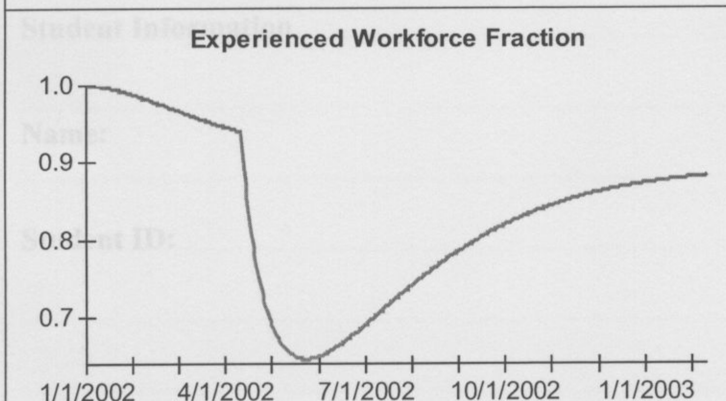
Experienced Workforce: The number of employees with capable of working on their own.

Total workforce: Total number of persons employed - experienced and unexperienced.

Newly hired Workforce: The most recently hired members of the workforce without experience.



Workforce Gap: The shortfall or excess number of employees, or the difference between the number of employees required and that already employed.



Experienced Workforce Fraction: The part of total workforce capable of working on its own.

3E: Test 1

The University of Sydney

Faculty of Economics and Business

Discipline of Business Information Systems

INFS 6001 Management Information Systems
Decision Support Systems Project
Test 1

Instructions

1. This is an individual test
2. It is assumed that you have read the case
3. Please fill your name and SID before you start the test
4. There are 6 questions
5. Each question carries 5 marks
6. Answer questions in the space provided
7. Rough work can be done on the last sheet marked as "rough work". Additional sheets are available from the instructor. These should be attached to your answer sheet.
8. Calculators and dictionaries are allowed
9. After answering each question, indicate your confidence level on the scale by placing an X (cross) mark

Student Information

Name:

Student ID:

Q1. Briefly describe three characteristics that you see as important which illustrate the functioning of Infosystems Corporation.

Q2. What in your view is causing the periodic oscillations in revenue (refer to figure 1, page 3 of case)?

How confident are you of the above answer? (Please rate on the scale)

0 10 25 50 75 90 100 %

Q3. Advise a long-term solution to this problem (described in Part-1 of case).

How confident are you of the above answer? (Please rate on the scale)

0 10 25 50 75 90 100 %

Q4. The following table shows the number of new employees hired (arrivals) and the number of employees who quit (departures) for the "XYZ" project.

Month	Arrivals	Departures
January 2004	13	0
February 2004	18	0
March 2004	10	4
April 2004	7	5
May 2004	10	7
June 2004	10	8
July 2004	12	0
August 2004	19	3
September 2004	11	9
October 2004	12	15
November 2004	8	17
December 2004	10	25
January 2005	8	28
February 2005	14	20

In which month did the largest number of employees work on the project?

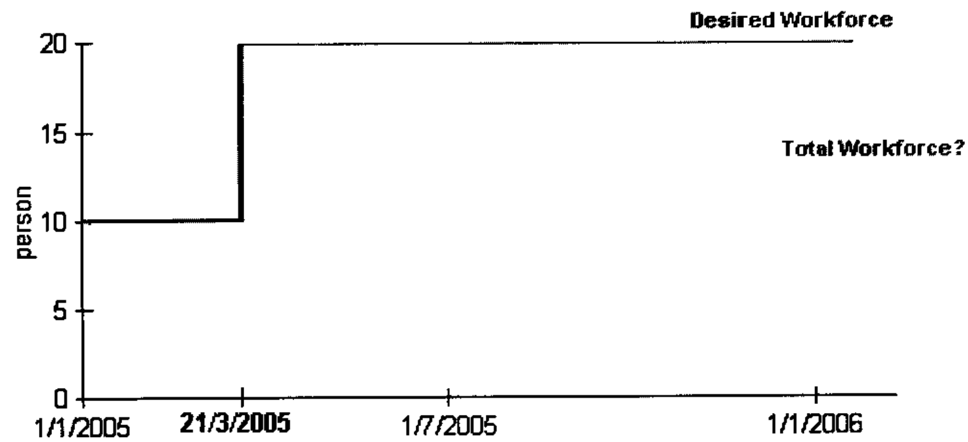
How did you arrive at the answer?

How confident are you on the accuracy of your answer?

0 10 25 50 75 90 100 %

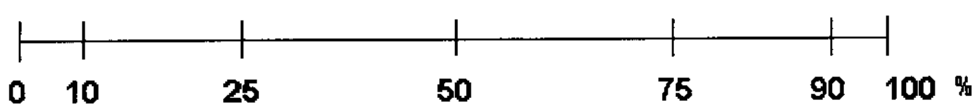
Answer questions 5 and 6 with respect to Part 2 of the case.

Q5. The project manager would like to increase the number of team members by 10 ("desired workforce") on the 80th day (as shown below). Predict the behaviour of the "total workforce" variable due to this change. Illustrate the graph of total workforce on the space below.



How did you arrive at the answer?

How confident are you on the accuracy of your graph?



Q6. The project manager would like to increase the number of team members by 10 ("desired workforce") on the 80th day (as shown in Q5). Predict the percentage of "experienced employees" on the project, during May 2005.

How did you arrive at the answer?

How confident are you on the accuracy of your answer?

0 10 25 50 75 90 100 %

Rough Work

4A: The following are the simple tasks that were provided to participants for the experiment reported in Chapter 4.

<u>Test 1</u>: The global demand for mobile telephony is growing. Government is installing mobile phone towers to enable a larger population to reap the benefit of mobile phones. The following table shows the number of projects on mobile towers started and the number of projects on towers completed by the government in each quarter.

Quarter	New projects	Completed projects
January-March 2001	0	0
April-June 2001	11	1
July-September 2001	44	2
October-December 2001	92	11
January-March 2002	53	17
April-June 2002	42	12
July-September 2002	37	25
October-December 2002	24	31
January-March 2003	47	55
April-June 2003	24	48
July-September 2003	15	59
October-December 2003	10	98
January-March 2004	16	35
April-June 2004	9	22
July-September 2004	4	8
October-December 2004	3	5
January-March 2005	1	2
April-June 2005	0	0

(i) In which quarter (e.g. April-June 2001) did the government have the maximum number of projects on hand?

____

(ii) How did you arrive at the answer?

____

____

____

____

____

<u>Test 2:</u> In all cities, there is generally a flow of people moving out of the city and people moving into the city in search of work. The table below shows the number of people arriving (inward-migration) and leaving the city (outward migration) in the small town of Emerald.

Year	In-migration (in hundreds)	Out-migration (in hundreds)
1900	15	0
1901	20	0
1902	12	6
1903	9	7
1904	12	9
1905	12	10
1906	14	0
1907	21	5
1908	13	11
1909	14	17
1910	10	19
1911	12	27
1912	10	30
1913	16	22

(i) In which year did the largest number of people stay in the city?

____

(ii) How did you arrive at the answer?

____

____

____

<u>Test 3:</u> The following table shows the number of new employees hired (arrivals) and the number of employees who quit (departures), for the "XYZ" project. Initially there are no employees on this project.

Month	Arrivals	Departures
January 2005	13	0
February 2005	18	0
March 2005	10	4
April 2005	7	5
May 2005	10	7
June 2005	10	8
July 2005	12	0
August 2005	19	3
September 2005	11	9
October 2005	12	15
November 2005	8	17
December 2005	10	25
January 2006	8	28
February 2006	14	30

(i) In which month did the largest number of employees work on the project?

____

(ii) How did you arrive at the answer?

____

____

____

____

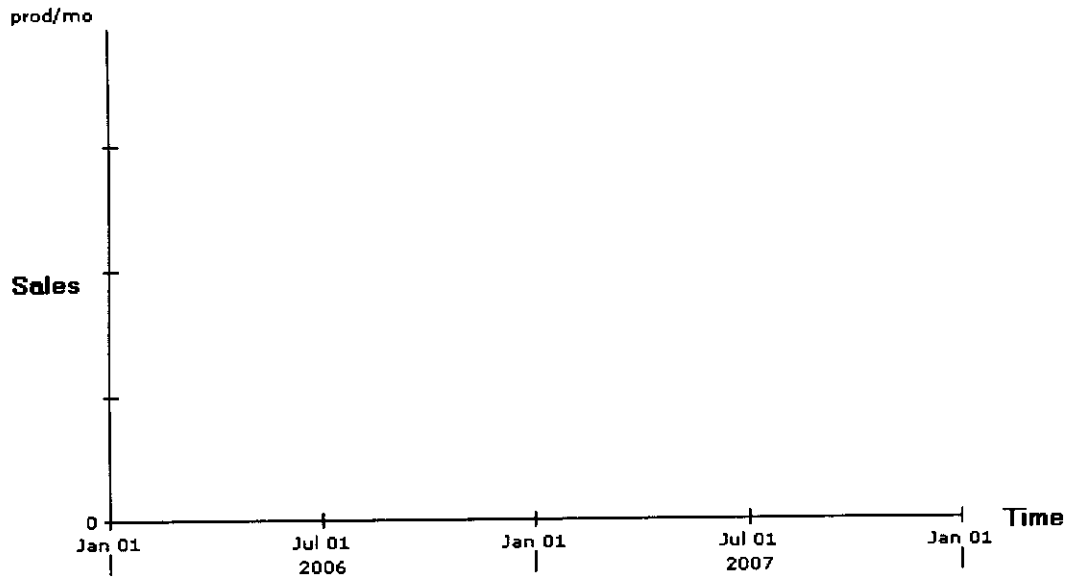
____

____

____

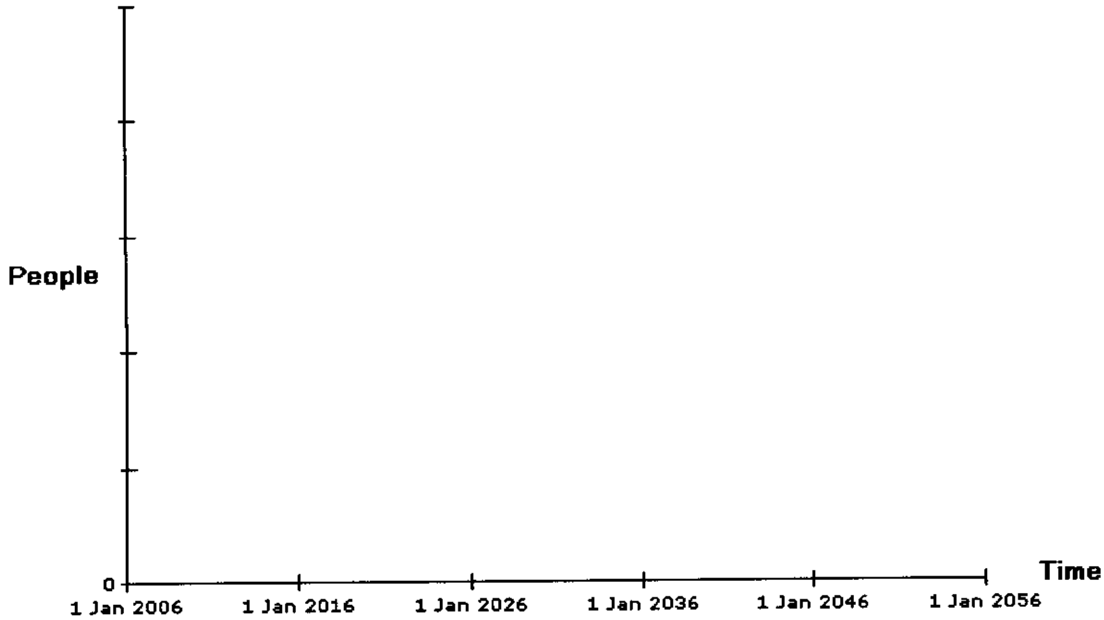
4B: The following are the complex tasks that were provided to participants for the experiment reported in Chapter 4.

<u>Test 1</u>: Draw the graph of the sales rate over of period of 2 years. Draw your answer in the space provided below.



List all the factors (directly or indirectly) on which sales of Aussierola in Byrock depend and illustrate how are these factors related to each other?

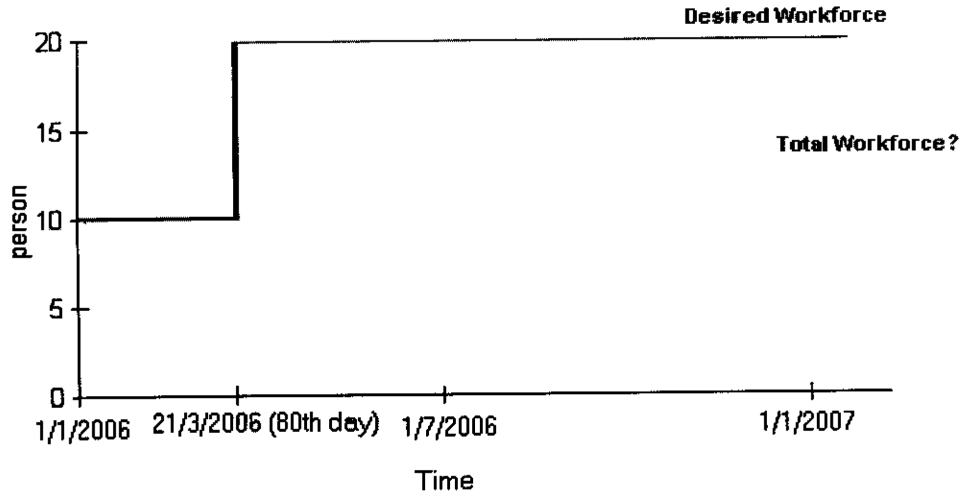
<u>Test 2</u>: Below is a graph of population versus time over a period of 50 years. Draw the graph of total population on the space provided below.



On what factors (directly or indirectly) does the total population of Winton depend? Illustrate your answer by listing the factors you took into account and how these factors relate to each other.

<u>Test 3</u>: The graph below depicts the increase in the number of team members by 10 ("desired workforce") on the 80th day from the start date of the project (as

described in the case). Draw the graph of the "total workforce", before and after this change. Draw your answer in the space provided below.



List all the factors (directly or indirectly) on which total workforce of Infosystems depends and illustrate how are these factors related to each other?

4C: The following is the mental effort rating scale that was provided to participants for the experiment reported in Chapter 4.

Write your Experiment ID number here (e.g. A1, B22):

____

For each of the following questions, please circle the response that best represents how you felt while you were completing the test.

I Overall, how easy or difficult did you find solving Question 1?
1. Extremely Easy 2. Very Easy 3. Moderately Easy 4. Slightly Easy 5. Neither Easy nor Difficult 6. Slightly Difficult 7. Moderately Difficult 8. Very difficult 9. Extremely Difficult
II Overall, how easy or difficult did you find solving Question 2?
1. Extremely Easy 2. Very Easy 3. Moderately Easy 4. Slightly Easy 5. Neither Easy nor Difficult 6. Slightly Difficult 7. Moderately Difficult 8. Very difficult 9. Extremely Difficult

4D: The following are the topics covered during system dynamics interventions

Qualitative System Dynamics

1. Introduction to SD
2. Brief history and application of SD
3. Definition of Qualitative SD/ systems thinking
4. Behaviour over time graphs
   i. Definition
   ii. Examples
   iii. Purpose
5. System
   i. Definition
   ii. Examples of systems
   iii. Class activity on differentiating system from collection
   iv. Characteristics of a system
   v. System boundary
   vi. Examples of dynamic problems associated with different systems
6. Endogenous view
   i. Definition
   ii. Examples
7. Holistic view of a problem
8. Linear thinking versus systems thinking
9. Feedback loops
   i. Cause and effect
   ii. Polarity
   iii. Symbols used in loops
10. Reinforcing loops
    i. Definition
    ii. Examples
    iii. Behaviour generated by reinforcing loops
    iv. Practice exercises
11. Balancing loops
    i. Definition
    ii. Examples
    iii. Behaviour generated by reinforcing loops
    iv. Causal loop diagrams
12. Causal Loop Diagrams (CLDs)
    i. Multiple loop systems
    ii. Explanation of behaviour of simple systems from its feedback structure
13. Loop dominance
14. Practice exercises (individual and group)
    i. Identifying polarity of cause and effect relationships
    ii. Identifying polarity of feedback loops
    iii. Drawing a feedback loop from a narrative
    iv. Drawing multiple feedback loops from a narrative
15. Delays
    i. Types of delays

ii. Representing delays in feedback loops

16. System Archetypes
    i. Definition
    ii. 9 types of system archetypes
    iii. Each archetype explained with example, feedback structure and the behaviour over time produced
17. Example of qualitative SD application – shown by instructor
    i. The case of FitCo (adapted from Daniel Kim)
18. Example of qualitative SD application – with active participation of participants
    i. Acme case study (adapted from Goodman)
19. Stock and Flows
    i. Principle of accumulation
    ii. Examples
20. Summary of qualitative SD tools

References

1. *The Fifth Discipline Field book* by Peter Senge et.al., Doubleday
2. *Introduction to Systems Thinking*, by Daniel Kim, Pegasus Communications
3. *Systems Thinking Tools*, by Daniel Kim, Pegasus Communications
4. The Systems Thinking Playbook, Linda Booth Sweeny and Dennis Meadows
5. "...to enable Freshmen to do what once strained Newton's powers..." Systems Thinking and Dynamic Modeling: a Conference for K-12 Education, 1998 by George Richardson

Quantitative System Dynamics

1. Introduction to SD
2. Brief history and application of SD
3. Definition of Quantitative SD
4. Behaviour over time graphs
    i. Definition
    ii. Examples
    iii. Purpose
5. System
    i. Definition
    ii. Examples of systems
    iii. Class activity on differentiating system from collection
    iv. Characteristics of a system
    v. System boundary
    vi. Examples of dynamic problems associated with different systems
6. Use of formal models
7. Stocks and flows
8. Representing feedback with stock and flow models
9. Powersim
    i. Getting started with Powersim
    ii. Introductory examples

iii. Activities with simple stocks and flows
iv. Creating a new simulation project
v. Creating units of measurement
vi. Creating variables, links and flows
vii. Defining variables and flows
viii. Setting up the simulation
ix. Creating data in- and output objects
x. Adding navigation capabilities and documentation
xi. Creating and saving reference data
xii. Analyzing the model

10. Delays
    i. Representing delays in feedback loops
11. Hands-on activity
    i. Creating and running models from simple to complex scenarios
12. Summary of quantitative SD tools

References

1. Roadmaps study guide – MIT SD collection
2. Business Dynamics, by John Sterman
3. Industrial Dynamics by J.W. Forrester
4. Powersim Online help
5. Powersim and system dynamics by Pal Davidsen

5A: The following is the questionnaire that was provided to the participants after the longitudinal experiment reported in Chapter 5.

Exp. ID:

Between the period 17th September, 2006 and 2nd February, 2007 were you involved in:

Using system thinking tools such as causal loop diagrams/feedback loops/system archetypes – Yes, No
If *Yes*, briefly describe in what form were you involved and how frequently.

____

____

Using stock and flow diagrams – Yes, No
If *Yes*, briefly describe if what form were you involved and how frequently.

____

____

Using behaviour over time graphs – Yes, No
If *Yes*, briefly describe where did you use them and how frequently.

____

____

Using any system dynamics software for modelling (e.g. Powersim, Vensim, IThink) – Yes, No
If *Yes*, briefly describe in what form were you involved and how frequently.

____

____

Using any system dynamics based simulation game - Yes, No
If *Yes*, briefly describe the name of the game, in what form were you involved and how frequently.

____

____

Please describe if you were involved in any other form of systems thinking/system dynamics in the period mentioned above (e.g. reading about, talking to people about systems thinking/system dynamics etc.)

Did the systems thinking/system dynamics interventions during September 2006 change your approach of dealing with problems? Yes, No
If *Yes*, briefly explain what were these changes

5B: Powersim^{TM} screenshots

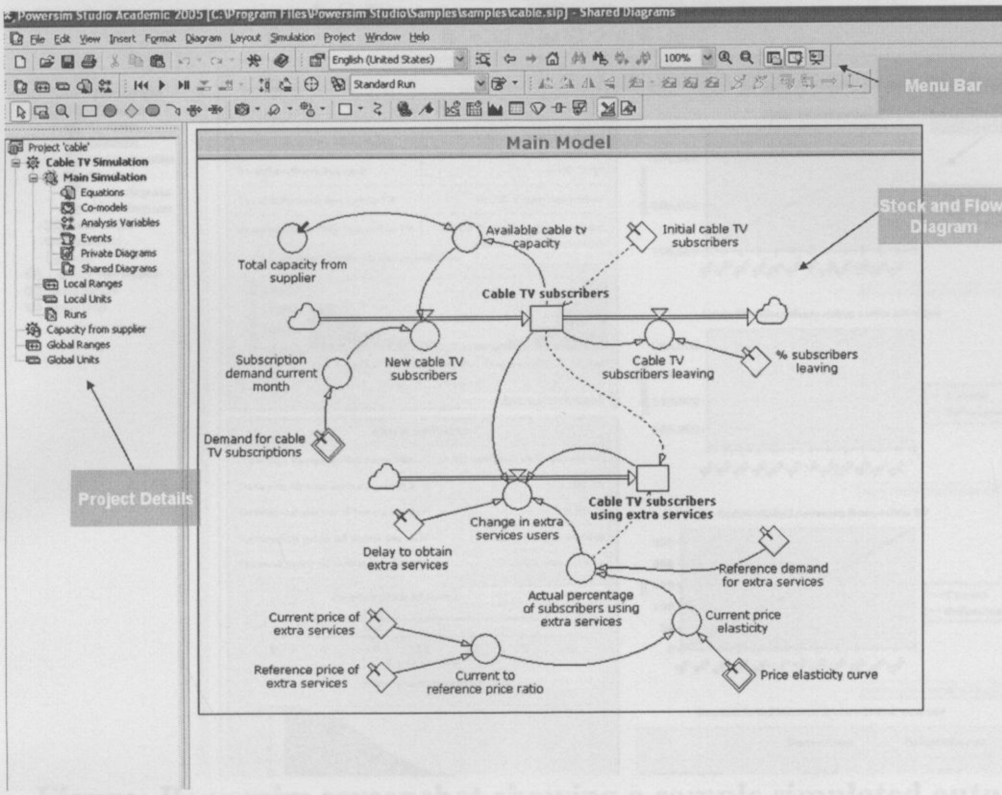


Figure: Powersim screenshot showing a sample stock and flow diagram

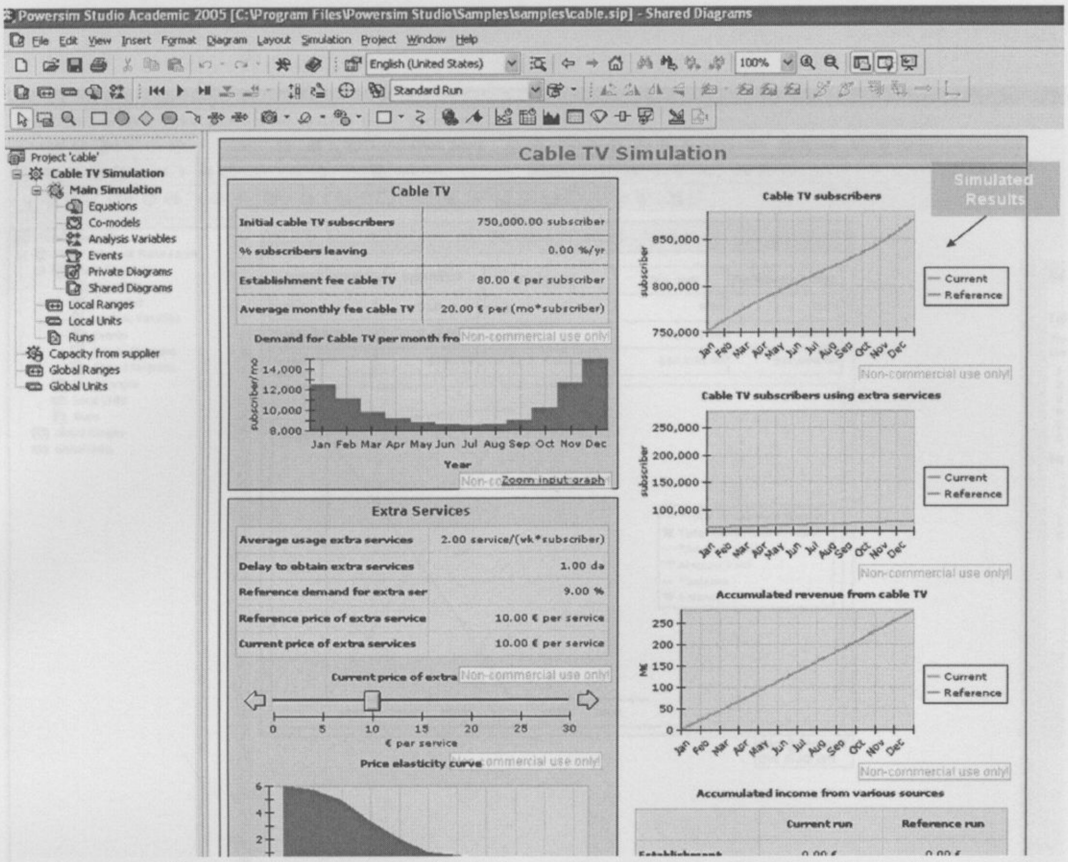


Figure: Powersim screenshot showing a sample simulated output

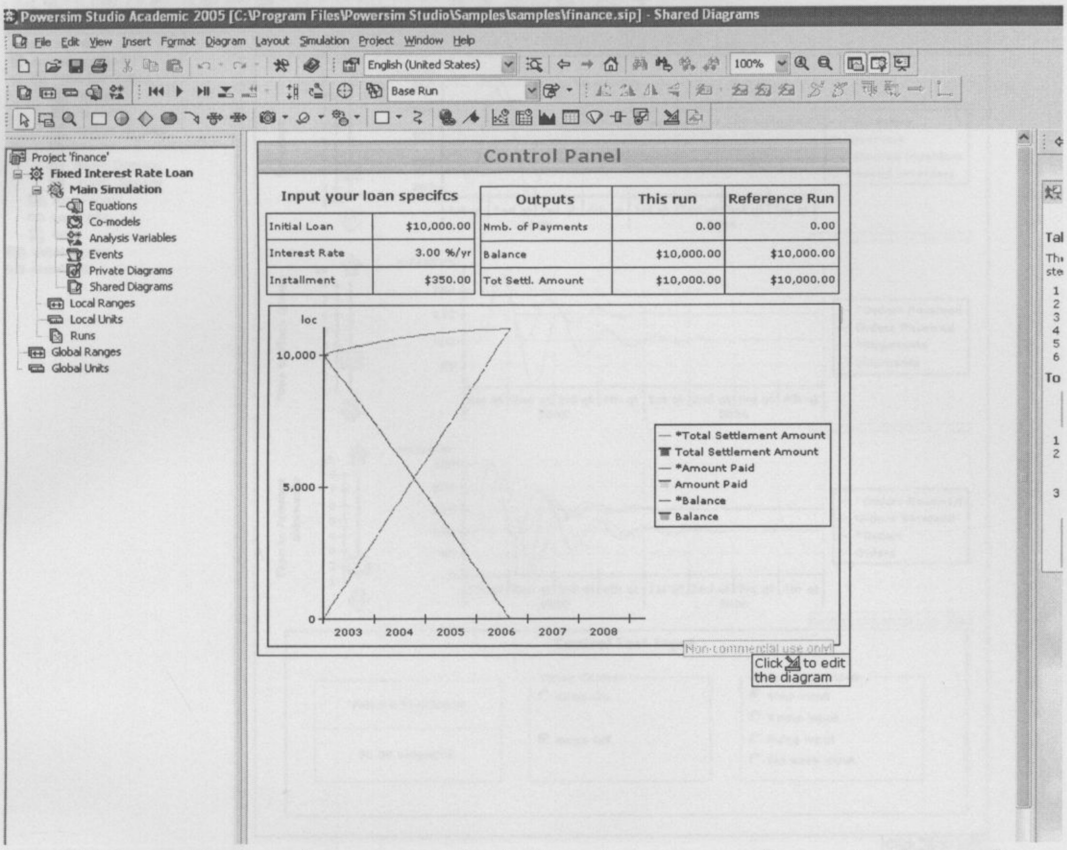


Figure: Powersim screenshot showing a sample control panel

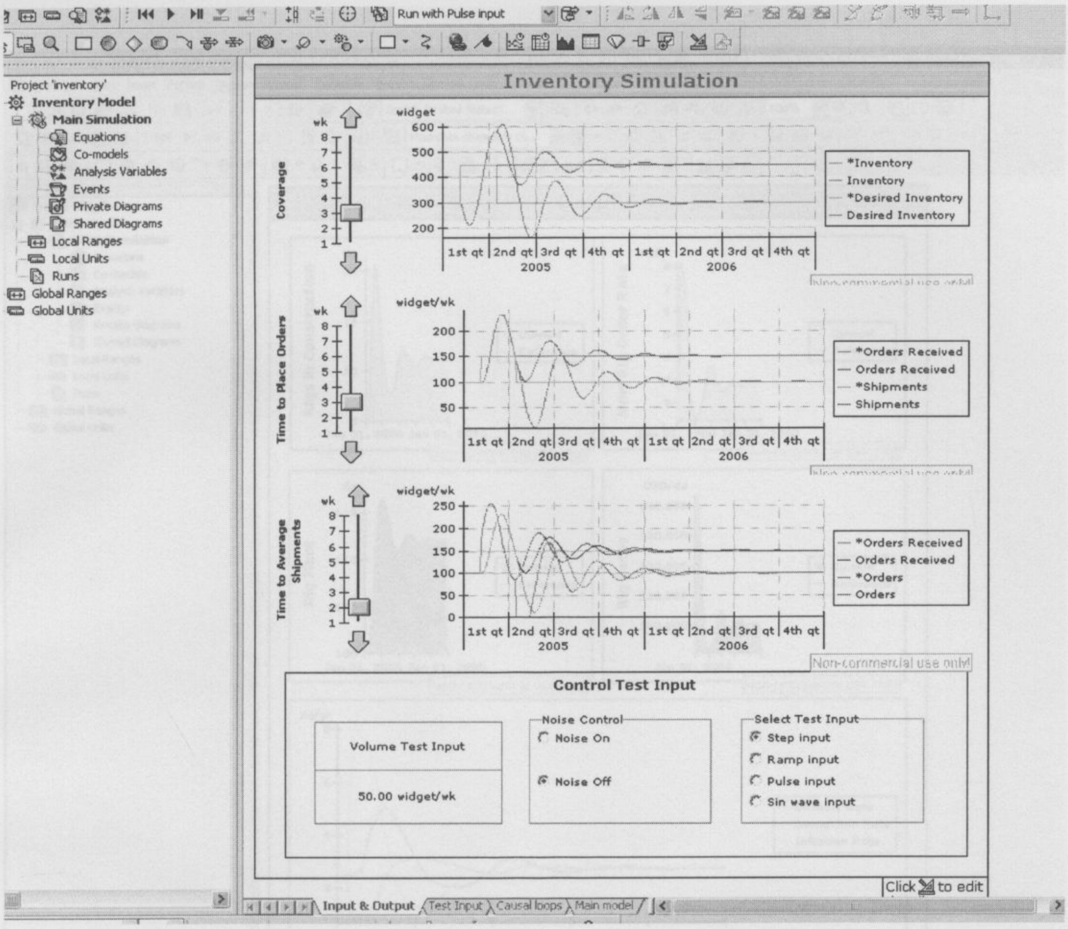


Figure: Powersim screenshot showing a sample simulated output

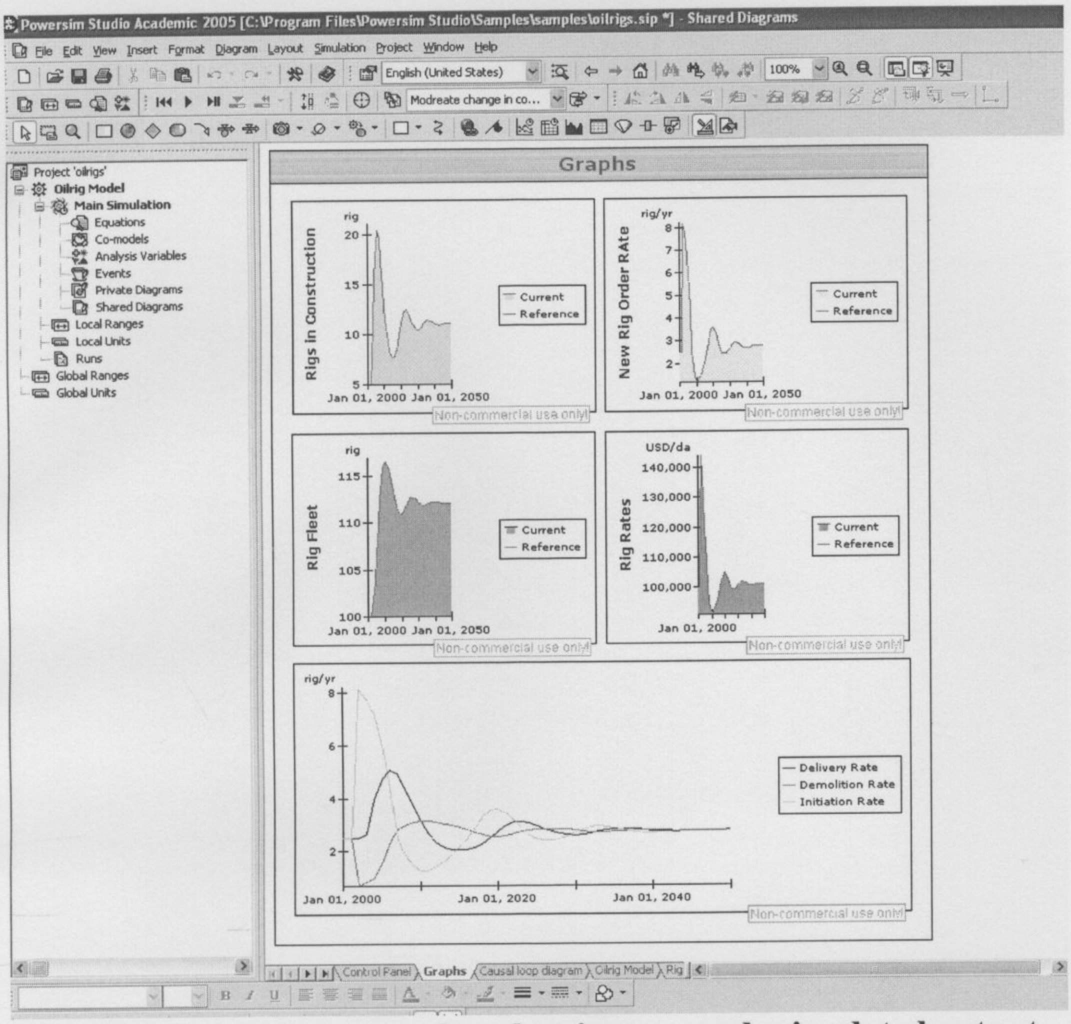


Figure: Powersim screenshot showing a sample simulated output