Media and technology effects in the forecasting task

Author:
Portocalis Van Toorn, Christine

Publication Date:
2004

DOI:
https://doi.org/10.26190/unsworks/9634

**The School of Information Systems,
Technology and Management,
Faculty of Commerce and Economics,
The University of New South Wales**

Media and Technology Effects in the Forecasting Task

Christine Portocalis-Van Toorn

A thesis submitted in fulfilment of the requirements for the
Degree of Master of Commerce (Hons).
2004

I hereby declare that this submission is my own work and to the best of my knowledge it contains no material previously published or written by another person, nor material to which a substantial extent has been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the Thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the Thesis.

I also declare that the intellectual content of this Thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

Christine Portocalis-Van Toorn
2004.

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ....

ACKNOWLEDGEMENTS

I owe a considerable vote of thanks to my family – my husband John, my children Tara and Zachary, my Mother Teresa, my Uncle Joe and Esther – for their patience, encouragement and support throughout the duration of this project… which we all thought would never end! I would also like to thank my later Father, John, for instilling in me the importance of education from a very early age.

I also owe a great deal of gratitude to my supervisor and mentor, Professor Marcus O'Connor who gave me more time, guidance and encouragement than I could ever have imagined. A significant proportion of my achievement is due to Marcus' efforts over the many months … years!! spent on this project. Marcus' never-ending enthusiasm for the area and his great depth of knowledge was a constant source of inspiration.

I would also like to thank my colleague Peter Parkin for his friendship and never-ending support throughout this Project. Peter was an amazing source of inspiration and encouragement over the years and was always there to provide input and assistance whenever necessary. There are also many other colleagues in the School and friends who have helped me over the years, to all of them, in particular, Dr. Meliha Handzic, I would like to say "thank you".

The last word goes to Professor Marcus O'Connor – thank you Marcus for always being there, I certainly could not have done this without you.

TABLE OF CONTENTS

ABSTRACT

A significant body of research has investigated the effects of *media* and related *technology* on task performance. This has been a much debated topic, and one which has become increasingly important over the past 30 years. Much of the interest in this area can be directly attributed to the proliferation of computers in the workplace and the introduction of automated decision support systems. Researchers have argued that the ever increasing emphasis on business productivity and efficiency has led to the demand for a quantifiable measure of performance improvement when using such automated tools. In turn, this demand has led to the emergence of a new research area, namely that of Task-Technology Fit. This area is dominated by the work of Goodhue and Thompson (1995) which sought to empirically test their proposed research model.

The overall aim of this study is to extend research in the area of TTF by conducting a series of experiments in order to review the impact of *media* and related *technology* on task performance. The primary variables of interest to this study are: the Tool – the *media* and related *technology*; Task Complexity – applying Campbell's (1988) classification; and the People, or Individuals undertaking the task.

Having identified forecasting as being an essential component of many business processes and a task that is common in decision making, we used the judgemental extrapolation of time series data as our Task or *object of interest*. This provided us with an objective measure of performance – the accuracy of forecast predictions. Using our proposed Research Model of TTF, and varying each of the variables of interest we conducted a series of controlled experiments to review the impact upon task performance

Results showed that accuracy gains were able to be achieved by varying the Tool – *media* and related *technology* – and through the use of more mature and/or educated People. However, results in the area of Task Complexity were inconclusive. Further studies need to be conducted in this latter area in order to identify a proven relationship between the variables in the proposed Research Model of TTF.

1. Introduction

1.1 Background to this study

The issue of presentation and working *media* or choice of optimum display format (*mode*) in order to maximise task performance has been a much debated topic. From Washburne's (1927) study which reviewed the adequacy of different information display formats for tasks of varying complexity, this issue has remained a much deliberated topic. In particular, this has become increasingly important over the past 30 years, since the proliferation of computers into the workplace and the introduction of automated decision support systems. Researchers have argued that the ever increasing emphasis on business productivity and efficiency has led to the introduction of an additional perspective, namely that of Task-Technology Fit (TTF) (Goodhue, 1995). Studies in the area of TTF have been dominated by Goodhue and Thompson (1995). The basic premise behind their proposed theoretical model is that:

> "Task-Technology Fit (TTF) is the degree to which a technology assists an individual in performing his or her portfolio of tasks. More specifically, TTF is the correspondence between task requirements, individual abilities, and the functionality of the technology." (p. 218).

It is the interaction between the task, the technology and the individual which is crucial to the framework of TTF. Goodhue and Thompson (1995) further propose that "as the gap between the requirements of a task and the functionalities of a technology widens, TTF is reduced" (p. 218).

There has been much research conducted regarding the relative merits of different display formats or *modes* of presentation, particularly so in the decision making environment. Typically these studies have concentrated on the relative merits between data presented in either graphical or tabular format (Benbasat and

Schroeder, 1977; Lucas and Nielsen, 1980; Lucas, 1981; Lawrence, 1983; De Sanctis, 1984; Angus-Leppan and Fatseas, 1986; Dickson, De Sanctis and McBride, 1986; Blocher, Moffie and Zmud (1986); Remus, 1984, 1987; Vessey, 1991; Harvey and Bolger, 1996 - for a more detailed review please refer to section 2.3.5).

In addition, there have also been numerous studies conducted on the relative merits of presenting the same data across two different *media* - namely paper-based and screen-supported. (A comprehensive review is provided in Dillon, 1992. A summary of previous research is provided in section 2.6.6.1 and a full listing of previous research is provided in Appendix 2). Research in the area of Media studies has tended to concentrate on two major tasks - namely proof-reading and reading for comprehension. Therefore, in terms of the range of tasks examined, research in the area of Media studies has not been extensive. This Thesis will seek to extend the area by introducing a new task or *object of interest* – the judgemental extrapolation of time series data.

The overall aim of this Thesis is to extend research in the area of TTF. Specifically, we are interested in reviewing the impact of *media* and the related *technology* on task performance. Of particular interest, are previous studies whose primary focus has been the influence of media and technology and the subsequent identification of variables used to measure and compare task performance across the different media.

The judgemental extrapolation of time series data was selected as our *object of interest* as it was deemed to be a task common in decision making. Forecasting is an essential component of many business decisions. It is used extensively by many areas in the business environment such as sales prediction; production scheduling; budgeting and financial planning. Research in the area of forecasting has also identified judgemental forecasting as the preferred method used by practitioners (Lawrence, 1983; Mentzer and Cox, 1984; Lawrence *et al*, 1985; Dalrymple, 1987; Taranto, 1989; Sanders and Manrodt, 1994; Goodwin, 1998 - for a more detailed review please refer to section 2.3).

In an attempt to review the impact of *media* and related *technology* on task performance, this Thesis will utilise results from Media studies and research in the areas of Task-Technology Fit, judgemental time series forecasting and Task itself. For the purpose of this Thesis, a "task" refers to the actual work being performed by the subject. The outcome or completion of the task results in the identification of an objective measure. In turn, this can be used to judge overall accuracy or performance between two or more treatment conditions.

The knowledge gained from previous studies will be used to suggest the impact of *media* and related *technology* on the judgemental extrapolation of time series data. The predicted impact will then be experimentally assessed. Specifically, this study will compare the accuracy of judgemental forecasts produced using paper-based time series representations, with those produced using software-supported time series displays. The primary focus is the impact of presentation and working *media* and related *technology* on the accuracy of predictions in the judgemental extrapolation of time series data. This Thesis will aim to provide some insight into the following question;

> Are there any differences in judgemental forecast accuracy that are due to the *media* and related *technology* used in the forecasting process?

1.2 Motivation behind this research

The overall motivation for this Thesis is to review the impact of *media* and related *technology* on task performance. Our aim is to extend research in the area of TTF by conducting a series of controlled experiments on a task that will;

provide an objective measure of performance; and is widely used in a business environment.

This Thesis will re-examine previous media studies and derive tentative explanations of how the use of screen-based *media* and related *technology*, as opposed to paper-based *media* and related *technology*, could affect task performance. Using these explanations, we will test the veracity of differences by conducting a series of controlled experiments involving the traditionally paper-based task of judgemental time series forecasting. Our aim is to compare the accuracy of predictions under each of the treatment conditions – paper-based and screen-based *media* and related *technology* – and thus provide empirical evidence for our proposed research model as presented in section 2.4.

1.3 Research Outline

Following this introduction, the next Chapter will present a review of the literature in relation to the variables of interest under observation in this Thesis. Our overall review will span across four major research areas; Task-Technology Fit (TTF) to be covered in section 2.2; judgemental forecasting – section 2.3; Task – in particular task complexity, to be covered in section 2.5; and Media studies- section 2.6.

Our proposed Research Model, outlining the relationship between each of the variables of interest to this Thesis will be presented in section 2.4.

Chapter 3 will present our research methodology. This chapter will cover the operationalisation of each of the variables of interest as well as providing summary details of the experimental studies to be undertaken as part of this Thesis.

A detailed discussion of each of the experimental studies undertaken as part of this Thesis will be presented in Chapters 4 to 7. Hypotheses developed for examination will be reviewed separately as part of each of the four experimental studies.

Last of all, in Chapter 8 we present the conclusions of our research and the implications for future research leading to a refinement of the model of TTF. Our ultimate aim being the development of a comprehensive model which may be used by researchers and practitioners seeking to measure the effectiveness of organisational information systems.

2. Literature Review

2.0 Review and Analysis of Related Literature

2.1. Introduction and Aim of the Research:

The overall aim of this Thesis is to extend research in the area of Task-Technology Fit. Specifically, we are interested in reviewing the impact of *media* and the related *technology* on task performance. Of particular interest, are previous studies whose primary focus has been the influence of media and technology and the subsequent identification of variables used to measure and compare task performance across the different media.

We identified the judgemental extrapolation of time series data as a suitable task, or *object of interest* for this Thesis. The literature review in this area will provide a definition of the task as well as a comprehensive review of previous studies conducted in the area of judgemental extrapolation of time series data. This will be presented in section 2.3.

A review of the literature and previous research in relation to Task itself is presented in section 2.5. This enables us to reassess the theoretical and empirical studies undertaken in the area of Task identification and definition. The performance of this review shall seek to identify a suitable taxonomy or theoretical model of tasks and their related attributes. This will be used as the basis to classify our experimental task.

An analysis of the Media literature indicates that over many years numerous studies have been undertaken to investigate the influence of different *media* and related *technologies* on task performance. Findings from our review of the literature are presented in section 2.6. The primary focus of many of these previous studies has been a comparison of performance *between* the media. Some researchers have attempted to attribute reported differences in performance

to specific factors. However, in general, researchers have tended to concentrate on the *tangible* differences between the media. In his 1992 review of the Media literature, Dillon advises that significantly more research is required before broader goals, such as designing software systems that can improve on paper, can be adequately addressed.
This Thesis will seek to extend this area further by not only reviewing performance differences between the media, but also *within* the media by varying the related *technologies*.

Goodhue (1988) also noted a shortfall in the literature in relation to the impact of *media* and *technology* on task performance. In an effort to address this shortfall, Goodhue proposed an initial model of Task-Technology Fit. This model was the basis for Goodhue and Thompson's 1995 model of TTF, an integral component of this Thesis. Our review of the literature in relation to Task-Technology Fit is presented and discussed in detail in section 2.2.

2.2 Task-Technology Fit – Review and Analysis of Related Literature

2.2.1 Introduction

The major focus of this Thesis is to review the impact of *Media* – paper versus screen-based – and the related *Technology*, on task performance, and to conduct such research on a task that is deemed to be relevant in the decision making context.

This section of the literature review will focus on studies in the area of Task-Technology Fit (TTF), a fairly "young" research area and one which is dominated by Goodhue and Thompson's 1995 study. The basic premise behind the proposed theoretical model is that:

> "Task-Technology Fit (TTF) is the degree to which a technology assists an individual in performing his or her portfolio of tasks. More specifically, TTF is the correspondence between task requirements, individual abilities, and the functionality of the technology." (p. 218).

It is the interaction between the task, the technology and the individual which is crucial to the framework of TTF. Goodhue (1995) maintains that TTF is essentially an additional perspective that is needed in order to satisfy the ever increasing emphasis on business productivity and efficiency.

Goodhue and Thompson (1995) propose that "as the gap between the requirements of a task and the functionalities of a technology widens, TTF is reduced" (p. 218). In other words, they maintain that if the working *medium* and corresponding technology do not represent an optimum solution, then task performance may well be compromised (regardless of the level of *utilization*). In turn, an increase in either task complexity, or a decrease in the functionality of technologies designed to support those tasks will lead to a subsequent decrease in TTF (p. 218).

Goodhue and Thompson further propose that "at any given level of utilization, a system with higher TTF will lead to better performance since it closely meets the task needs of the individual"(p. 218). They define *utilization* as "the behaviour of employing the technology in completing the tasks" (p. 218). As mentioned in the previous paragraph, the researchers propose that the primary focus of their study is not to link the *level* of utilization with task performance, but to identify individual factors which may contribute to achieving positive performance impacts. They maintain that much of the "utilization" research has tended to *assume* that increased utilization of Information Systems will generally lead to positive performance impact (presented in Goodhue and Thompson; 1995 p. 214). The researchers propose that this is often not the case and that positive performance impacts can only be obtained when the technology provides features and corresponding support that are both deemed to be a good "fit" to the requirements of the task. Information systems will "...have a positive impact on performance only when there is correspondence between their functionality and the task requirements of users" (p. 214). They maintain that if the task and technology do not represent a good "fit", no amount of *utilization* will lead to positive performance impacts.

Goodhue and Thompson acknowledge that much of the previous research in the area of "fit" has centered on the graphs versus tables literature. In particular, research in relation to individual decision-making performance. Citing the work of Vessey 1991 (and others p. 214 – many of whom will be referenced in section 3), and her theory of "cognitive fit" they acknowledge her valuable contribution to the research area. The essence of Vessey's theory being – that mismatches between technology characteristics and tasks have an adverse effect on the decision–making process, and in turn, performance. In his 1995 study, Goodhue also argued that Vessey's theory of "cognitive fit" could essentially be seen as a more general case of the TTF perspective (p. 1831). He maintains that whilst research in the area of "cognitive fit" "... tends to focus narrowly on cognitive task processes, TTF applies to a more macro task domain including not only cognitive but also other task processes such as the mechanical process of accessing data ..." (p. 1830).

Goodhue and Thompson (1995) specify that for the purpose of their research and their proposed research model, *technology* can generally be viewed as the tools that are used by individuals whilst carrying out their designated tasks. When viewed in the context of IS research, the technology generally refers to the computer systems as a whole. Typically, this includes the hardware, software, the manipulation mechanism which allows the user to articulate their needs, as well as associated services – all serving to assist users with the performance of their task/s (p. 216).

The researchers maintain that although there is a proliferation of Information Systems all deemed to assist the individual with their decision-making – particularly in an organisational context – there is still much concern about the lack of knowledge in this research area. Much of the concern has centered on the need to more fully understand and be able to evaluate the linkage between information systems *use* and individual *performance.*

The work undertaken by Goodhue and Thompson (1995) sought to expand upon the basic model of TTF as presented by Goodhue 1988. The fundamental premise behind Goodhue's 1988 model is the assumption that a positive correlation between task needs and system functionality – defined as TTF – will result in positive performance impacts (Goodhue 1998, p. 107). The basic model of Task-Technology Fit, as proposed by Goodhue (1988) is presented in the next section as figure 2.2.2.1.

2.2.2 The Basic Model of Task-Technology Fit – Goodhue (1988)

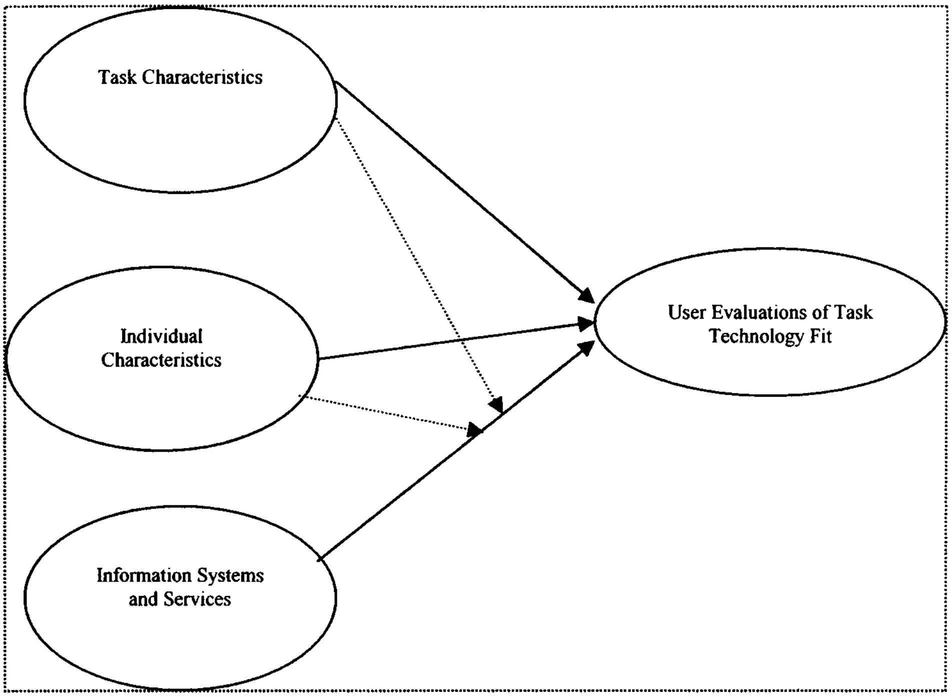


Figure 2.2.2.1 The basic model of Task-Technology Fit (Goodhue 1988) (dotted lines are "moderating" interaction effects)

This initial model identifies three variables as having an impact upon performance – measured in terms of User Evaluations of TTF. The variables in question are (1) Task characteristics; (2) Individual (person) characteristics and (3) Information Systems and Services. Of particular interest to the proposed study is the area of Task characteristics – this will be reviewed in detail in section 2.5 and the area of Information Systems in so far as it relates to the *Media* – paper versus screen-based – and associated *Technology* being used in order to undertake a given task – this will be covered in section 2.6.

Goodhue and Thompson's 1995 study sought to expand upon Goodhue's 1988 model, and further extend the notion "that performance impacts will result from task-technology fit – that is, when a technology provides features and support that "fit" the requirements of a task" (p. 214). In an effort to provide support for their proposed theoretical model – presented below as Figure 2.2.3.1, the researchers analysed data obtained from over 600 individuals, using 25 different information technologies across 26 different departments in two companies (p. 214). Their study sought to highlight "… the importance of the fit between technologies and user's tasks in achieving individual performance impacts from information technology" (p. 213). It also sought to "decompose" TTF into its individual components – 16 initially – thus enabling analysis to be conducted on the data in order to review the impact of individual components on task performance, as well as identify and measure any interaction effects. Their final model incorporated only 8 of the original TTF dimensions (a copy of the dimensions can be found at Appendix 1).

What must be noted at the outset here is that given the nature of Goodhue and Thompson's study – a field study – it can be acclaimed in terms of its external validity. However one of the major limitations of their study is that they were not able to obtain an *objective* measure of performance in the field context. They themselves admit that no objective measure would have been compatible across such a large number of individuals with many diverse task portfolios. Instead, they had to measure performance *subjectively* – in terms of "perceived" performance impacts as rated by task-doers i.e. User evaluations of TTF.

2.2.3 Expanded Model of Task-Technology Fit – Goodhue and Thompson (1995)

The expanded model of Task-Technology Fit as proposed by Goodhue and Thompson (1995) is presented here as Figure 2.2.3.1.

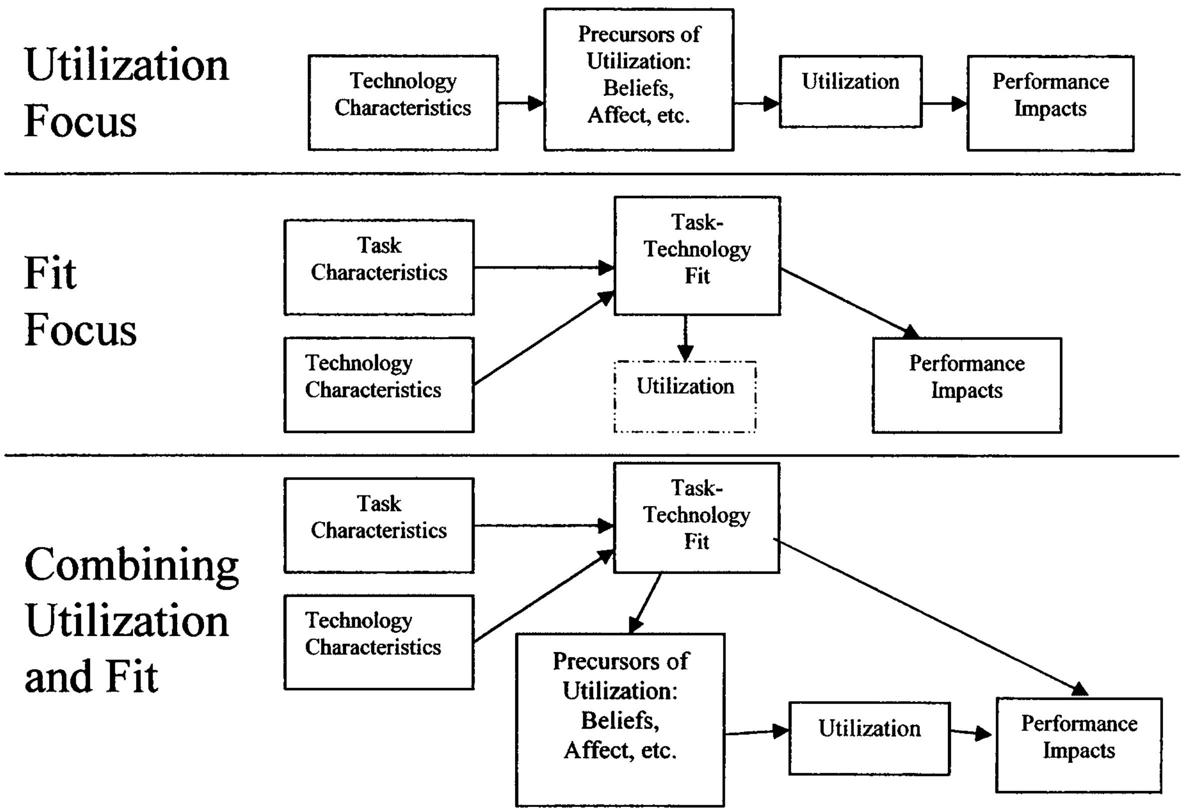


Figure 2.2.3.1 Three Models of the Link From Technology to Performance (Goodhue, D.L. and Thompson, R.L. 1995 p. 215)

The researchers cite and attribute credit to previous work undertaken in the broad area of TTF – specifically research seeking to link Technology to Performance. They identify two complementary research streams, or areas of focus – that of *utilization* and that of *fit*.

2.2.4 Utilization Focus Research

A model of the traditional "utilization focus" stream is represented by the top section of Figure 2.2.2. It highlights the relationship between the Technology Characteristics and Performance Impacts.

In their review of the literature in relation to *utilization,* Goodhue and Thompson (1995) note that "stated or unstated, the implication is that increased utilization will lead to positive performance impacts" (p. 214). However, the researchers maintain and seek to prove that if the task and technology do not represent a good "fit", no amount of *utilization* will lead to positive performance impacts. In addition, Goodhue and Thompson (1995) note that often utilization of an Information System may not necessarily be "voluntary" – an individual may *have to* utilise the system as part of their job description or function. Goodhue (1995) advises that the basic premise behind TTF is that users will evaluate the technology options available to them, but may not necessarily have a choice about whether or not they use them (p. 1830). Goodhue and Thompson (1995) maintain that utilization alone is not a worthy indicator of how good a system is, in terms of satisfying users' needs and leading to positive performance impacts. They note that "utilization of a poor system ... will not improve performance" and that even when utilization is voluntary, these systems may still "...be utilized extensively due to social factors, habit, ignorance, availability, etc. (p. 216).

Noting the deficiency of the "utilization focus" stream of the literature, the researchers sought to incorporate the notion of "fit" into their proposed theoretical model.

Previous research in the "utilization focus" stream has tended to adopt user attitudes and beliefs in order to predict utilization for information systems. In order to empirically test their proposition, and give recognition to the fact that utilization may be in fact mandatory, Goodhue and Thompson assigned and measured "overall dependence on systems" as their measure of utilization (p. 223).

2.2.5 Fit Focus Research

In their review of the literature in the "fit focus" stream, Goodhue and Thompson (1995) strongly cite previous work undertaken by Dickson *et al.* (1986), Jarvenpaa (1989) and Vessey (1991). They note that research in this area has added to previous research in the "utilization focus" stream, by highlighting the importance of both Task Characteristics, as well as Technology Characteristics, in measuring Performance Impacts. The model incorporating the "fit focus" stream is represented by the middle section of the model in Figure 2.2.3.1 – in addition, there is the suggestion of the link between *fit* and *utilization* as denoted by the dotted arrow. Goodhue (1995) initially noted one of the limitations of the "fit" stream of research – the major focus being the "...impact of information *representations* on individual decision making performance" (p. 1830 – emphasis added). Goodhue and Thompson (1995) further highlighted some of the major limitations of the "fit focus" stream, citing that previous studies "...focusing on fit alone do not give sufficient attention to the fact that systems must be *utilized* before they can deliver performance impacts" (p. 216 – emphasis added). In addition, they reiterate that utilization is a complex outcome, one which is often linked to other social and situational factors – such as habit, ignorance or availability. They conclude that the *fit* model can only serve to "...benefit from the addition of this richer understanding of utilization and its impact on performance" (p. 216).

The resulting model as proposed by Goodhue and Thompson is represented in the bottom section of Figure 2.2.3.1 – and is further expanded upon in Figure 2.2.5.1 below. In combining the *utilization* and *fit* focus, the researchers' aim was to acknowledge that task performance could be influenced, and in turn determined, by both utilization and TTF.

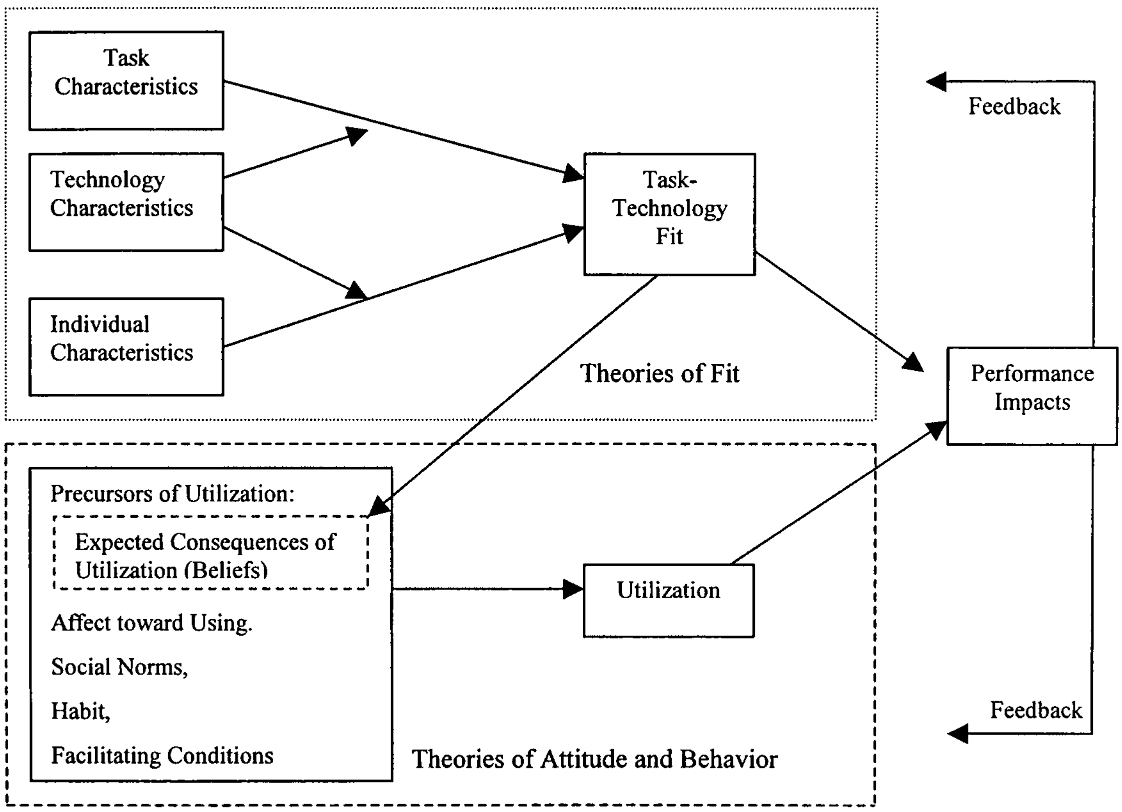


Figure 2.2.5.1 The Technology-to-Performance Chain
Goodhue, D.L. and Thompson, R.L. 1995 p. 217.

Goodhue and Thompson initially refer to this model as the Technology-to-Performance-Chain – deeming it to be a "...model of the way in which technologies lead to performance impacts at the individual level" (p. 216). They further note that the TPC model seeks to combine theories focusing on task–

system fit and utilization; introduce an additional variable – Individual Characteristics – and provide a definition for each of the major variables (on p.216):

Technologies:	tools used by individuals in carrying out their tasks;
Tasks:	actions carried out by individuals in turning inputs into outputs;
TTF:	the degree to which a technology assists an individual in performing his or her portfolio of tasks;

The inclusion of "Individual Characteristics" as an additional variable in the TPC Model serves to highlight the fact that it is the individuals which who are using the technology to assist them in carrying out their tasks. The authors note that "Characteristics of the individual (training, computer experience, motivation) could affect how easily or well he or she will utilize the technology" (p. 216).

Goodhue and Thompson's TPC model seeks to combine all the major variables from previous models from both the *fit* and *utilization* streams. The resulting model seeks to determine the impact upon performance in terms of Task Characteristics; Technology Characteristics; Individual Characteristics and Utilization. However, it must be noted it is difficult to fully understand the nature of the link between TTF and Utilization (the arrow from TTF to Utilization). The TPC model appears to indicate that TTF may well be a predictor of utilization – that in some way it will predict or determine the expected consequences or beliefs of utilization. The researchers validate this link by maintaining that "...TTF should be one important determinant of whether systems are believed to be more useful, more important, or give more relative advantage" (p. 218). However, this is not in keeping with their previous line of argument – that utilization of an Information System may not necessarily be "voluntary", an individual may *have to* utilise the system as part of their job description or function. It is questionable whether the TPC model would apply to such situations of mandatory use. The TTF may well be found to be low, but the *utilization*, by nature of it being mandatory, would still need to be ranked as fairly high – in terms of Goodhue and Thompson's measure of utilization, "overall dependence on systems".

The impact of "Individual Characteristics" is also unclear. On the one hand, the researchers advise that this variable could affect how "easily and well" task-doers will *utilize* the technology (p. 216). On the other hand, there is no direct link representing or recognising this relationship between "Individual Characteristics" and *utilization*. In addition, given Goodhue and Thompson's proposition that "at any given level of utilization, a system with higher TTF will lead to better performance" (p. 218) the relationship between TTF and utilization, and subsequent impact on task performance remains unclear.

However, for the purpose of this Thesis, utilization is not a variable of interest. As previously noted, the major focus is to review the impact of *Media* and the related *Technology* on Task performance. As such, utilization is not deemed to be a variable for consideration. Having previously identified a major limitation of Goodhue and Thompson's study – the fact that they were not able to obtain an objective measure of performance in the field context – this Thesis will aim to address this limitation. We propose to extend the area, by conducting a controlled laboratory experiment on a task that will; provide an objective measure of performance; and is widely used in a business environment, namely – the judgemental extrapolation of time series data.

A full description of the task identified as being the *object of interest* for this Thesis – the judgemental extrapolation of time series data – is provided in section 2.3.

2.3 Judgemental Time Series Forecasting – Review and Analysis of Related Literature
Task Description and Measurement

Forecasting is an essential component of many business decisions, e.g. sales prediction, production scheduling, budgeting and financial planning all require accurate forecasts. The basic task of forecasting can be described as "the gathering and analysing of repeated observations", (Makridakis and Wheelwright 1989, p.9) and the prediction of future events. When those observations are presented as an ordered sequence of values, observed at equally spaced time intervals, the representation is deemed to be a "time series". In general forecasting terminology, the use of past patterns and historical values in order to determine future values, is associated with time series extrapolation.

2.3.1 Methods of Forecasting:

The methods used to record and analyse observations and derive predictions determines whether the forecasting task is judgemental (subjective) or statistical (objective). In addition, judgemental forecasters may use a number of techniques in order to try to predict future events. If observational information is recorded directly in human memory and then used intuitively for prediction purposes, the forecast is referred to as judgemental (Makridakis and Wheelwright 1989, p.9). Of particular interest to this Thesis is the use of historical patterns and relationships as a basis for predicting future values, commonly referred to as extrapolation.

Armstrong (1985, p.152) describes that extrapolation, in its judgemental form, is known as "eyeballing" and relies on the assumption that future events will conform to the available historical data. Lawrence *et al.* (1985) define "eyeballing" as the means by which to reach a cognitive understanding of the behaviour of the time series, and then using that as a base from which to predict future values. In addition, Armstrong (2001, p.790) defines judgemental extrapolation as "a subjective extension of time-series data".

Over the past twenty years there has been much debate over the value and role of judgement in the forecasting process. Research to-date has produced inconclusive findings. As maintained by Webby *et al.* (2001), the only thing that forecasting researchers can agree on, is "...that there is no single best method of forecasting."

The forecasting research area remains divided, with staunch supporters of the value and use of objective or statistical methods on one side (Ashton, 1982, Armstrong, 1985, 1986; Carbone and Gorr, 1986, Makridakis, 1988) and champions for the use of subjective methods on the other. Supporters of subjective methods for forecasting, report that many studies have shown judgemental forecasting methods to be far more widely and frequently used by practitioners (Lawrence, 1983; Mentzer and Cox, 1984; Lawrence *et al.*, 1985; Dalrymple, 1987; Taranto, 1989; Sanders and Manrodt, 1994; Goodwin, 1998). They maintain that in practice, forecasting is typically performed through the exercise of human judgement. Given the increase in sophistication of forecasting aids, in particular for objective methods, this remains an interesting finding.

More recently, a review of judgemental time-series forecasting by Webby, O'Connor and Lawrence (2001), highlighted a particular set of circumstances under which judgemental forecasting was deemed to be the most appropriate method of forecasting. The focus of their study was to "...examine when and how judgement should be used in time-series forecasting..." also, to determine "...the contribution of judgement to the forecasting process...and the way in which judgemental knowledge can be used effectively". Findings from their review identified the importance of domain knowledge in the judgemental forecasting process.

Webby *et al.* (2001) define domain knowledge as being "an attribute of the forecaster..." basically, knowledge that is possessed by the forecaster, knowledge that the forecaster has about the task at hand or "the variable of interest". They go to great lengths to present a distinction between the use of domain knowledge and that of contextual information. Webby *et al.* (2001) stress that "domain knowledge is the result of applying human interpretation to contextual (or environmental) information" and that "contextual information may not always produce corresponding domain knowledge" (Contextual information will be discussed further in Section 2.3.3.4 addressing Time Series Characteristics). Webby *et al.* (2001) conclude that where domain knowledge is deemed to be crucial in the forecasting process, judgemental forecasting is the preferred method, producing the most accurate results.

When reviewing the impact or the importance of domain knowledge on the judgemental forecasting task, it may often be easier to view the task as a two-stage process. The first stage consisting of time-series extrapolation in its pure form, or "eyeballing", as it is commonly referred to; and the second stage, being the refinement of the prediced forecast through the application of domain knowledge. For the purpose of this Thesis, the primary focus will be the first stage of the judgemental forecasting task – the task of eyeballing, relying on the assumption that future events will conform to available historical data.

2.3.2 Judgemental Forecasting – Task Characteristics:

A comprehensive, multi-factor review of the literature in the area of judgemental versus objective time series forecasting was undertaken by Webby and O'Connor in 1996. One of the major objectives of Webby and O'Connor's study was to "determine whether the effectiveness of subjective versus objective methods is dependent upon task, human and environmental differences".

The study reviewed the impact of a number of both human and task characteristics on the accuracy of forecasts. The task was time series forecasting and typically characteristics related to this task included the following; trend, seasonality, noise, instability, the number of historical data points provided, the number of periods to be predicted, length of forecast horizon, feedback and presentation type – graphical versus tabular. Experience and motivation were among the human characteristics that were examined. In addition, the provision or omission of contextual information was also examined and was deemed to be an "environmental difference".

Webby and O'Connor's study identified a number of factors which had an effect (generally adverse) on judgemental ability. These included trend, instability, number of historical data points and length of forecast horizon. The factors of interest to this study will be defined and discussed later in this section. Overall, general findings supported those of Sanders and Manrodt's 1994 study and reported that "subjective techniques generally represent about 40-50% of the total techniques used in time series forecasting". Again, mention is made of the fact that given the increase in technology and the emergence of sophisticated computer supported forecasting aids; this was an interesting, although perhaps, an unexpected finding.

2.3.3 Time Series Characteristics:

The following sections will provide definitions for the following time series characteristics deemed to be relevant to this Thesis; trend, seasonality, time horizon, contextual information, and instability and noise.

2.3.3.1 Trend:

In a time series, the trend factor represents the long-term behaviour of the data. A trend pattern can therefore be identified as the general increase or decrease in the value of the measured variable over a period of time. Trend can be presented as a linear relationship between the data points. This is achieved by fitting a line to the past series of values – using trend analysis, with time as the independent variable. This line may also be projected into the future and used as the forecast.

According to Andreassen and Kraus (1990), there is a general tendency to continue a trend, in turn, this may be enhanced depending on the length and consistency of the trend. Webby and O'Connor's 1996 review of the literature report that judgemental forecasts of trended series can be expected to produce reasonable results. In terms of accuracy, there are conflicting results based on the "nature" of the trend, i.e. whether it is an upward or a downward trend. However, in general, forecast accuracy for downward sloping trends has been found to be inferior to upward sloping trends (Lawrence and Makridakis, 1989, O'Connor *et al.*, 1993).

This study will incorporate trend as a factor in the experimental design, specifically the use of graphical presentation of trended time series data. The rationale for the selection of graphical presentation will be discussed at length in Section 2.3.5. However, it is in keeping with the general principle of judgemental forecast accuracy which maintains that "judgemental forecasts based on trended series presented graphically are much less biased" (Harvey 2001, p. 64).

2.3.3.2 Seasonality:

Davey and Flores 1993, define seasonality as a "structured pattern of changes within a year". In general, it is defined as a pattern which repeats itself over a series of fixed intervals in time. Typically these may be either related to the day of the week, or may be monthly, quarterly, half-yearly or yearly. For example, sales of ice cream, soft drinks or heating fuels typically exhibit a seasonal pattern.

Variations caused by seasonal fluctuations within a data set can often cause substantial fluctuations in the time series data patterns. This can be supported by findings from Lawrence *et al.'s* 1985 study which showed that judgemental forecasts of seasonal time series produced less accurate results than those for non-seasonal time series. In addition, Webby 1993, found that seasonality generally tended to increase the forecast errors of human judges.

In order to make sense of the data and the related data patterns, it is often necessary to identify the seasonal component within the data set. Once identified, it can be analysed and interpreted separately in order to enable the forecaster to make an accurate value judgement about the likely occurrence of future events.

This study will incorporate seasonality as a factor within the experimental design.

2.3.3.3 Time Horizon:

The time horizon refers to the period of time over which a decision will have an impact. An important aspect is the number of periods for which a forecast is desired as well as the number of periods or data points which have been provided.

The tendency for forecasting accuracy to decrease as the forecast horizon increases is widely reported in the forecasting literature (Dalrymple and King, 1981; Armstrong, 1985). This has been shown to occur in both judgemental and statistical forecasting (Lawrence *et al.*, 1985; O'Connor, Remus and Griggs, 1993).

The current study will aim to determine the extent, if any, of differences between the accuracy of predictions across the time horizon. In order to achieve this, average MAPEs for groups of close-in forecast periods (periods 1-6), outer periods (7 to 12), as well as all twelve periods will be derived. This is in keeping with previous research in the area. The horizon MAPE grouping is identical to that used by Edmundson and Terry (1986) when they compared various approaches for forecasting similar time series to those in this Thesis.

2.3.3.4 Contextual Information:

Webby *et al.* (2001) broadly define contextual information as an "attribute" of the general forecasting environment, adding an extra dimension to Webby and O'Connor's (1996) definition of – "information, other than the time series and general experience, which helps in the explanation, interpretation and anticipation of time series behaviour". A definition of contextual information is also provided in Armstrong (2001), here it is deemed to be "information about explanatory factors that could affect a time-series forecast" (p. 774).

O'Connor and Lawrence (1998, p.71) report that the results of a large number of forecasting studies emphasise the importance of the subjects' knowledge of non-time-series or contextual information. This is further supported by findings from Lawrence *et al.'s* (2000) study, which emphasised the importance of contextual information, specifically its relevance to the task at hand.

However, when designing an experimental task there remains the major problem of exactly how much contextual information to provide. Handzic's research (1997) shows that when subjects are provided with multiple pieces of causal non-time series information, they under-utilise the full-potential of that information. Also, Connolly and Serre (1984) describe that forecasting subjects often have trouble assessing the relative validity of such information.

Rich information environments generate considerable diversity between subjects due to subjects' varying abilities to both utilise that information and cope with information overload. However, minimising the amount of contextual information made available is also problematic as it has the effect of unduly emphasising that which is given – as shown by Armstrong (1983). He presented people with essentially the same graphs, but with different labels and found that people made different forecasts depending on the labels presented to them (one group was presented with a graph labelled "US production of automobiles", the other Production of Product X in Transylvania").

Goodwin (1998) identifies that even the seemingly innocent act of labelling a graph "Sales" can have a profound effect. The possibility of different reactions to such labels by individual subjects is supported by the observation that, in the context of advertising sales, executives have a natural tendency to 'underbudget' (Ashton, 1984). This is consistent with Whitecotton's (1996) conclusion that "real-world experience" can contribute to forecasting bias, and generally in keeping with Tversky and Kahneman's 1981 findings that "choices involving gains are often risk averse and choices involving losses are often risk taking". In addition, these findings further add weight to the importance of task context in the area of cognitive decision making and support Payne's 1982 studies which identified that when faced with different task contexts, people make different decisions. Given the above, contextual information could therefore be seen to significantly contribute to any between-subject accuracy variations in an experimental setting. In many of the judgemental forecasting studies (such as Lawrence *et al.,* 1985) the focus has been to conceal the context of the time series. The purpose of this deliberate action was to enable forecasters to concentrate solely on the task at hand – judgemental time series forecasting – without disruption or interference from external or domain factors. In addition, this action was also seen as instrumental in trying to achieve equivalence with quantitative methods.

The high level of internal control allowed researchers to concentrate solely on the effects of time series characteristics and, their impact on judgemental performance.

Other studies have also found that, in the absence of contextual information, novices can perform as well as expert forecasters when undertaking a basic judgemental forecasting task such as "eyeball" extrapolation (Edmundson, Lawrence and O'Connor, 1988; Sanders and Ritzman, 1992).

Given the above findings, and the view presented earlier, that the task of judgemental forecasting may be able to be viewed as a two-stage process, this Thesis will focus on the first stage of the judgemental forecasting task, commonly referred to as judgemental eyeballing. The experimental design for this Thesis will not provide subjects with any contextual information. The aim of the research will be to concentrate on the effects of the time series characteristics and, their impact on judgemental forecast accuracy.

2.3.3.5 Instability:

The presence of a discontinuity or temporal disturbance in a time series is referred to as "an instability" (Webby and O'Connor, 1996). It can also be considered as a change occurring in the time-series pattern which can't easily be directly attributed to any particular cause.

In general, judgemental forecast accuracy is found to be diminished when the time series displays a fairly high level of instability. In particular, this is so in the absence of contextual information, which may aid the user in explaining any reasons for the disturbance in the time series pattern/s.

O'Connor *et al.,* 1993, identified another interesting factor when they reviewed the impact of instability on the accuracy of judgemental extrapolation.
In general, instead of examining the time series as a whole, subjects tended to concentrate on individual movements between the periods. It appeared that subjects had a tendency to "over-react" to the most recent data. Findings from the study indicate that prior instability in a time series affected subsequent judgemental accuracy, particularly so for those forecasters identified as being "baddies" (p. 170). In an effort to "fine tune" their predictions, these subjects changed their forecasts to a much greater extent than those identified as being "goodies", producing an overall less accurate result. Perhaps in their quest to detect a pattern upon which to anchor their judgement, the "baddies" were not able to detach themselves and review the time series as a whole. In turn, their behaviour led to an overall less accurate forecast. This type of behaviour having been observed and previously reported by Andreassen, 1988.

However, another factor that must be noted here is the use of computer-supported media in the O'Connor *et al.* (1993) study. As the researchers themselves commented, perhaps the technology took some of the subject's attention away from the overall behaviour of the time series (p. 171). This is one of the key areas of focus of this Thesis.

2.3.3.6 Noise:

Noise is that element of a time series that can be defined as "randomness", sometimes also referred to as "white noise". It is believed that that "people have a poor conception of randomness; they don't recognise it when they see it and cannot produce it when they try." (Slovic *et al.,* 1974, p. 192).

Studies in the area of judgemental extrapolation, have found that the presence of randomness in a time series can often render traditional methods and strategies to be less effective, resulting in less accurate predictions (Goodwin and Wright 1993; O'Connor *et al.,* 1993; Harvey *et al.,* 1997). In an effort to identify the underlying pattern in the time series and thus provide an "anchor" for pattern extraction and future predictions, people often experience difficulty in separating or distinguishing between the random and non-random components of the time-series. When faced with increased levels of noise in the time series, people have tended to overreact to the most recent data – such as the value of the last observation. This has long been established by Kahneman and Tversky (1973) who identified that in general, when making predictions, there was a tendency for people to place too much weight on the most "salient" information (often the most recent), and not enough on the less "salient", such as long-term trends or patterns, which was often the more important information source.

Instead of reviewing the time series as a whole and trying to determine the underlying pattern, people have a tendency to place too much focus on the movements of the series from one period to the next. As O'Connor *et al.,* 1993 concluded, in general, people have difficulty "appreciating the concept of randomness and its influence on behaviour" (p. 171).

Findings from Webby and O'Connor's 1996 review support the notion that the presence of random noise has a detrimental effect on forecast accuracy when judgemental extrapolation is the method of choice (Sanders 1992; Lawrence and O'Connor 1992; O'Connor *et al.,* 1993). Andreassen and Kraus's experimental studies (1990) support the notion that increased noise reduces people's ability to perceive and interpret patterns in the time series. In turn, this impairs their ability to make accurate judgements when using extrapolative forecasting methods (Harvey *et al.,* 1997).

2.3.4 Decomposition of Time-Series Data:

In time series forecasting, decomposition is the term used to denote the separation of the data into several unique components or subpatterns – trend, seasonality, noise and cycle. Decomposition then enables each of these components to be projected into the future. In general, decomposition methods have been deemed to be the most direct and explicit in identifying and isolating the trend and seasonal components within a set of data.

It has often been said that the basic principle behind decomposition, is the general notion of "divide and conquer" (Raiffa, 1968). By decomposing a task into several smaller components or subsets, it allows the subject to concentrate on one thing at a time. Findings from MacGregor and Armstrong's 1994 study support this notion, they reported that as a decision-analytic tool, decomposition enables a complex task to be reduced into smaller and "cognitively more manageable parts" (p. 32).

Thus by definition, decomposition may be seen to be decreasing the complexity of that task and in turn decreasing overall cognitive load. This has also been suggested by Webby and O'Connor in their 1996 review of the literature (p. 106) and was one of the major focus points for Vessey's 1991 research, a review of presentation format in relation to decision-making performance.

Vessey theory based analysis was centered on the notion of "cognitive fit", maintaining that "problem solving with cognitive fit results in increased problem-solving efficiency and effectiveness" (p. 220). She essentially presented a review of the graphs versus tables literature and sought to explain performance outcome in a decision-making environment, in terms of "cognitive fit". Citing Newell and Simon's 1970's work in the area of human information processing, Vessey purports that "since humans are limited information processors, more effective problem solving will result when the complexity in the task environment is reduced" (p. 220).

Vessey's work concentrated on decision-making tasks, both simple and complex tasks. Identifying tasks involving judgement or inference as being "complex tasks", ones which can, in turn, be further decomposed into several subtasks (p. 225). Vessey seeks to develop a theory to support this and proposes the notion that task complexity can be reduced by both;

(a) decomposing a task into several sub-components, and
(b) finding the best "mode of representation" for that particular task (p. 220).

Vessey refers to this concept as the "notion of cognitive fit". This is also a pivotal area of interest for this Thesis and will be addressed further in Section 2.4.1 where a review of the literature in the area of task complexity will be conducted.

In terms of decomposition and its effect on performance outcome, studies in this area continue to provide inconclusive and mixed results. This is supported by findings in Webby and O'Connor's 1996 review, they conclude that more investigation is required into the impact of decomposition on the relationship between task complexity and decision quality.

A basic premise behind decomposition, the notion of *structure*, has been the subject of many research studies. Findings from Armstrong *et al.'s* 1975 study showed that in general, the use of a decomposed approach led to improvements in people's judgement decisions. In terms of forecasting accuracy, Edmundson and Terry's 1986 study found that accuracy increased with the use of a Decision Support System. The DSS – later to be named GRAFFECT – allowed for the time series data to be decomposed and thus presented in a more structured manner. In turn, that view has been supported by MacGregor *et al.'s* 1988 findings which showed that when greater structure was imposed on the decision aid, accuracy increased. Overall, the notion that judgemental decomposition adds structure to the task, has been supported by findings as reported in Webby and O'Connor's 1996 review of the literature.

In general, it has also been noted that decomposition is especially useful when dealing with the naïve user – one who knows relatively little about the topic in question. However, findings by Edmundson (1990) showed that both novice and expert forecasters produced more accurate predictions when aided by the decomposition DSS.

Citing research evidence gathered "over four decades of research in human judgement and decision making" (p. 110), MacGregor maintains that "decomposition improves judgemental performance over holistic or unaided judgement". In a 2001 study, MacGregor found that decomposition improved judgemental forecasting accuracy in situations where task uncertainty was high. This had been previously identified by Armstrong *et al.* (1975) almost thirty years ago. Armstrong found decomposition to be particularly useful when subjects knew comparatively little about the subject matter or the task. Research findings support the notion that in order for decomposition to be an effective strategy, one must first identify the "level" of uncertainty at which the use of decomposition becomes appropriate (MacGregor and Armstrong 1984). In his summary, MacGregor (2001) deemed decomposition to be "... an effective strategy for improving the quality of judgemental forecasts" (p. 121), particularly for tasks with high uncertainty.

The current research study will seek to decompose time series representations and review the impact of structure on the accuracy of predictions in a judgemental forecasting task. In addition, it will also seek to review the influence of the *mode* of representation, i.e. the media, and thus extend the focus of prior research conducted in the area of task presentation.

2.3.5 The Impact of Presentation Mode in a JFS Task

Much research has been undertaken in an effort to determine the impact of presentation format or mode of output, on the decision making process. Generally, studies have concentrated on the relative merits of graphical versus tabular data displays (Benbasat and Schroeder, 1977; Lucas and Nielsen, 1980; Lucas, 1981; Lawrence, 1983; De Sanctis, 1984; Angus-Leppan and Fatseas, 1986; Dickson, De Sanctis and McBride, 1986; Blocher, Moffie and Zmud, 1986; Remus, 1984, 1987; Vessey, 1991; Harvey and Bolger, 1996).

It was De Sanctis (1984) who argued that more research needed to be undertaken in this area in order to identify the environmental conditions under which one presentation format produced more accurate forecasts than the other i.e. graphs versus tables. Remus (1984, 1987) reported a correlation between environment/task complexity and presentation format. He reported that given an environment of low complexity, tables produced more accurate forecasts than graphs; however, in an environment of intermediate complexity, graphics produced superior results.

Dickson *et al.* (1986) also reported a task-related effect, concluding that "for a task activity that involves seeing time dependent patterns in a large amount of data, graphs are a good choice of format." (p. 46). However, to date, no proven theoretical framework exists to support these findings.

In general, results from studies in this area have proven to be inconclusive, with Remus (1987, p. 1200) declaring that the empirical literature in this area remains "in disarray". However, what must be noted at this stage was that there was no theoretical framework grounding Remus' 1987 empirical study. In addition, his literature review citing "... at least nine conflicting studies on the relative impact of graphical and tabular displays on managerial decision making" (p. 1200) was not presented or categorised in terms of display *media*; that is hard-copy or *paper*, and CRT or *screen-based*. Results from studies in the broad area of graphs versus tables literature included studies conducted across several different types of display *media*. Remus' study did not seek to categorise and present findings according to display media or media type. Perhaps such categorisation may have presented a more streamlined set of overall results. Indeed this is one of the main areas of interest to this Thesis and will be expanded upon later in section 2.4.2.

Given the lack of rigour in Remus' review, it is therefore surprising to find him declaring that the empirical literature in this area remains "in disarray" (p. 1200). However, to-date it appears as though the question of optimum presentation format remains largely unanswered, with researchers still debating the relative merits of one presentation format over the other (Harvey and Bolger, 1996, p. 134).

Findings from Lawrence's (1983) study supported the view that graphical presentation of time-series data was better for short-term forecasting and produced more accurate – although not significantly so – results. However, given the small sample size of this study it is difficult to generalise these results. Dickson *et al.'s* (1986) study concluded that forecast accuracy in the graphical presentation format was significantly better than in the tabular format (for eight of the nine forecasts). However, given that only three time series were used, and that all three were trended, it is once again difficult to generalise these results. This is also supported by Dickson *et al.'s* general conclusion that "more refined examination ... is a subject for future research" (p. 47).

Angus-Leppan and Fatseas' (1986) study reported similar findings to those of Lawrence (1983) and Dickson *et al.* (1986). Their findings supporting the notion that graphical presentation produced more accurate results for short-term forecasts (MAPE was significantly less for graphical than tabular predictions for periods 1-6; no significant difference was reported for periods 7-12). However, there were also limitations with Angus-Leppan and Fatseas' study. The major one being that the same subjects were used to predict forecast accuracy for each of the treatment conditions. The subjects were first asked to make extrapolative estimates using the table format. They were then asked to repeat the exercise using the graph format. It can be argued that given their previous exposure to the data in table format, they were already familiar with the nature of the time series data by the time they were asked to make extrapolative estimates in the graph format. Given these limitations, the generalisability of the results of the Angus-Leppan and Fatseas (1986) study is questionable.

In terms of time series characteristics, there has been much research conducted to determine the impact, if any, of one presentation format compared to the other (graphs versus tables). Past studies have shown the influence of trend on the accuracy of time series extrapolation. Of particular interest are the findings reported by Harvey and Bolger's 1996 study, that graphical presentation of data for linearly trended series enabled people to produce more accurate judgemental forecasts. These results further supporting those of Dickson *et al.*, 1986. They identified that when using judgement to forecast linearly trended series, people were found to perform better when the data was presented in graphical format. In general, people's tendency to underestimate trend was less prominent than when they were presented with a tabular format.

However, the major difference between Harvey and Bolger's (1996) study and previous studies in this area, was the "rigour" of the experimental design, thus enabling generalization of the results. As identified by Harvey and Bolger, many of the previous studies had failed in this area. In an attempt to rectify the situation, Harvey and Bolger incorporated the following into their experimental design; balanced within-subject design, half the subjects received graphical presentation first and half received tabular presentation first. Experimental conditions differed only in terms of presentation format – identical series were presented in both table and graphical format; untrended as well as trended series were studied. Subjects were presented with 44 different 20-point time series and were asked to predict the 21^{st} and 22^{nd} points of each one. The task took approximately 40 minutes to complete.

In recognising the "rigour" of Harvey and Bolger's (1996) experimental design, several questions also arise. Subjects were only asked to predict short term forecasts, this was in keeping with Lawrence's 1983 findings. However, one question that arises is the validity of the forecast "horizon". One could question whether the prediction of *only two* forecast periods by each subject, for each of the time series, is enough data upon which to judge subject's performance in terms of accuracy. Another issue is the use of the root mean square error (RMSE). This metric has been reported to be an unreliable and inappropriate measure for comparing accuracy across time series (Armstrong and Collopy, 1992; Chatfield, 1988; Fildes, 1992). Although Harvey and Bolger recognise this as a limitation of their design, they still maintain it to be a useful measure in comparing the effects of different presentation formats on forecast accuracy. Finally, given the use of 44 different time series, presented to each subject in a task of 40 minutes duration, the issue of subject fatigue must also be raised. It would be interesting to note if subjects' forecast accuracy was significantly different across time series – based on order of presentation. Harvey and Bolger's results do not address these issues, thereby perhaps compromising the generalisability of their experimental findings.

A very interesting aspect of Harvey and Bolger's study is their assertion that in order to ensure generalisability of overall findings, research on the effects of presentation format should embrace the theoretical aspects of this area. They declare that to-date, research in this area "has been empirically rather than theoretically driven" (p.122). In addition, they quote that something that would be extremely useful, and something that is missing in the research area, is an understanding of the underlying "cognitive processes" of the effects of presentation format (Harvey and Bolger 1996 p. 122 - quoting Ganzach, 1993). In general, Harvey and Bolger maintain that:

> "Experiments should be designed and analysed to cast some light on the cognitive processes responsible for any effect of presentation format" (p. 122).

This is also one of the major aspects of the focus of this Thesis – to propose a theoretical model, and design and conduct a series of experiments to help further research in this area. In fact, to further extend the literature by not only concentrating on presentation format (or *mode*) – graphs versus tables, but also incorporating the presentation media – screen-based versus hard-copy or paper.

A number of major studies have concentrated on the effect of presentation format on judgemental time series extrapolation (Lawrence *et al.*, 1985; O'Connor, 1993). Most of these studies concentrated initially on the comparison of judgemental versus statistical techniques, but also addressed the issue of graphs versus tables

Findings from Lawrence *et al.* (1985) indicate that the use of graphical display formats produced better results for judgemental forecasting under the following conditions; accuracy of predictions for non-seasonal time series was greater than those for seasonal time series and accuracy of predictions for short-term forecasts outperformed those for long-term forecasts. In contrast, O'Connor *et al.* (1993) found that judgemental forecasting produced significantly worse results than statistical when a graphical display format was used. The researchers qualify their results by identifying that the use of different media across the two studies may have had a significant effect on the results. Lawrence *et al.* (1985), used the traditional media; paper-and-pencil; whilst O'Connor *et al.* (1993) adopted a more modern approach, using a computer screen and mouse-based input. O'Connor *et al.* (1993) maintain that "perhaps the technology took some of their attention away from the overall behaviour of the time series" and observe that this is indeed an area for future study. When analysing their results, O'Connor *et al.* (1993) identified a "bizarre conclusion, that the people who performed best in the task spent the least at it... there was no association between the time taken and the accuracy of the forecasts". This is also an area for future study, expanding the area from presentation *media* and *mode* to include the factor of *time*.

Findings from Harvey and Bolger's 1996 study, which incorporated a computer-based presentation media and both graphical and tabular presentation modes, left them with the following question. They disputed whether the use of different *media*, that is, the use of computer-based displays, as opposed to hard-copy display, for both presentation *modes* would have produced different results. Harvey and Bolger (p. 134) maintain that the following question remains unanswered; "would our results generalize to forecasters using pencil and paper?" Or put another way, would the *media* have made a difference to the results? Harvey and Bolger further maintain that the literature in this latter area – "the effects of computer-based versus hard-copy display" – appears to be in just as much "disarray" as the literature on graphical versus tabular presentation format (p. 134).

Questions posed from the findings of Dickson *et al.'s* 1986 study, namely that the "... effectiveness of the data display format is largely a function of the characteristics of the *task* at hand ..." remain largely unanswered (p. 40, emphasis added). In their concluding comments, Dickson *et al.* maintain that "more refined examination of the dimensions of task environment and their relationship to the use of graphics" is needed (p.47).

This is a research area in which to-date, empirical results remain unclear. The overall question still remains; how to best identify the optimum display *media* and *mode* for a given purpose (Meyer *et al.,* 1997)? That is, to identify the best display *media* and *mode* in relation to any particular *task*. This area is the object of interest for this Thesis. The research question being;

> Does the presentation *media* as well as the *mode* have any impact on *task accuracy* in judgemental forecasting?

Of interest to this study is the potential impact and the direct nature or effect of any such impact. In addition, if an impact is found to exist, this study will seek to isolate and identify the relevant variables. Findings from this study aim to aid in the development of decision support systems utilizing graphical, computer-based displays to assist with the judgemental forecasting task in a commercial environment.

The next section will present a review of the literature in relation to Task. Of particular interest is the area of task complexity. We will then seek to relate findings from the review of task literature to the *object of interest* of this Thesis – the judgemental extrapolation of time series data. Our aim is to provide a classification of the complexity of the proposed experimental task.

2.4 Research Model of TTF for Judgemental Time Series Forecasting

This section will propose a Research Model of TTF for the experimental task to be undertaken as part of this Thesis and will outline the relationship between each of the variables of interest.

The TTF model as proposed by Goodhue and Thompson (1995) has been used as the basis for the proposed Research Model for the judgemental extrapolation of time series data. According to the researchers:

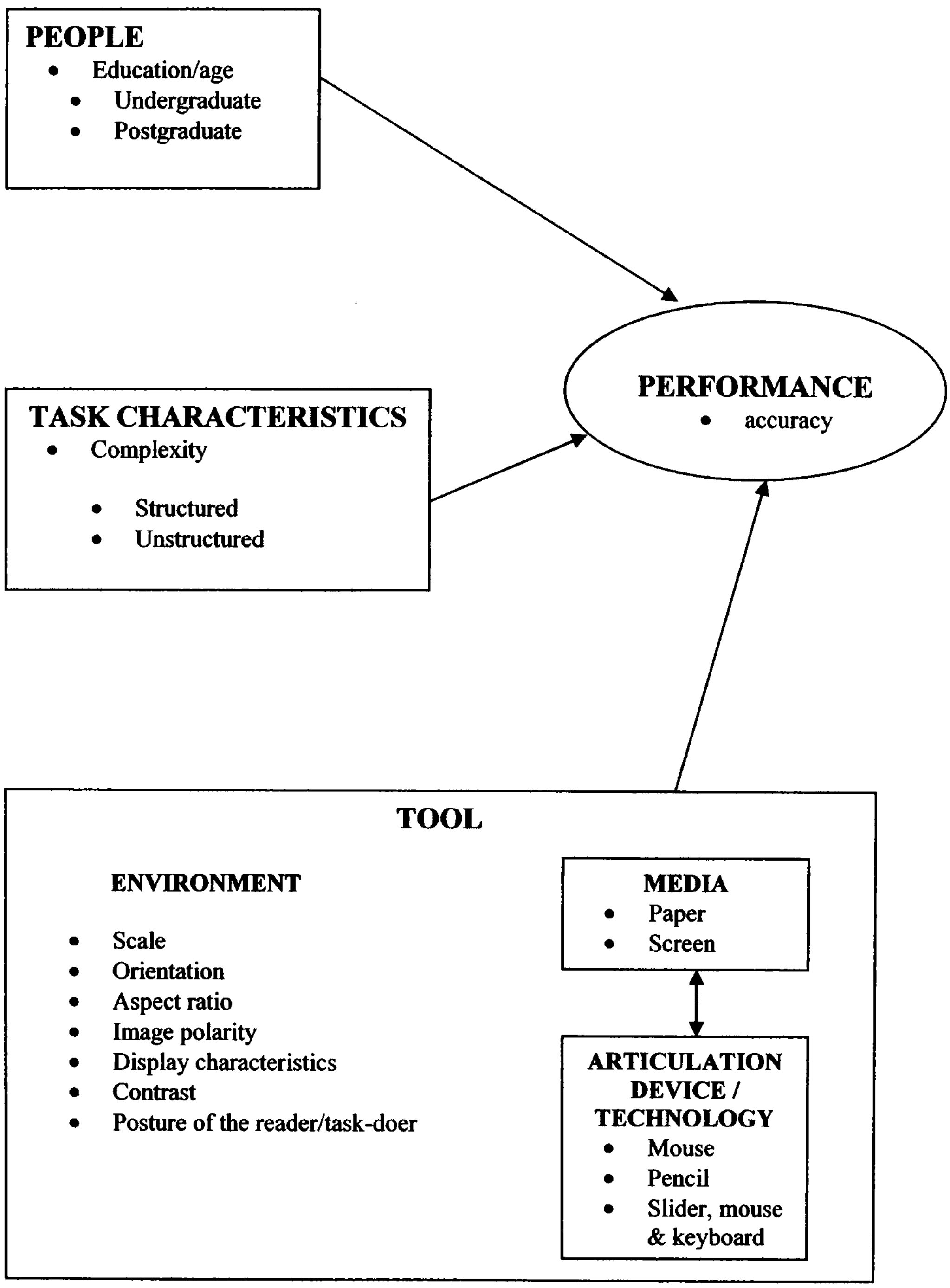
> "Task-Technology Fit (TTF) is the degree to which a technology assists an individual in performing his or her portfolio of tasks. More specifically, TTF is the correspondence between task requirements, individual abilities, and the functionality of the technology." (p. 218).

The basic premise behind the proposed theoretical model and experimental studies is to seek to identify a link between the components of the Research Model namely, the Tool – the *media* and accompanying *technology*, the People undertaking the task, and Task Complexity. The level of TTF will be measured in terms of task performance, or more precisely, the accuracy of predictions in the judgemental extrapolation of time series data.

A review of the literature in relation to Task-Technology Fit has already been conducted in section 2.2. This identified Goodhue and Thompson's 1995 model of TTF as being fundamental to this Thesis. The judgemental forecasting literature has been reviewed in section 2.3 and that in relation to Task and Media studies will be reviewed in the next two sections (2.5 and 2.6). This will provide the basis for tailoring the experimental studies to be conducted as part of this Thesis.

The proposed research model for this Thesis is outlined in Figure 2.4.1 below:

Figure 2.4.1 Research Model
Media and Technology Effects in the Forecasting Task



2.5 Task – Review and Analysis of Related Literature
A review of Task Characteristics and Task Complexity

2.5.1 Introduction

The major focus of this Thesis is the effect of *Media* and the related *Technology* on the accuracy of predictions in a judgemental forecasting task. In section 2.3 we identified the basis for selecting judgemental time series extrapolation as the *object of interest* or experimental task, for this study. A review of the literature in relation to Task-Technology Fit (TTF) was conducted in section 2.2 and was highlighted as the basis for the proposed Research Model as presented in section 2.4. We must now look to the literature and previous research in relation to Task itself. This will enable us to review the theoretical and empirical studies undertaken in the area of Task identification and definition. The performance of this review shall seek to identify a suitable taxonomy or theoretical model of tasks and their related attributes – one which can then be used to classify our experimental task.

Fleishman and Quaintance (1984) have defined a task as a set of conditions that elicit specific activities or processes. Others broadly define tasks as being related to the actions carried out by individuals when turning inputs into outputs, specifically when trying to satisfy their information needs (Goodhue and Thomson, 1995; D'Ambra and Rice, 2001). However, to date, there appears to be no adequate theoretical model or taxonomy to describe a task or to explain how tasks differ from one another (Ferguson, 1956; Hackman, 1969: Weick, 1965).

For the purpose of this Thesis, a task refers to the actual work being performed by the subject. The outcome or completion of the task results in the identification of an objective measure. In turn, this can be used to judge overall accuracy or performance between two or more treatment conditions. The task being undertaken in this study is a judgemental time series forecasting task, specifically the judgemental extrapolation of time series data. This task has been selected as the *object of interest* for the current study as it not only also serves to provide an objective measure – the accuracy of predictions – but also represents a task that is deemed to be relevant in a decision making context.

2.5.2 Task Complexity – Theory and Definition

The literature in relation to task complexity is of primary interest to this Section. It is important at the outset, to make a distinction between task complexity and task difficulty. It is often noted that the two terms are used interchangeably, when in fact they actually refer to two different things (Jacko and Ward, 1996).

Shaw (1976) has defined the difficulty of a task in terms of the amount of effort required to complete the task. Jacko and Ward (1996) acknowledge that a task may be deemed to be difficult, without necessarily being seen to be complex, because of the sheer amount of effort required by the person in order to undertake and complete the task. Others (Campbell, 1988; Locke, Shaw and Saari and Latham, 1981) propose that a task may be perceived to be difficult or complex, because the task actually requires the person to possess a high level of skill or knowledge in order to undertake and complete the task. According to Campbell (1988), task difficulty and complexity are in fact two different things which can be represented and measured objectively. Campbell proposed the notion that task *difficulty* is an attribute that represents a *person x task interaction*, whereas task *complexity* can be viewed as an *independent measure*.

> "A task of specified complexity may be difficult for one person, but not for another. For example, flying an aeroplane is easy for a veteran pilot although a new student may find it difficult – even though the objective characteristics of the task are identical" (p.45).

Campbell maintains that within the literature, the notion of task complexity is not well understood. He further advises that task complexity has generally been seen and treated as being either one of three things (p.40):

- an interaction between task and person characteristics;
- primarily a psychological experience; or,
- as a function of objective task characteristics.

Campbell's 1988 review and analysis of task complexity, provides a framework by identifying those qualities which make a task complex. Campbell's work is greatly influenced by the work of Schroder, Driver and Streufert (1967); they identified three primary aspects of a complex task:

- *Information Load* – the number of dimensions of information requiring attention.
- *Information Diversity* – the number of alternatives associated with each dimension
- *Degree of Uncertainty Involved* – the rate of information change.

In trying to provide a definition of task complexity, Campbell looks to specific task attributes or qualities, and maintains that:

> "Task complexity can be directly related to the task attributes that increase information load, diversity or rate of change. Thus complexity, can be defined objectively, and it can be determined independently of any particular task-doer" (p. 43).

Having previously identified the individual components of time series characteristics in section 2.3.3 one can see that the use of Campbell's definition allows us to provide an "objective measure" of the complexity of our *object of interest* – the judgemental extrapolation of time series data, the experimental task for this Thesis. This association shall be explored further throughout this section of the Literature review.

Others, such as Jacko and Salvendy (1996) have tried to define task complexity in relation to mental workload. Using the Subjective Workload Assessment Technique developed for scaling air transport pilots' mental workload, they propose the following differing levels task complexity:

- *Low Complexity* – tasks are typically performed automatically.
- *Moderate Complexity* – tasks associated with uncertainty, unpredictability or unfamiliarity.
- *Highly Complex* – tasks require extensive mental effort, skilled planning, and total attention to task.

Attempting to classify the task of Judgemental Extrapolation of time series data in terms of Jacko and Salvendy's definitions, it appears that the task falls into the *Moderate Complexity* category.

Wood's 1986 research aims to present a set of constructs which could be used to describe task complexity. He maintains that to-date the majority of studies concentrating on the analysis of tasks have been based on an empirical approach. In general, task characteristics are "inductively derived and are not based on a formal definition of tasks" (p. 61). It is this formal definition which has still proven to be somewhat "elusive" in the literature to-date.

Wood further maintains that all too frequently task characteristics as identified by researchers, have tended to be confounded by the inclusion of both task and non-task elements. In particular, Wood notes the failure of researchers to make the distinction between attributes of the task and attributes of the individual performing the task. This can be likened to the distinction made by other researchers when defining a task in terms of difficulty and/or complexity (Campbell, 1988; Locke, Shaw and Saari and Latham, 1981; Shaw 1986; Jacko and Ward; 1996).

Wood proposes a theoretical model of tasks, and identifies *products*, *required acts*, and *information cues* as the three essential components. Wood's definition of task complexity is concerned with, and is expressed in terms of three different types of complexity:

- *Component Complexity* – "A direct function of the number of distinct acts that need to be executed in the performance of the task and the number of distinct information cues that must be processed in the performance of those acts" (p. 66).
- *Coordinative Complexity* – "Refers to the nature of the relationships between task inputs and task products" (p. 68).
- *Dynamic Complexity* – "Due to changes in the states of the world which have an effect on the relationships between task inputs and products" (p. 71).

Wood's theoretical model is in part based upon the work of Hackman (1969), and represents a partial combination of Hackman's "task qua task" and "task as behavioral requirements" approaches. Wood proposes that his theoretical model, allows the researcher to calculate the total complexity of a task, including the knowledge and skills required of the individual in order to perform the required task. He maintains that his theoretical model is able to achieve this by incorporating the three types of complexity – component, coordinative and dynamic – and by the presentation of formulas or indexes to allow and enable quantification of these constructs. However, to-date, there has not been widespread acceptance or adoption of Wood's 1986 theoretical model, with Wood himself acknowledging that in practice, the calculation of some of the proposed indexes may well be "beyond our present capability" (p. 81).

Whilst this literature section would not be complete without acknowledging Wood's contribution to the development of a theoretical model of tasks, it must be noted that this study will not be adopting Wood's model. Rather than express task complexity in terms of the different *types* of complexity – as Wood has – this study will seek to analyse the objective complexity of any given task by identifying the individual components or "sources of complexity" – as defined by Campbell's complexity framework.

In terms of the complexity of a Task, other researchers such as Campbell and Gingrich (1986) have attempted to define a task in terms of whether it is deemed to be a *simple* or a *complex* task. They propose that a *simple* task can be identified as one which has a small number of independent sub-parts, one which places "minimum cognitive demands on the task-doer for comprehension and execution". In contrast, they identify a *complex* task as one which may have several, often conflicting elements to satisfy, and one which places "substantial cognitive demands on the task-doer for comprehension and execution" (p. 164). Baron (1986) also provides some insight into the inherent differences between simple and complex tasks, advising that simple tasks require the processing of fewer cues than complex tasks, and are therefore less likely to challenge the cognitive capacity of the decision maker. Baron also proposes that decision makers undertaking a simple task have excess cognitive capacity, whilst those undertaking a more complex task have little, if any, excess cognitive capacity.

Jacko and Ward (1996) report that:

> "Traditionally, the complexity of a task has been expressed on a continuum ranging from simple to complex, and researchers have assumed that the underlying attributes of tasks that determine complexity are understood" (p. 534).

Jacko and Ward maintain that their examination of the relevant literature in relation to task complexity showed that, in many cases, task attributes were not clearly defined or understood. Dickson, DeSanctis and McBride (1986) previously identified a similar weakness when they reviewed the decision making literature. They report that, the literature tells us that "task" is very important to decision making, and that factors such as task content, complexity and degree of structure, all play a role in decision making. However, they advise that there is generally no agreement on definitions of these factors, or, on what methods should be used to operationalize them (p. 40).

Wood (1986) echoes the same sentiments, he believes that major problems exist in the classification and design of tasks used in research.

> "The empirical approach to the study of tasks has failed to provide definitions of task characteristics with sufficient construct validity to either reliably describe how tasks differ from one another or validly predict the effects that are due to variances in tasks" (p. 61).

A review of the relevant literature suggests that there exists a lack of consensus in the identification of characteristics or attributes which are deemed to affect task complexity. The need to identify a suitable operational definition for task complexity thus remains a challenge for researchers in this area.

Jacko and Salvendy (1996) provide a brief summary of their review of the literature (p. 1189). This has been adapted and is presented as Table 2.5.2.1. The Table represents a somewhat comprehensive list of the characteristics which may affect the complexity of a task. The characteristics that are of interest to this Thesis are those identified by Campbell (1984, 1988); Payne (1976) and Schroder, Driver and Streufert (1967). However what must be noted is that this list is in no way exhaustive.

Table 2.5.2.1 Characteristics which may affect the Complexity of a Task

Task Characteristics	Reference
Number of dimensions requiring attention, number of alternatives associated with each dimension, rate of information exchange.	Schroder, Driver and Streufert (1967)
Spatial correspondence between user communication interface and prompt for response.	Boggs and Simon (1968)
Magnitude and variation of stimulation, number of sensory modalities affected	Schwab and Cummings (1976)
Number of cues, cue intercorrelations, cue validities, function forms, cue variability.	Steinmann (1976)
Number of alternatives available, number of dimensions of information available.	Payne (1976)
Path-goal multiplicity	Terborg and Miller (1978)
Uncertainty, unpredictability, unfamiliarity	Reid, Schingledecker and Eggemeier (1981)
Number of commands necessary, use of separate storage areas, homogeneity of edits, number of strategies employed.	Boyd (1983)
Path-goal multiplicity, decision verifiability, solution multiplicity.	Campbell (1984)
Products, required acts, and information cues.	Wood (1986)
Multiple paths, multiple outcomes, conflicting interdependence among paths, uncertain or probabilistic linkages.	Campbell (1988)
Procedure-oriented versus choice-oriented tasks and their compatibility with task presentation.	Zhao (1992)

Data in the Table above presents a brief review of the Literature identifying the nature of task characteristics deemed to be relevant in the determination of task complexity. The following sections will present Campbell's typology of tasks and will seek to describe or "fit", the task of Judgemental extrapolation of time series data – the *object of interest* for the current study – using Campbell's framework.

2.5.3 Objective Task Characteristics

Vessey (1994) references the works of Campbell (1988), Fleishman (1982) and Wood (1986), when acknowledging that there is no comprehensive theory of tasks. However, she notes that it is helpful to determine the abstract characteristics of a task which are influenced by different types of representation, or *modes*, of that task (eg; graphical versus tabular). This was also referred to by Campbell (1988) when he referenced the work of Hammond (1986). Campbell advises that:

> "Hammond pointed out how the complexity of essentially similar tasks varies as a function of the task's *mode* of representation" (p. 41, emphasis added).

Campbell continues to propose that task complexity may be affected or influenced by, not only the *mode* of representation, but also by the *physical form*. This is the primary focus of this Thesis; the *media* or *physical form,* as it's referred to by Campbell. He continues to advise that when dealing with a task requiring judgement, a distinction must be made between situations where past information is made available, and those in which judgement is based primarily on memory. Campbell concludes that, taken as a whole, prior studies in the area of task complexity highlight the need for researchers to broaden the scope of their studies to include other areas of interest, such as such as short-term memory and span of attention. His suggestion is that the issue of task complexity cannot be considered in isolation of these other factors.

In essence this is the basis for the research undertaken by Jacko and Ward (1996). In their study, they sought to validate a short-term memory link between Campbell's (1988) framework – presented below – and the information processing model for computer-based psychomotor tasks as developed by Salvendy and Knight (1982). Their validated information-processing model will be presented later in this section as will an explanation and discussion of the term "psychomotor".

In order to try to encompass some of the "other" research areas as identified by Campbell, this Thesis will seek to identify the influences of abstract task characteristics. In order to seek to make a valuable contribution to the research area, this Thesis will concentrate on presentation *media* (i.e. the physical form) rather than *mode* of representation.

In seeking a suitable basis for the identification of task characteristics, we look to the framework developed by Campbell (1988). This framework involves the following four fundamental task attributes or characteristics – which will be discussed in detail later in this section:

- **Multiple Paths**
  The number of possible ways to arrive at desired outcomes.
- **Multiple Outcomes**
  The number of desired outcomes of a task.
- **Conflicting Interdependence Among Paths**
  When achieving one desired outcome conflicts with achieving another desired outcome.
- **Uncertainty or Probabilistic Linkages**
  Uncertainty of whether a particular path leads to a desired outcome or not.

Campbell's framework has been recognised as being appropriate for objectively assessing the complexity of often dissimilar tasks (Mykytyn and Green, 1992; Schütte and Jordan, 1996). This appears to apply particularly to tasks from the broad areas of information systems and behavioural decision making, Umanath (1994).

Campbell introduces a "typology of complex tasks", used to categorise a task by identifying the fundamental task characteristics which contribute to its complexity. He maintains that in order to analyse the objective characteristics of a task, we must first identify the individual components or "sources of complexity".

1. ***Multiple Paths*** – "An increase in the number of possible ways to arrive at desired outcome increases information load, and thus increases complexity" (p. 43)

2. ***Multiple Outcomes*** – "As the number of desired outcomes of a task increases, complexity also increases" (p. 43)

3. ***Conflicting Interdependence Among Paths*** – "If achieving one desired outcome conflicts with achieving another desired outcome, complexity will increase" (p.44)

4. ***Uncertainty or Probabilistic Linkages*** – "Uncertainty can increase complexity by enlarging the pool of potential paths to a desired outcome" (p. 45)

Campbell then groups these task categories into four general task classifications which he labels; decision (eg; employee selection); judgement (eg; stock market analysis); problem (eg; scheduling); and fuzzy tasks (eg; business ventures).

He further advises that complexity is determined by both the *number* of attributes or sources possessed by the task and the *degree* to which a task incorporates each of the individual attributes or sources. This can be presented in the following tabular format:

Table 2.5.3.1 Campbell's Complex Task Classifications: *

Task Type	Paths Source 1	Outcomes Source 2	Interdependencies Source 3	Uncertainty Source 4
Decision	—	X	X	X
Judgement	—	—	X	X
Problem	X	—	X	X
Fuzzy Tasks	X	X	X	X

* (adapted from Campbell (1988) Table 2 p.46 and Table 3 p.47)

On the face of it, Campbell's use of paths and outcomes to define his attributes implies that the framework is most appropriate for the complexity analysis of problem-solving tasks. However, Campbell describes that an outcome may be a desired end-state or, alternatively, a task dimension requiring attention and which entails a separate information processing stream. In decision-making tasks, these information processing streams are consistent with the handling of particular groups of cues by the decision-maker. According to Speier, Valacich and Vessey (1997), and Baron (1986) this forms a quantitative basis for decision-making task complexity.

Campbell's framework is based, in part, on the earlier work of Schroder, Driver and Streufert (1967). Schroder *et al.* identified the number of dimensions of information requiring attention, the number of alternatives associated with each dimension, and the rate of information change as primary aspects of a complex task. The association of outcomes with information dimensions suggests that Campbell's paths are related to Schroder *et al.'s* dimension alternatives. This is consistent with a generalisation of Jacko and Salvendy's (1996) description that, for menu retrieval tasks, a path is the successive choices made while traversing menus.

Using the above interpretation of outcomes and paths, Campbell's "probabilistic linkages" represent the complexity associated with establishing which information dimensions and which individual informational elements should be utilised when undertaking a particular task.

Jacko and Ward (1996) describe this framework as "one of the most comprehensive attempts to integrate different conceptualizations of the complexity of a task into a concise representation" p. 534. The primary aim of their 1996 study was to validate a short-term memory link between Campbell's (1988) framework and the information processing model for computer-based psychomotor tasks as developed by Salvendy and Knight (1982). Their validated information processing model is presented here as Figure 2.5.3.1. It seeks to provide some insight into task performance, given a certain level of task complexity – measured by the presence and / or absence of, or combination of, certain identified task characteristics (as per Campbell's 1988 framework).

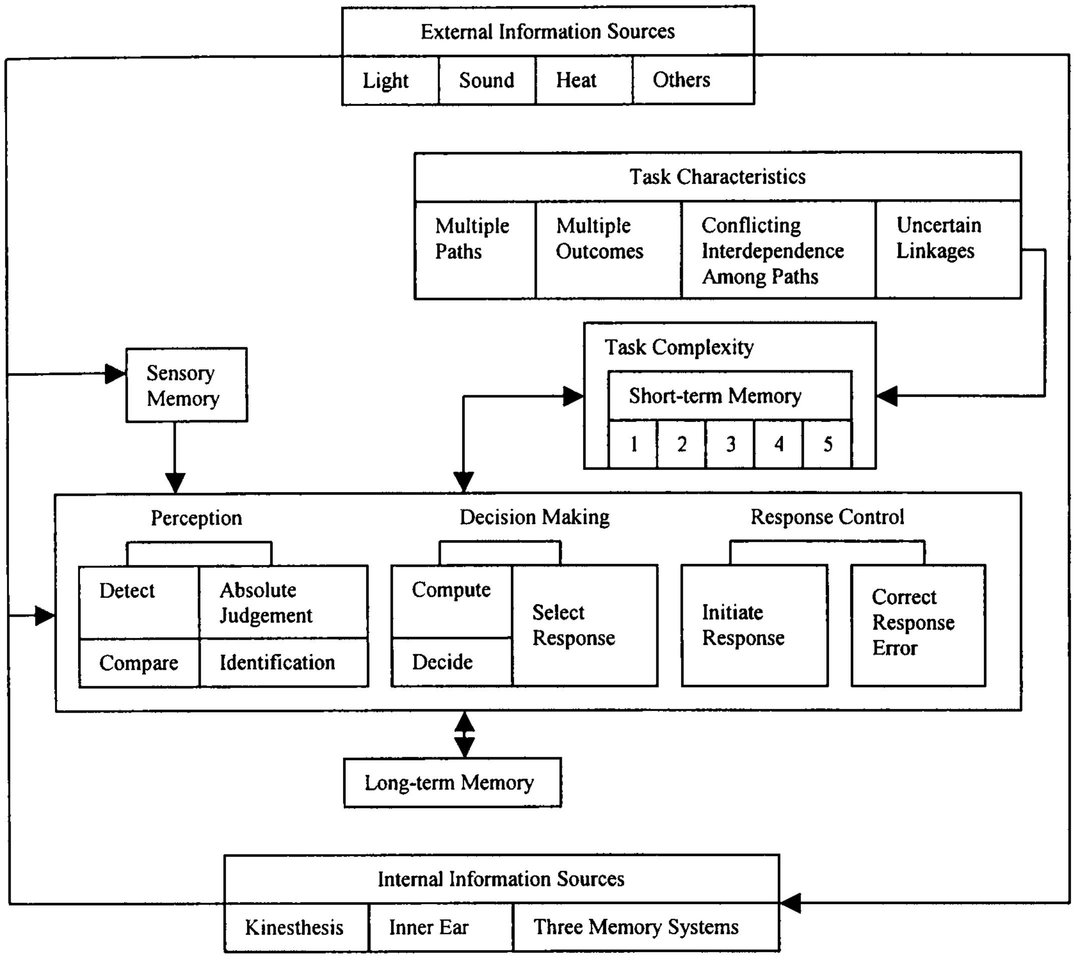


Figure 2.5.3.1 Validated Information Processing Model (modified from Salvendy & Knight, 1982)

Jacko, J.A. and Ward K.G, 1996 p. 536

Jacko and Ward's study sought to examine the relationship between task performance and the absence and / or presence of, complexity characteristics as identified by Campbell. Their object of interest was psychomotor tasks, or tasks which require the person to use controlled movements involving their body, or some instrument or device as an extension of their body, in order to accomplish the task. Their findings identified that an increase in the presence of objective task characteristics – as identified by Campbell – could be equated to an increase in information load. In turn, this could challenge task-doer's short-term memory given the additional burden requiring them to attend to more and more things. Jacko and Ward propose that this increased information load could be directly translated to represent an increase in task complexity (p. 535).

Having provided an interpretation of Campbell's Task Complexity framework and having incorporated Jacko and Ward's validation of Salvendy and Knight's information processing model, the next section will now seek to provide a "fit", for the task of judgemental extrapolation of time series data – the *object of interest* for the current study.

2.5.4 Complexity Characteristics of the Task – Judgemental Extrapolation of time series data

In order to identify the individual source/s of complexity inherent in this study's task, the presence and / or absence of each of the four individual components – as identified in Campbell's framework – will be examined in detail. In addition, this study will seek to provide a "fit" or categorisation for the task of Judgemental extrapolation of time series data. Each of Campbell's four components or "sources" of complexity will be examined in detail and will be deemed to be either present or absent in our chosen task.

1. ***Multiple Paths***

According to Campbell's definition, the presence of *multiple paths* can lead to an increase in task complexity. "An increase in the number of possible ways to arrive at desired outcome increases information load, and thus increases complexity" (p. 43).

Our experimental task of judgemental extrapolation was previously defined as "a subjective extension of time-series data" (Armstrong, 2001, p. 790). In particular, it was noted that in time series extrapolation, task-doers make extensive use of historical patterns and relationships as the basis for predicting future values. Given this definition and brief description of the nature of task-doer's activity, the notion of multiple paths does not appear to be of relevance to this task. This classification is in keeping with Campbell's findings as presented in Table 2 (p.46). In his discussion on the classification of "Judgement Tasks", he also noted that "multiple paths are not relevant to tasks in this category" (p. 47).

2. *Multiple Outcomes*

Campbell reports that "As the number of desired outcomes of a task increases, complexity also increases" (p. 43).

In terms of task outcomes, the proposed study will seek to determine the effects of *media* and related *technology* on task performance – as measured by a single variable, accuracy. Given that task-doers will be advised that the accuracy of their predictions will be the only variable that will be captured and the only indicator which will be used in order to judge performance, there appear to be no multiple outcomes for our experimental task.

In contrast, had task-doers been advised that both the time taken to complete the task as well as their level of accuracy would be measured; this in turn would identify a potential conflict in terms of desired *outcomes*, namely, a speed/accuracy trade off. In an effort to complete the task in the most efficient and timely manner, whilst still achieving a high degree of accuracy, the desire to achieve multiple outcomes would effectively increase task complexity. However, given that we will only be measuring the accuracy of predictions, the issue of *multiple outcomes* as a source of complexity would not appear to be relevant to this Thesis. Once again, this is consistent with Campbell's findings as presented in Table 2 (p. 46). In his discussion on the classification of "Judgement Tasks", he also noted that "multiple desired outcomes ... are not relevant to tasks in this category" (p. 47).

3. Conflicting Interdependence Among Paths

According to Campbell's definition, the presence of *Conflicting Interdependence Among Paths* can lead to an increase in task complexity. He notes that "If achieving one desired outcome conflicts with achieving another desired outcome, complexity will increase" (p. 44).

For our judgemental extrapolation task, it is particularly difficult to try to assess the magnitude of the interdependencies among the historical values, as perceived by task-doers. For example, it is difficult to determine just how a "typical" task-doer would approach the task at hand. Would they view the historical values as one big sequence? Or would they view them as a number of small localised sequences – perhaps, concentrated in areas where the historical time series has a stable trend? Or, perhaps the task-doer would view them as separate individual values.

In addition, there exists great difficulty in distinguishing, with certainty, the effects of signal (seasonality and trend) from noise. For a detailed discussion of these time series characteristics, please refer to section 2.3.3.

In relating the "components" of the time series to the complexity source, it must be noted that task-doers may choose a number of potential attributes or "task-associated information" (Campbell p. 47) when trying to predict future value. They may choose to rely on the slope of the time series, the shape, the absolute value of previous points or even the relative value/s of a group of points. In essence, there appears to be a large pool of potential attributes to use in anchoring their predictions, with a moderate to high level of conflicting level of interdependence among the paths.

Another point of interest is that in a very changeable series, many short sequences of adjacent historical values would need to be considered, these in essence, would add to *interdependence* complexity. Under the assumption that the time series for this study's forecasting task are reasonably changeable, there would probably be a fairly high degree of complexity generated from task-doers deciding how to utilise the interdependencies they perceive between the historical values.

The identification of the presence of conflicting interdependence in our task is once again in keeping with Campbell's categorisation for "Judgement Tasks" as presented in Table 2 (p. 47).

4. ***Uncertainty or Probabilistic Linkages***

Campbell reports that "Uncertainty can increase complexity by enlarging the pool of potential paths to a desired outcome" (p. 45).

He further advises that a task-doer will experience an increase in information load if they are not able to establish a clear relationship between the information elements provided and the ultimate goal, or desired outcome. Given that the proposed experimental task is one involving the use of human judgement, it would be expected that the level of uncertainty, and thus the level of complexity, would be fairly high.

Typically, whilst undertaking the task, in order to facilitate prediction of each new period value, task-doers must decide which information elements to utilise. They must also decide *how* to use them. For example, task-doers might need to determine relative weights to apply to them. The unclear relationship between the information elements provided and the desired outcome – level of accuracy, and the general lack of feedback, would imply that task-doers experience a fairly high degree of uncertainty.

This can be further supported by the notion that the presence of random noise has a detrimental effect on forecast accuracy when judgemental extrapolation is the method of choice (section 3.3.6 from Webby and O'Connor's 1996 findings, based on Sanders 1992; Lawrence and O'Connor 1992; O'Connor *et al.,* 1993). Given that the time series to be used for the experimental task are in essence real time series as used in a commercial environment, it would be expected that they would be subjected to varying levels of "noise" (for a more detailed discussion, please refer to section 2.3.3.6).

Webby and O'Connor's 1996 review confirmed Andreassen and Kraus' 1990 findings, that increased noise reduces people's ability to perceive and interpret patterns in the time series and impairs their ability to make accurate judgements when using extrapolative forecasting methods. Given these findings, in trying to assess the level of uncertainty or probabilistic linkages for the judgemental extrapolation task, one must conclude that the presence of this source must be fairly high. This also corresponds to Campbell's categorisation for "Judgement Tasks" as presented in Table 2 (p. 47).

In summary, applying Campbell's task classification, the complexity characteristics of this study's judgemental extrapolation task can be presented in Table 2.5.4.1 below.

Table 2.5.4.1

Fundamental Complexity – Judgemental Extrapolation of time series data

Task	Paths Source 1	Outcomes Source 2	Interdependencies Source 3	Uncertainty Source 4
Judgemental Extrapolation	—	—	Moderate-High	Moderate-High

As could be expected, when applying Campbell's general task classification, our forecasting task can best be described as a "judgement" task which focuses on finding the best way to incorporate available task information in order to achieve the desired outcome. The task is complicated by uncertainties involved in determining a path's specific relationship to an outcome.

The major focus of this Thesis is the effect of *media* and related *technology* on the accuracy of predictions in a judgemental extrapolation task. A review of the literature in the areas of Judgemental Forecasting and Task has already been presented in sections 2.3 and 2.5 above. The next section will present a detailed review of the Media literature. It will seek to identify and relate findings from empirical media studies to the actual tasks undertaken.

The overall aim of the current study, being to review the impact of *media* and related *technology* on task performance and to then seek to extend the area by conducting further research on a task that is deemed to be relevant in a decision making context – the judgemental extrapolation of time series data.

2.6 Media Studies – Review and Analysis of Related Literature: Reading and Proof-Reading Studies – Task Description and Measurement

2.6.1 Introduction

The major focus of this Thesis is the effect of *Media* and related *Technology* on the accuracy of predictions in a judgemental forecasting task. In section 2.3 we identified the basis for selecting judgemental extrapolation of time series data as the *object of interest* or, experimental task, for this study. The literature in relation to both Task-Technology Fit (TTF) and Task itself, were reviewed in sections 2.2 and 2.5. Having proposed a suitable taxonomy of Task/s and their related attributes in section 2.5, this was then used to classify our experimental task.

In this section, we conduct a detailed review of the literature and previous research in relation to *Media*. More specifically, theoretical and empirical studies undertaken in order to determine the effect of *Media* on task performance. The primary aim of this Thesis is to extend the area by conducting further research on a task that is deemed to be relevant in a decision making context – the judgemental extrapolation of time series data.

The issue of presentation and working medium (*media*) or choice of optimum display format (*mode*) in order to maximise task performance has been a much debated topic. One of the oldest and largest studies to review the adequacy of different information display formats for tasks of varying complexity was conducted by Washburne (1927). He used an accuracy measure of information recall in order to compare results. Over the past thirty years, this has become an increasingly important area, especially since the introduction of computers into the workplace.

Many of the studies that followed tended to concentrate on the issue of display format or *mode* of presentation. (For a detailed review please refer to Section 2.3.5). Much research in this area was undertaken in an effort to determine the impact of presentation format or *mode* of output, on the decision making process. Generally, studies tended to concentrate on the relative merits of different data display formats – i.e. graphical versus tabular.

Yet another related area – being the area of interest for this current section – is the effect of presentation *media* on task performance. For the purposes of this section, this area shall be referred to as "Media studies". However, much of the research that has been conducted in Media studies over the past twenty years, has tended to concentrate on a very narrow range of tasks, namely reading for comprehension and proof-reading. These studies have used a variety of experimental paradigms, making it problematic to interpret and compare results across studies (Hansen and Haas, 1988). The tendency of many researchers to analyse both reading and proof-reading results together, as if they were essentially one task, has certainly exacerbated interpretation difficulties. Hansen and Haas, Muter and Maurutto (1991), and Dillon (in his 1992 critical review of media studies) have all, in some way, succumbed to this tendency. However, the above researchers have also admitted that, "reading and proof-reading share some component processes, but other processes are unique to each skill" (Muter and Maurutto, 1991 p. 257).

The assumption that proof-reading is a substantial and precursory component of the task of reading for comprehension is refuted by Mills and Weldon's (1987) findings. They report that for adult readers, legibility and word identification (i.e. major proof-reading processes) play only a minor role in reading. Consequently, this research will view proof-reading and reading as two separate tasks and analyse media studies in the literature accordingly.

By presenting empirical findings from previous Media studies according to "Task Type", this section of the literature review will seek to show that, contrary to Hansen and Haas' 1988 view, one is able to interpret and compare results across studies. This section will also highlight both similarities and difference between the experimental task to be undertaken (the judgemental extrapolation of time series data) and the tasks being reviewed as part of the Media literature review – proof-reading and reading for comprehension.

2.6.2 Media – Presentation and Working Medium Definitions:

This section will provide definitions for the presentation and working Media used in the following research; Paper and Screen.

2.6.2.1 Paper:

The term "paper" has generally been used to describe tangible or "hard-copy" materials. The task information is presented on some form of physical paper material which approximates the pages of a book or journal. The natural orientation for this medium is horizontal, with the task-doer generally looking down onto the medium. However, the manipulation facilities are vast, allowing the user to pick up the paper and either re-position it or manipulate it to suit their own needs. If the task information is required to be directly edited, this is achieved using devices such as a pencil and eraser.

2.6.2.2 Screen:

"Screen" medium involves an image of the task information or portions of it being projected onto a visual display unit (VDU) which is one of the input/output devices connected to a computer. Editing task information in this medium can involve a task-doer using a keyboard or manipulating a pointing / selecting device (eg. a computer mouse, light pen or joystick). In this medium, the interaction of editing device with the displayed task information is mediated via some form of computer software. The general orientation for this medium is for the screen to be in an upright position, facing the task-doer. Although, with new technological advances, horizontal and varying "degrees" of vertical orientation (allowing tilting of the screen) are now available. However, in general, manipulation facilities remain limited.

2.6.3 Proof-Reading – a Definition

Mills and Weldon (1987) define proof-reading in terms of "legibility", which generally refers to the ease of identification of text items, and is measured in terms of the number of items identified (p. 331). As an experimental task, proof-reading is often chosen because it is a visually intensive task which occurs in many real-world applications and can be accurately and objectively scored (Wright and Lickorish 1983, Gould and Grischkowsky 1984, Creed *et al.,* 1987).

2.6.4 Reading for Comprehension – a Definition

Cushman (1986) refers to this task as "reading for content" as opposed to reading for the identification of errors. Waern (1981) states that "free-recalls" are the most commonly used approach to assess comprehension. However, Dillon (1992) reports that, rather than "free-recalls" most media-related experimental studies used post-reading questionnaires (sometimes involving multiple choice questions, such as Muter *et al.,* 1982). Typically, as in Askwall (1985), the questionnaires test specific factual and inference information from the article(s). As the comprehension questionnaires follow the reading itself, task information manipulation consists mainly of paging within articles. Consequently, the reading task typically requires minimal manipulation to complete.

2.6.5 Task Measurement

In his review of reading and proof-reading studies, Dillon (1992) describes speed, accuracy (of proof-reading), fatigue and comprehension (*accuracy* of reading) as outcome measures which are used to assess task behaviour. For proof-reading, accuracy is determined by the number of errors in the text which are correctly detected (or remain undetected, Wright and Lickorish, 1983).

However, Dillon advises that it is more difficult to derive a suitable means to quantify reading accuracy (comprehension). Waern and Rollenhagen, 1983, propose that reading accuracy (comprehension) can be seen as a quantifiable measurement of the participant's performance in terms of "the output produced by a reader after having finished reading" (p. 454). Applying the "obtaining of meaning" description of reading, accuracy is the extent of recall of meanings (both fact and inference) from a text supplied rather than, say, a photographic recall of the text itself. Comprehension questionnaire scores are only a surrogate means of measuring recalled meaning.

A wide range of measures have been used to used to gauge comprehension, such as requesting task-doers to complete missing parts of a message, or a sentence; or asking them to provide answers to questions based upon the message. Many of these types of measures have been used in the comprehension media studies presented in Table 2.6.6.2.1 (a full list of the types of measures as categorised by Carroll (1972) can be found at Appendix 3).

Speed also has a different basis when the two tasks are compared. A task-doer's proof-reading speed is the time taken to both examine the text and record the errors. In reading for comprehension, speed only refers to the time taken to examine the text.

In order to simplify the experimental design and minimise any potentially confounding factors, the current study will employ the use of a single variable in order to measure task performance. Task measurement for the experimental task – judgemental extrapolation of time series data – will be restricted to only one variable, that of accuracy. This will be measured by the accuracy of predictions. Given the possibility of a speed / accuracy trade-off, the use of a single variable to provide an objective measure of performance is deemed to be a primary focus of the current study. This should help to ensure that task-doers are able to undertake the task without the presence of Campbell's "Multiple Outcomes" task complexity source (as discussed in section 2.5.1). In turn, this should help to at least stabilise, if not minimise, task complexity.

2.6.6 Empirical Studies: Media Studies – Proof-Reading and Reading

The following section will provide a review of empirical studies conducted over the past twenty five years. Specifically, the review will seek to classify Media studies into either proof-reading or reading studies. A detailed summary of the review is presented as Appendix 2 and a brief summary presented as Table 2.6.6.2.1 is included in this Section.

2.6.6.1 Proof-Reading – Empirical Studies

Studies in relation to proof-reading accuracy, have produced varying results, Wright and Lickorish (1983), Gould and Grischkowsky (1984), Gould *et al.,* 1987a and Gould *et al.,* 1987b, reported no significant differences between the media, whereas Creed *et al.* (1987) and Wilkinson and Robinshaw (1987) found screens to be significantly less accurate than paper. According to Dillon (1992), accuracy "most commonly refers to an individual's ability to identify errors in a proof-reading exercise" (p. 1300). His 1992 critical review attempts to provide some form of explanation for the decrease in accuracy in the screen medium, by linking it to the task itself. Dillon notes that "... on typical VDUs, accuracy may be lessened for cognitively demanding tasks" (p. 1305). This proposed relationship between accuracy and task complexity when using a screen-based medium, has previously been suggested by others. However, few, if any studies have sought to provide empirical evidence to test this proposition – this will be discussed further in section 2.6.7 below.

A review of the proof-reading literature identifies that conditions which have been examined, involve misspelled words and errors in punctuation. These occur through missing or additional spaces or letters, double or triple transpositions, misfits or inappropriate characters, and missing or inappropriate capitals (Wright and Lickorish, 1983; Gould and Grischkowsky, 1984; Wilkinson and Robinshaw, 1987; Creed *et al.,* 1987). The errors in the studies were "seeded" by the researchers; however, task-doers were normally only given vague information about the number of errors in an article. For example, Gould *et al.* (1987a) and (1987b) advised task-doers that there were from one to ten misspelled words to be identified.

One of the founding studies in relation to the media debate for proof-reading was by Wright and Lickorish (1983), who were primarily concerned with the adequacy of the screen for a proof-reading task, specifically, the refereeing of electronic journals.

Wright and Lickorish (1983), and Gould and Grischkowsky (1984) both concentrated on the identification of misspelled words (being one component of the proof-reading task). In Gould and Grischkowsky's study, task-doers were specifically instructed to look for one of four types of misspellings: letter omissions, substitutions, transpositions and additions. Both studies found significant speed differences between the media, with task-doers performing the task better in the paper environment. However, accuracy measures produced different results with both studies reporting no significant differences between the media (Wright and Lickorish reported a greater accuracy in the paper medium, though not significantly so; whilst Gould and Grischkowsky found accuracy between the media to be almost equivalent).

Wilkinson and Robinshaw (1987) argued that the task selected by Gould and Grischkowsky does not equate to proof-reading but is merely an exercise in the identification of spelling mistakes. They attempted to avoid spelling or contextual mistakes during their study, and concentrated on the following five types of errors: missing or additional spaces or letters, double or triple reversions, misfits or inappropriate characters, and missing or inappropriate capitals. They believed their approach to be more relevant to the task demands of proof-reading. Findings from their study indicated a significant difference between the media for accuracy, speed and fatigue, with task-doers producing better results in the paper medium than the screen.

Proof-reading studies conducted by Creed *et al.* (1987), isolated differences in terms of three major types of errors:

- *Visually Similar* – Single letters substituted resulting in a 'non-word' e.g. *e* for *c*
- *Visually Dissimilar* – Single letters substituted resulting in a 'non-word' e.g. *s* for *j*
- *Syntactic* – For example, altering the tense of a word resulting in incorrect grammar.

They reported significant differences in the media in terms of accuracy, with task-doers performing better in the paper medium. In terms of speed, task-doers did perform the task faster in the paper medium. However, this reported difference was not significant.

In general, even though the types of errors to be identified may have differed slightly between the various studies, the overall aim has remained the same – to provide a quantifiable measurement of the participant's performance in terms of the number of errors identified by the task-doer.

Typically, studies required task-doers to proof-read an article and to "mark" the error upon identification of a set of pre-determined conditions. When using the paper medium, the error marking was achieved through either manipulation of a pencil (Wright and Lickorish, 1983; Gould and Grischkowsky, 1984; and Creed *et al.*, 1987) or verbally reporting the error (Wilkinson and Robinshaw, 1987; Gould *et al.*, 1987a and 1987b). In the screen medium, the variety of error marking approaches and devices adopted in the studies included; using a joystick (Creed *et al.*, 1987), a light-pen (Gould and Grischkowsky 1984), a keyboard (experiment S+S, Wright and Lickorish 1983), recording on paper using a pencil (experiment S+P, Wright and Lickorish 1983), and verbal reporting (Wilkinson and Robinshaw, 1987; Gould *et al.*, 1987a and 1987b).

Given that many of the studies above produced varying results (as shown in Table 2.6.6.2.1) and given that a number of different error-marking mechanisms were employed, this represents an additional area of interest for the current study.

Relating our experimental task to those previously undertaken as part of Media studies, it is interesting to note that task-doers are required to "interact" with both forms of media (paper and screen) in both our experimental task, and the proof-reading task. In the paper-based medium, task-doers typically use a pencil and eraser whilst in the screen-supported medium task-doers will mark their predictions using a "manipulation mechanism" or "device" – mouse, keyboard, slider or combination of all. This "interaction" with the media is similar to that experienced by task-doers undertaking proof-reading experiments, where they were required to mark the error/s using a designated "mechanism". This is an area of interest for this Thesis. Our experimental task, the judgemental extrapolation of time series data, appears to have something in common with the proof-reading task. It would be interesting to note whether experimental findings from the current study would support previous findings from Media proof-reading studies. This will be reviewed further in Section 2.6.7 below which will present a review of the complexity of the proof-reading task by applying Campbell's (1988) framework.

2.6.6.2 Reading for Comprehension – Empirical Studies

Generally, most media reading experiments involve task-doers reading one or more articles (typically between one and four) with accuracy being measured by comprehension tests given after completion of the reading activity. The majority of studies have used meaningful but fictional article content, such as short stories.

Askwall (1985) describes that this approach helps to keep content familiarity constant between subjects. Typically, articles ranged from 200 words (Kak 1981) to up to 2,000 words (Kruk and Muter, 1984; Muter and Maurutto, 1991).

In terms of "reading" as an experimental task, Dillon (1992) believes that it is a limited and often distorted view of what people actually do when they read. It is generally accepted that people read for pleasure, entertainment or just to learn. Dillon argues that this traditional notion of reading bears little or no resemblance to the "experimental" reading task and that "a more realistic conceptualisation of human reading is required" (p. 1323). Mills and Weldon (1987) defined reading in terms of the "ease with which the meaning of the text can be comprehended" (p. 331), and referred to this as the "readability" of the text.

General performance measures for comparing reading performance across media studies have been related to speed or comprehension or to a combination of both (Mills and Weldon 1987). Other measures such as fatigue (Muter *et al.*, 1982, Cushman 1986) have also been used but are not of primary interest to this research. Various studies have used many different methods to measure comprehension but the general aim has remained the same – to provide a quantifiable measurement of the participant's performance in terms of "the output produced by a reader after having finished reading" (Waern and Rollenhagen 1983).

In terms of accuracy differences between the media, studies have produced overwhelmingly similar results. All of the studies reviewed as part of this research reported no significant differences in the media in terms of comprehension, with task performance in the paper medium being equivalent to that in the screen medium (Kak, 1981; Muter *et al.,* 1982; Switchenko, 1984; Kruk and Muter, 1984; Askwall, 1985; Cushman, 1986; Oborne and Holton, 1988; and Muter and Maurutto, 1991).

However, it is interesting to note that Muter and Maurutto's 1991 study found task accuracy in the screen medium to be higher for eight out of the twelve participants (although not significantly so).

In general, it is difficult to interpret and compare results across studies. This is due in part to both the complexity of the cognitive tasks involved and to the large number of experimental paradigms adopted by the various studies (Helander *et al.*, 1984; Hansen and Haas, 1988 and Dillon, 1992). However, in his 1992 critical review, Dillon concludes that "fears of ... reduced levels of comprehension as a result of reading from VDUs, would ... seem unfounded" (p. 1305). Although the present review of the literature does not indicate a resounding support for Dillon's conclusion, perhaps future research in this area will.

Although speed differences between the media are not of primary interest to present study, it is worth noting that a review of the literature indicates mixed results – as presented in Table 2.6.6.2.1. A number of studies have reported a significant difference between the media, with reading speed on paper being faster than screen (Kak, 1981; Muter *et al.*, 1982; Kruk and Muter 1984). Whilst others have found reading speed to be unaffected by the presentation medium (Switchenko, 1984; Askwall, 1985; Cushman, 1986; Oborne and Holton, 1988; Muter and Maurutto, 1991).

The research by Kruk and Muter suggests that one of the critical factors contributing to the speed difference between the media is the difference in the amount of information displayed per page. Most computer screens tend to display a reduced amount of information per page compared to paper.

Although primarily concerned with proof-reading, Gould *et al.'s* (1987a) study also included a brief experiment on reading for comprehension (as part of his experiment 3).

The study compared reading speed under three separate conditions; Paper (printed in the traditional way), Paper-Rotated (aspect ratio was similar to the screen) and Screen. Each task-doer read the same number of words under each treatment condition. Results indicated that, where the paper condition was made to resemble the screen (aspect ratio was similar to the screen), there was no significant difference in reading speed between the treatment conditions. This approach however, was criticised by Dillon who proposed that the intent of media studies was not to produce a "degraded" paper condition, but to seek to identify inherent differences *between* the media with a view to improving the screen-based medium in order to help maximise task performance.

Applying Mills and Weldon's (1987) interpretation of readability as "...the ease with which the meaning of the text can be comprehended" (p.331), one can see that in order to provide an accuracy measure for "comprehension", task-doers must be able to grasp the context of the passage being presented to them. In turn, the task could not be undertaken without applying some form of strategy, either gleaning meaning from the entire passage by skimming, or by a detailed review of all sections of the passage, or by means of various word groupings. Whichever way the task-doers chose to undertake the task at hand, a strategy would need to be developed and adhered to. Reading for comprehension would require the passage to be reviewed in its entirety, perhaps to be read in some structured sequence, in order to grasp the general meaning of the passage material. As such, task-doers undertaking the reading task would be required to use all the available information presented to them in order to complete the task. In essence, this could be strongly related to task-doers undertaking a judgemental extrapolation task, where they are deemed to be using all the data presented – the historical patterns and relationships – as the basis for predicting future values.

This is an additional area of interest for the current study. Our experimental task (the judgemental extrapolation of time series data) appears to have something in common with the task of reading for comprehension. It would be interesting to note whether experimental findings from the current study support previous findings from Media reading studies. This will be reviewed further in Section 2.6.7 below which will present a review of the complexity of the reading task by applying Campbell's (1988) framework.

The overall results of empirical studies which examined the paper versus screen media differences for the tasks of proof-reading and reading are presented in Table 2.6.6.2.1 below. The table summarises task performance in terms of accuracy (*aka* comprehension for the reading task), speed, and to a lesser extent fatigue.

Table 2.6.6.2.1 Summary of the Findings of Media Performance Studies: Proof-Reading and Reading for Comprehension

Proof-Reading:

AUTHORS	Year	ACCURACY	SPEED	FATIGUE
Wright and Lickorish	1983	p > s	P > S	N/A
Gould and Grischkowsky	1984	p ≅ s	P > S	p ≅ s
Creed, Dennis and Newstead	1987	P > S	p > s	N/A
Gould, Alfaro, Barnes, Finn, Grischkowsky and Minuto	1987a	N/A	P > S	N/A
Wilkinson and Robinshaw	1987	P > S	P > S	P > S
Gould, Alfaro, Finn, Haupt and Minuto	1987b	p ≅ s	p ≅ s	N/A
Picking	1997	p > s	N/A	N/A

Reading for Comprehension:

AUTHORS	YEAR	ACCURACY	SPEED	FATIGUE
Kak	1981	p ≅ s	P > S	p ≅ s
Muter, Latremouille, Treuniet and Beam	1982	p ≅ s	P > S	p ≅ s
Switchenko	1984	p ≅ s	p ≅ s	N/A
Kruk and Muter	1984	p ≅ s	P > S	N/A
Askwall	1985	p ≅ s	p ≅ s	N/A
Cushman	1986	p ≅ s	p ≅ s	P > S [a]
Gould, Alfaro, Barnes, Finn, Grischkowsky and Minuto	1987a	N/A	p ≅ s	N/A
Oborne and Holton	1988	p ≅ s	p ≅ s	N/A
Muter and Maurutto	1991	p ≅ s	p ≅ s	N/A

Legend: P = Paper; S = Screen;

P > S indicates significantly better (i.e., greater accuracy or shorter time or lesser fatigue).
P > S [a] indicates significantly better but only in relation to +ve polarity (dark characters on a light background).
p > s indicates better, but not significantly so.
p ≅ s indicates no significant difference.
N/A indicates not applicable to study.

The proof-reading studies in Table 2.6.6.2.1 show mixed results for all three performance measures across media (accuracy, speed and fatigue).

When considering the accuracy component of task performance, the table shows that accuracy was unaffected by the choice of media in all nine reading studies. However, results for proof-reading studies show extremely mixed accuracy variations between the media. Among these studies, the paper medium typically gives superior and sometimes significant accuracy improvements over the screen medium.

Overall results of Media studies show that when modern screen technology is used, a predominant difference between reading and proof-reading occurs. Findings indicate that reading accuracy does not deteriorate when screen media is used compared to paper, whereas proof-reading accuracy can often be compromised by the use of screen media. It would be interesting to undertake further studies using our experimental task – the judgemental extrapolation of time series data – and to compare findings with those of previous Media studies.

In terms of speed differences, findings in Table 2.6.6.2.1 show that the paper medium provided superior speed performance compared to screen in most reading studies prior to 1985 but studies since, have not recorded any speed differences. Perhaps this could be attributed to the improved quality of screen-based displays, or even the increased sophistication of the screen manipulation/paginating mechanisms over the years. Or, perhaps this can be attributed to the increased use of computers and people's increased familiarity with the screen-based medium. Gould *et al.'s* (1987b) study is the latest proof-reading study to measure speed and the only one who found no apparent speed differences between the media. Gould *et al.* (1987b) explain that their results diverge from previous findings because of the improved quality in the computer displays they used. Their suggestion that improvements in screen technology result in a more paper-like display seems plausible. However, the fact remains that for the majority of proof-reading studies, speed differences (many significant) were recorded. However, what must be noted that speed is not a variable that will be measured or analysed as part of this Thesis. The issue of speed differences in proof-reading media studies is therefore deemed to not be of interest to the current study and will not be reviewed further.

Findings reported in Table 2.6.6.2.1, identify that task fatigue has seldom been recorded in both proof-reading and reading studies. Only two out of the seven proof-reading studies and three out of the nine reading for comprehension studies adopted fatigue as a variable of interest. The low number of studies and variations in their findings indicate that for both tasks it remains unclear whether the choice of media affects fatigue when undertaking the task. In turn, the issue of task fatigue in media studies is deemed to not be of interest to this Thesis and will not be reviewed further.

Given the above, in order to review the influence of media on the judgemental extrapolation of time series data, task performance could potentially be viewed from one or a combination of factors such as; accuracy, speed or fatigue, or even some other perspective. However, the above summary helps to support the notion that the selection of accuracy as our proposed task measure – namely the accuracy of predictions – would appear to be an appropriate and fruitful position for this study to take.

This position is supported by accuracy being an "intentional" outcome measure of performance. Task-doers undertaking media experiments were all told ahead of time that their task accuracy would be assessed. On the other hand, not all studies identify whether task-doers were informed whether speed would be a performance measure. Even when task-doers were instructed "to read through the test texts as quickly and as accurately as possible looking for errors as they went", the speed/accuracy to performance relationship is still inexplicit (Creed *et al.*, 1987, p. 7 and 1988, p. 471). For instance, it is debateable whether task-doers thought that there was a speed/accuracy trade-off, or whether they though the experimenters were merely suggesting they don't dally whilst undertaking the task.

2.6.7 Complexity of the Proof-Reading and Reading Tasks

The previous section concluded that proof-reading accuracy is typically reduced if the task is conducted in a screen-based medium rather than a paper one. However, there is no accuracy reduction for the reading task. A reasonable assumption is that proof-reading and reading for comprehension are essentially two different tasks. As such, it would be useful to identify the "abstract characteristics" (Vessey 1994) of each task, and in turn the complexity of each task (using Campbell's 1988 framework).

One possible explanation for the marked difference in media results to-date is the proposition that the complexity of each task is different under different media conditions. That is, the reduced accuracy for proof-reading studies, may result from proof-reading becoming somehow more complex when undertaken on screen. To investigate this effect, an appropriate starting point is to classify and compare the complexity characteristics of both proof-reading and reading for comprehension.

In order to achieve this, Campbell's (1988) framework which, was introduced in Section 2.6.6.2 above shall be used. Campbell's framework has long been recognised as being appropriate for objectively assessing the complexity of often dissimilar tasks (Mykytyn and Green, 1992; Schütte and Jordan, 1996). His framework involves the following four fundamental task attributes or sources of task complexity.

1. **Multiple Paths**
The number of possible ways to arrive at desired outcomes.

2. **Multiple Outcomes**
The number of desired outcomes of a task.

3. **Conflicting Interdependence Among Paths**
When achieving one desired outcome conflicts with achieving another desired outcome.

4. **Uncertainty or Probabilistic Linkages**
Uncertainty of whether a particular path leads to a desired outcome or not.

Campbell also introduced a "typology of complex tasks" – using four general task classifications labelled; decision, judgement, problem and fuzzy. He then identified the "sources" of task complexity by relating them to the four fundamental task attributes – the essence of his theoretical model. For ease of reference, this will presented in the following tabular format (previously presented in section 2.5.1):

Table 2.6.7.1: Campbell's Complex Task Classifications: *

Task Type	Paths Source 1	Outcomes Source 2	Interdependencies Source 3	Uncertainty Source 4
Decision	—	X	X	X
Judgement	—	—	X	X
Problem	X	—	X	X
Fuzzy Tasks	X	X	X	X

* (adapted from Campbell (1988) Table 2 p.46 and Table 3 p.47)

In media studies, essentially the same task – reading for comprehension and/or proof-reading; is undertaken by task-doers for each medium. The fundamental task characteristics for reading and/or proof-reading will be unaltered across media. We must therefore look to explain any difference in results in terms of *media influence*, that is, a relationship between the task *and* the media.

Looking to Schroder *et al.'s* definition of the components of task complexity, it would appear that "media influence" falls into the category of their so-called "secondary" properties of the task situation. These include organisational and environmental factors such as noxity, eucity and the degree to which the situation refutes or disorients the task-doer. Schroder *et al.* describe that secondary properties can manifest themselves by combining with the fundamental characteristics to produce changes in the actual complexity experienced by task-doers. The media study results from the previous section as presented in Table 2.6.6.2.1 suggest that the fundamental complexity characteristics of proof-reading are more susceptible to combining with particular media properties than those of reading for comprehension.

2.6.7.1 Proof-Reading – Task Complexity

Examining proof-reading in terms of Campbell's framework, a task-doer's desired end-states involve identification of a small number of misspelt words contained in a particular article. The manner in which the misspellings were seeded suggests that proof-reading's single information dimension consists of the individual words in an article rather than various word groupings as was the case for reading. Whilst undertaking the task of proof-reading an article written in English, task-doers would be expected to commence at the top of the article and work their way through it, reading from top-to-bottom and left-to-right. In terms of *multiple paths*, it would not seem to be a major issue for the task of proof-reading.

As task-doers were generally instructed to read through the passage and mark and identify as many errors as possible within the allotted time frame, there would appear to be some conflict in terms of desired *outcomes*. In an effort to complete the task in the most efficient manner possible, whilst still achieving a high degree of accuracy, *multiple outcomes* would probably be a factor affecting task complexity. However, with only two obvious outcomes to achieve, one would expect the speed/accuracy trade-off to equate to a low contributing factor.

Low dependencies between words allow a task-doer to achieve the task by focusing on individual words in isolation, one at a time. Article size has varied among proof-reading studies, ranging from 75 lines (perhaps 700 words, Creed *et al.,* 1987), to around 1000 words (Gould and Grischkowsky, 1984; and Gould *et al.,* 1987b), up to 600 lines or over 5000 words (Wilkinson and Robinshaw, 1987). In essence, task *outcomes* and *interdependencies* in the proof-reading task can be considered to be low. The complexity associated with determining whether a particular word is in error or not, equates to *uncertainty* in proof-reading. The very low incidence of false-positives among errors reported by Gould *et al.* (1987b) and the nature of the misspellings, suggest that this determination activity, although present, is largely trivial.

2.6.7.2 Reading – Task Complexity

In terms of Campbell's framework, a task-doer's desired end-states when reading a particular article, would be an understanding of meanings contained in the various passages of text within the article. Additional end-states would also be the comprehension of the overall themes of an article resulting from a synthesis of the passage meanings. When the task is presented in experimental situations (as in the reading-media studies), task-doers would probably assume that they would be tested on their understanding of particularly prominent or summarised meanings present in the article/s.

Unfortunately, meaning is derived from grouping of words (i.e. sentences and paragraphs) which form the basis of this task's information dimensions. In addition, a single word in a group can negate its entire meaning which indicates that, at some level, a task-doer must process every word. Also, as word groups can potentially contain several meanings or interpretations, there are often at least a few alternatives to each information dimension. Many *interdependencies* may also exist between the word groupings as this is the manner in which overall themes present themselves. This suggests that the complexity of reading is significantly contributed to by the number of information dimensions and the range of *interdependencies* among and within these dimensions.

As with the proof-reading experiment, to read an article written in English a task-doer could be expected to read the passages on a page in a top-down, left-to-right fashion. However, in an experimental situation with a comprehension test following, task-doers would probably re-read passages they thought important or didn't fully understand on first reading. Some task-doers may also vary from the typical top-down approach and read passages in some other order. This suggests that there is a measure of flexibility in the approaches taken by task-doers when reading. However, in general, one would say that the number of *paths* would be fairly limited, which in turn, would equate to very low, or generally not applicable.

Considerable uncertainty exists for task-doers in a reading for comprehension experiment as, from their perspective, they may be later tested on *any* meanings present in an article. Task-doers must use their own judgement to determine which passages are more important and worthy of more attention than others. This indicates a moderate to high level of *uncertainty* in this task.

Task-doers were generally instructed to read through the passage and were advised that the object of interest of the study was both the time taken to complete the task and their level of comprehension after they had completed the task. As previously identified for the proof-reading task, there would appear to be some conflict in terms of desired *outcomes*. In an effort to complete the reading task in the most efficient manner possible, whilst still achieving a high degree of accuracy, *multiple outcomes* would probably be a factor affecting task complexity. However, with only two obvious outcomes to achieve, one would expect the speed / accuracy trade-off to equate to a low contributing factor.

2.6.7.3 Complexity Characteristics of the Proof-Reading and Reading Tasks

Applying Campbell's task classification, the complexity characteristics of the two tasks can be summarised in the following Table below:

Table 2.6.7.3.1 The Fundamental Complexity of Proof-Reading and Reading

Task	**Paths Source 1**	**Outcomes Source 2**	**Interdependencies Source 3**	**Uncertainty Source 4**
Proof-Reading	—	Low	Low	Low
Reading for Comprehension	—	Low	Moderate-High	Moderate-High

Looking to Campbell's general task classification scheme – with the presence of source 2,3 and 4, and a low ranking for all three sources – the proof-reading task can be considered a relatively simple "decision" task involving examination of all words and looking for possible solutions (i.e. errors). Whilst undertaking the task, the task-doer would be trying to identify as many errors within the allocated time.

On the other hand – with the presence of source 3 and 4, with a moderate to high ranking, and source 3, with a low ranking – the task of reading for comprehension is identified as possessing similar, yet different characteristics to that of proof-reading. This is in keeping with Mills and Weldon's (1987) findings, where they reported that for adult readers, legibility and word identification (i.e. major proof-reading processes) play only a minor role in reading for comprehension or pleasure. They defined reading in terms of the "ease with which the meaning of the text can be comprehended" (p. 331), and referred to this as the "readability" of the text. Mills and Weldon further maintained that "legibility is involved in readability, but that readability involves additional skills" (p. 331).

Given the identification of complexity characteristics for reading, as outlined above, and given Mills and Weldon's definition of reading, it would be fair to propose that the task of reading for comprehension could be equated to being a reasonably complex "judgement" task. Essentially, the task-doer will be required to consider and integrate a number of sources of information – individual words, groups of words, paragraphs, sentences – and to make a judgement on the best strategy to adopt in order to complete the task – read the entire article, read one paragraph at a time, re-read certain sections, skim the entire article.

In summary, applying Campbell's task classification scheme we can identify proof-reading to exhibit the characteristics of a "decision" task, whilst reading for comprehension may be likened to a "judgement" task. These classifications, identifying each task as possessing distinct properties or characteristics, would help to explain the reported differences in Media study results over the past twenty years. What remains to be seen, is whether these reported differences can be explained in terms of, and be related to, a theoretical model. Essentially, one which could be used to help predict the relationship between task type and the potential influence of presentation *media* – paper versus screen-based – not just *mode*, which is what many of the past studies have tended to concentrate on over the past twenty years.

2.6.8 The Relationship between Task Complexity and Accuracy in Media Studies

As could be expected, Table 2.6.7.3.1 in the previous subsection identifies that reading for comprehension is a more complex task than proof-reading. However, Table 2.6.6.2.1 – which summarises experimental results – shows that it was actually the less complex proof-reading task for which accuracy was often affected by media factors. This is an interesting, yet unexpected finding.

The proof-reading description given in the previous subsection suggests that as uncertainty is low, a task-doer would normally be involved in sifting through large numbers of words one at a time in a methodical and routine manner. This indicates that proof-reading has considerable propensity for automatic or subconscious processing where responses by the task-doer are reflexive in nature (Downton and Leedham, 1993). Automatic processing is usually associated with relatively fast task responses. Downton and Leedham describe that this results from the normally low capacity channel between sensory registers and short-term-memory (STM) having a much higher *data rate* when under unconscious control. Ericsson and Simon (1980) suggest that, with automation, the intermediate processing steps and products do not require the use of STM, indicating that fast task responses may be more due to decreased loads on STM than channel capacity increases.

In contrast to proof-reading, reading for comprehension requires the integration of all levels of representation of the text (Stevens and Rumelhart, 1975) which means that this process makes considerable demands on human working memory (Just and Carpenter, 1992).

Reading is altogether a much more engaging and attentive process for the task-doer, suggesting that significant amounts of conscious processing are needed as the task-doer constructs and integrates ideas from the stream of successive words in the text.

Perhaps an examination of the nature of the "interaction" between the *task*, the *media* and the *task-doer* can help to shed some light on some possible causes for the marked differences in results of proof-reading and reading studies. In turn, this could be further explored using this study's experimental task – the judgemental extrapolation of time series data.

Our review of empirical studies indicated a vast difference in the "manipulation device" or instrument between the two tasks. In the proof-reading studies, task-doers were asked to identify and "mark" errors as they came across them whilst reading the passage. However, in the reading studies, task-doers were left to read the entire article and were required to answer questions about the content after they had completed reading the passage. Perhaps this is one of the key factors in helping to explain the inherent differences in proof-reading and reading media studies – as presented in Table 2.6.6.2.1. Perhaps the interaction with the manipulation device plays a vital role in helping to explain the marked difference in results across the studies. This is one of the key areas of interest to the current study.

In essence, the experimental task shares characteristics which are common to both reading and proof-reading. Similar to reading-for-comprehension, our judgemental extrapolation task is deemed to be a fairly complex "judgement task". However, it is also deemed to be similar to the task of proof-reading, in that there is a high degree of interaction with the "manipulation mechanism" or device. Task-doers in both proof-reading and judgemental extrapolation have to interact with the "manipulation mechanism" in order to both "mark" proof-reading errors and to "predict" forecast values. Perhaps it is the "interaction" with the "manipulation mechanism" which holds the key to the reported performance differences in media studies. Just and Carpenter (1992) provide an insight into possible explanations for the reported differences. They advise that from the processing perspective, media influences on proof-reading seemingly defy a common accepted assumption in the automaticity literature; that more subconscious or automatic processes are less dependent on attention and thus are normally less vulnerable to any kind of interference. Speier, Valacich and Vessey (1997) conducted further analysis into the effects of task interruptions and information presentation on task performance. Using Corragio's 1990 definition of interruption – an "externally-generated …event that breaks the continuity of cognitive focus on a primary task" (p. 19), they advise that interruptions can create both capacity and structural interference (as reported by Kahneman, 1973). Citing the work of Baecker *et al.,* 1995, they mention that in today's general work environment, computer-based tasks often involve "high cognitive loads" and that in turn this might make the task more prone to interference from interruptions (p. 21).

Speier, Valacich and Vessey's (1997) proposed research model is included below. It serves to provide a framework by which the "cognitive fit" of the experimental variables can be determined in terms of decision performance – measured in terms of accuracy and /or time taken to complete the task. The experimental study sought to support the notion that "cognitive fit facilitates decision making because the specific process used to act on the problem representation is the same as that needed to solve the problem" (p. 23). Findings from the study identified that decision-makers performing tasks of a complex nature have little or no excess cognitive capacity. In turn, interruptions tended to lead to a deterioration in task performance.

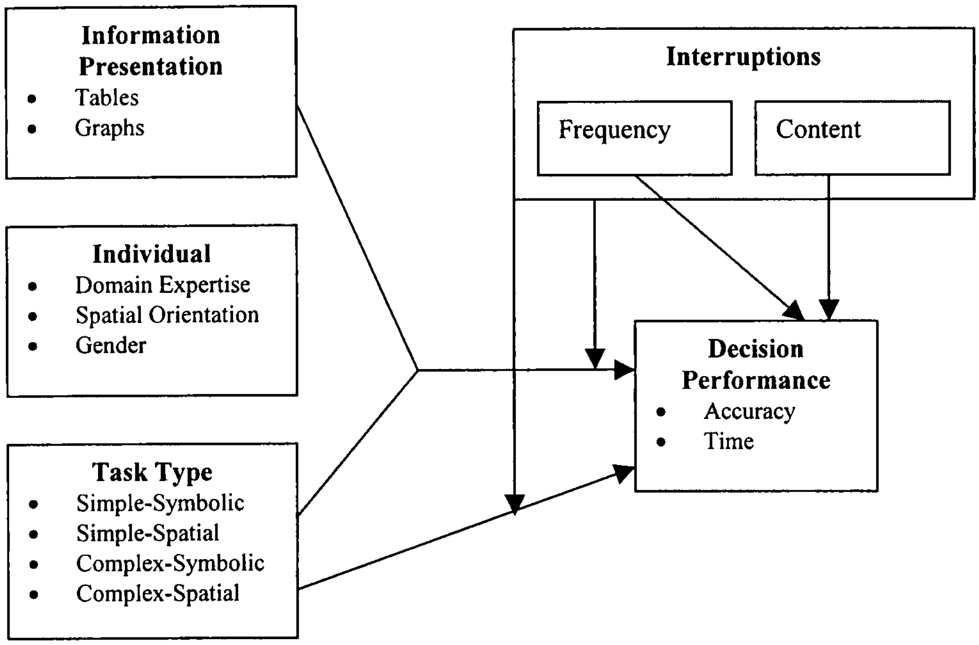


Figure 2.6.8.1 Research Model and Propositions

Speier, C., Valacich, J.S. and Vessey, I. 1997 p.25.

Findings from Speier *et al.'s* studies empirically confirmed Baron's Distraction / Conflict Theory (Baron, 1986), which indicates that "distractions facilitate performance on simple tasks and inhibit performance on complex tasks" (p. 22). They also state that, "...because both distractions and interruptions disrupt and potentially overload the finite cognitive capabilities of the decision-maker, it is likely that the effects of interruptions on decision performance will be similar to those of distractions" (p.22). Much of the work undertaken by Speier *et al.* sought to add to Vessey's (1991) study, which presented a theory based analysis centered on the notion of "cognitive fit" – presented in section 2.5.3. As previously noted, Vessey's theoretical model proposed that "cognitive fit facilitates decision making because the specific process used to act on the problem representation is the same as that needed to solve the problem" (Speier *et al.*, 1997, p. 23). It was also proposed that in situations where "cognitive fit" was found to be poor, results often reported decreased performance, in terms of task accuracy and / or speed. That is, where the match between the presentation format and the task is deemed to be suboptimal, there is increased pressure placed upon the task-doer. In order to understand the information presented, and undertake the task at hand, the task-doer must exert greater cognitive effort. It is this necessary increase in cognitive effort, which will often manifest itself in a decrease in task performance – measured in terms of accuracy and /or speed (Speier *et al.,* 1997 p. 24).

This is the basis for Goodhue and Thompson's argument in proposing their 1995 Theory of Task-Technology Fit (TTF). Although the primary focus of Goodhue and Thompson's model is the "fit" between the technology and the users' tasks, the basic premise remains the same, in order to "... have a positive impact on individual performance, the technology: ... must be a good fit with the tasks it supports." (p. 213). For a more detailed discussion please refer to section 2.2.

If we accept Baron's theory, that distractions facilitate performance on simple or automatic type tasks, and inhibit performance on complex or more "conscious" tasks, then we could also look to Cohen, Dunbar and McClelland (1990) proposition that there is a continuum of human processing between automatic and controlled or conscious processing. Using the results of Stroop effect experiments, they believe that rather than a particular process being discretely one or the other, it is best measured along a continuum, and that even relatively automatic processes may be susceptible to interference. Applying this conclusion, what may be happening in the proof-reading media experiments is that subjects undertaking the task on screen may process less automatically than those using the paper-based medium. That is; the "nature" of the task has changed across each of the presentation media, as has the effect of "interruptions" on task performance. In turn, it would be interesting to see if this basic assumption could be empirically tested using our experimental task – the judgemental extrapolation of time series data.

The following question needs to be asked: Would there be any reported difference between the accuracy of forecast predictions for subjects undertaking the experimental task in the screen-based medium, compared to those undertaking the task in the paper-based medium? In addition, if there were a reported difference, would this performance difference be deemed to be significant?

Another plausible explanation for the reported difference in task performance for the proof-reading task, is that "interference" or interruptions caused by task-doers' interaction with the manipulation device/s are having a greater impact on task performance than initially though possible. In terms of task performance, this decreased automaticity could then result in task accuracy reductions due to relatively lower channel capacities available from sensory registers to STM, or, perhaps an increased load on STM.

While the previous paragraph offers a possible explanation of why screen-based proof-reading is less accurate than paper-based, there remains the question of what factors contribute to the postulated decrease in unconscious or automatic processing in the screen-based version of the task. The wide range of potential factors may include characteristics such as the tangibility of the medium (being able to pick up and easily move around the paper or pencil compared to equivalent devices in the screen medium, Dillon 1992); the working orientation and resolution of the screen and its text compared to paper. However, as previously noted one of the more salient differences between the media in proof-reading and reading studies is the manipulation mechanism involved in "marking" an error. In general, task-doers in proof-reading studies were required to "mark" the error as they came across it whilst reading the passage. Potentially this could be seen as a "distraction". They were also required to use a manipulation device in order to paginate in the screen-based medium. Typically, these manipulation devices varied from the use of a keyboard, joy stick, light pen, through to the use of "verbal" identification of errors as they were identified. However, in the reading-for-comprehension studies, task-doers were able to undertake the task – read the passage – without any obvious "distraction" from a manipulation device. Comprehension tests were given after the completion of the task. There were minimal interruptions whilst the task was being undertaken. Task-doers in the reading studies were also required to use a manipulation device to paginate in the screen-based medium. However, this was the only distraction", and only occurred at the end of every page and could therefore not be classified as a "constant" interruption or distraction.

In trying to seek some insight into the possible causes for the differences in the results of Media studies as presented in Table 2.6.6.2.1, perhaps we could look to Waller's 1986 editorial comments. He maintains that a "distraction" occurs as readers need to articulate their needs in manipulating electronic texts, that is, think about moving the mouse or making an entry onto the keyboard rather than just "automatically" turning the page (p. 73). Waller advises that what we don't yet fully understand is what effect this "distraction" has or could have in terms of task performance. In essence, "...an intellectual construct has replaced the physical object" (p. 73). We need to further examine whether the "articulation" or "distraction" is putting an extra load on cognitive resources and in turn, whether this is having an effect on task outcomes or task performance.

This suggested "influence" of manipulation device has some support from the findings reported in Table 2.6.6.2.1. Two of the three studies in which a computer device was utilised in error marking in the screen-based medium (namely those by Wright and Lickorish and Creed *et al.*, 1987) reported reduced accuracy in the screen-based medium than on paper. From a pure manipulation perspective, using a keyboard or joystick seems to be less accurate compared to pencil.

The third study by Gould and Grischkowsky (1984), which instructed task-doers to mark errors with a light-pen, reported no differences between the media. This might suggest the plausible conclusion that light-pen manipulation in the screen-based medium is reasonably similar to using a pencil on paper.

The influence of manipulation devices has attracted some debate among researchers. Wright and Lickorish propose that it could be an important factor in the speed of proof-reading. However, Creed *et al.* (1984) suggest that the media differences in their study are due to factors other than the manipulation mechanism.

Gould *et al.* (1987a) also conducted studies in this area. Their focus was on speed differences rather than accuracy differences between the media. Gould *et al.'s* (1987a) error reporting mechanism was also decidedly different – using verbal reporting of errors for both screen-based and paper medium. Their results reveal that, even when the response mechanism is constant, task speed differences occur between the two media. This gives some indication that manipulation may not be the only media factor influencing task performance. Like Gould *et al.*, (1987a), Wilkinson and Robinshaw (1987) also conducted studies using verbal reporting of errors for both screen-based and paper medium. However, Wilkinson and Robinshaw's studies reviewed differences in the media in terms of speed, accuracy and fatigue. They reported a significant difference across all three performance measures, with results indicating poorer performance in the screen-based medium for all three measures – speed, accuracy and fatigue.

However, what must be noted here, is that by modifying the original task to have a different medium for error reporting – verbal rather than a direct "manipulation" device, this form of proof-reading could be considered to be a new task, or at the very least a "different" task compared to those undertaken in previous studies. According to Nisbett and Wilson (1977), verbalisation itself may sometimes interfere with normal task processing. This indicates that it may be problematic to apply findings of verbal-reported proof-reading experiments to wholly screen or paper-based proof-reading tasks. Therefore, the generalisability of Gould *et al.'s* (1987a) and Wilkinson and Robinshaw's (1987) studies may be questionable.

If the influence of manipulation on the proof-reading task cannot be ruled out, it is worth speculating how it could affect task accuracy. For the proof-reading task the impact is clearly indirect as task-doers use the manipulation mechanism only after they identify an error (apart from paging within the article). In screen-based experiments which rely exclusively on keyboard interaction (such as some of those by Wright and Lickorish), task-doers would need to recall various keystroke sequences in order to navigate the cursor to a word on screen and mark it as an error. It is likely that the required keystroke sequences would not be particularly meaningful to task-doers, suggesting that they would be difficult to learn and remember (Norman, 1988). The need to remember keystroke sequences and marking methods in order to detect errors in proof-reading (rather than mark the errors) would form extrinsic memory load. The maintenance of extrinsic load has been shown to interfere with the performance of tasks such as comprehension (Just and Carpenter, 1992).

Extrinsic memory load does not seem to account for the impact of Creed *et al.*'s (1987) joystick usage on proof-reading accuracy. However, the researchers have identified the use of a joystick – with red buttons to "shoot" words in error and a green button to paginate through the articles – as a limitation of their experimental design. Given this "odd" error marking mechanism, given that task-doers in Creed *et al.*'s studies had very little time to get used to this mechanism (they did a short practice exercise and a small number of trials taking approximately six minutes each), the experimental situation is quite likely to have disoriented many task-doers. This disorientation seems to be a possible explanation for the screen medium being less accurate than paper for proof-reading in Creed *et al.*'s experiments.

In summary, it is proposed that proof-reading may often be less accurate in the screen-based medium compared to paper because the manipulation mechanism or method of interaction with the media may be "distracting" the task-doer and having an adverse effect on cognitive load. Further, accuracy seems to be either indirectly (through reduced automaticity), or perhaps directly, impacted upon by the manipulation mechanism chosen for a screen-based version of the task. Disorientation or extrinsic memory load requirements may be the vehicles by which the manipulation mechanism affects accuracy.

Generalising the above proposal suggests that, if a particular task is normally done largely automatically and/or involves the task-doer in a considerable amount of interaction using a manipulation device, it may then be less accurately accomplished on screen compared to paper. Interaction compromises accuracy if the interaction method on screen is significantly more complicated or off-putting than using a pencil in the paper medium. The previous section also indicates that inherent automaticity in a task can be indirectly assessed by identifying the fundamental complexity characteristics for that task. Typically, tasks with low uncertainty, probabilistic linkages and conflicting path interdependencies will allow many task-doers to incorporate more automatic processing into their completion of the task and hence cross-media accuracy effects may occur with such tasks.

This is identified as an area where future research needs to be conducted. In fact, Goodhue and Thompson (1995) conclude that although their study provided a successful validation of their proposed TTF research model, much more work was still required in order to empirically test the model. Goodhue and Thompson concluded: "It is also important to go beyond *perceived* performance impacts, perhaps by constructing a laboratory experiment in which the model can be tested with objective measures of performance." (p. 231).

This is one of the major desires of the current study; to further the area by constructing and conducting a laboratory experiment which will provide an objective measure of performance. The key focus is to expand the Media "debate" and to further research in the area of TTF by introducing another task, one which will provide an objective measure of performance, namely – the judgemental extrapolation of time series data.

3. Research Methodology

3. Research Methodology

3.1 Introduction

The overall aim of this Thesis is to review the impact of *media* and the related *technology* on task performance and to extend the area by conducting such research on a task that is deemed to be relevant in a decision making context. As previously noted, the chosen task must also be considered to be an essential component of many business decisions and must provide an objective measure of task performance. In attempting to satisfy all of the above requirements, the task of judgemental extrapolation of time series data was selected.

In particular, this Thesis will seek to review the influence of presentation and working *media* and the related *technology* on the accuracy of the time series extrapolation decision. A series of four (4) experiments was undertaken which sequentially addressed issues related to the interaction of *media* and the related *technology*.

The experiments compared the accuracy of judgemental forecasting predictions using paper-based time series representations, with the accuracy of those produced using screen-supported time series display. For the paper-based treatment condition, task-doers were required to use a pencil and eraser in order to articulate their needs and register their forecast predictions. In the two screen-supported treatment conditions, task-doers were required to use either a combination of slider, mouse and keyboard or, the mouse, in order to articulate their needs and register their forecast predictions.

A number of laboratory experiments, four in total, were undertaken. The Analysis of Variance (ANOVA) statistical procedure was employed to facilitate data analysis.

The dependent variable for all the studies undertaken was the accuracy of predictions as entered by task-doers, using the Mean Absolute Percentage Error (MAPE) in order to assess forecast accuracy across treatment conditions. The independent variables included:

1. The Tool – the *media* and related *technology.*
   - 1.1 *Media* – which may also be referred to as the *presentation and working medium* (paper-based versus screen-based).
   - 1.2 *Technology* – the related "manipulation device" for each *media*
     - 1.2.1 Pencil and eraser
     - 1.2.2 Slider, mouse and keyboard
     - 1.2.3 Mouse

   (Where the pencil and eraser (1.2.1) are related to the paper-based media, and the slider, mouse and keyboard (1.2.2) and the use of the mouse (1.2.3) are related to two separate screen-based media conditions).

2. Task Complexity – applying Campbell's (1988) classification:
   - 2.1 Unstructured, with no decomposition – more complex
   - 2.2 Structured, decomposed – less complex

3. People – the emphasis being on education/age:
   - 3.1 Undergraduate
   - 3.2 Postgraduate

4. The time-horizon for predicted values:
   - 4.1 Periods 1-6 (first six)
   - 4.2 Periods 7-12 (last six)
   - 4.3 Periods 1-12 (All 12)

A discussion of the operationalisation of each of these variables will be presented in the following section. Specifically the experiments seek to provide some insight to the following question;

> Are there any differences in judgemental forecasting accuracy that are due to the Tool – i.e. the *media* and related *technology* – that is used in the forecasting process?

Whilst this is the overall research question for this Thesis, we will explore this question in different task complexity conditions (no decomposition versus decomposition) and with different cohorts of subjects (undergraduate students versus postgraduate students). A reference to the overall Research Model (presented in section 2.4) will suggest that the first question relates to the task environment or the Tool itself, the second to the nature of the task, and the third, to the nature of the people engaged in the judgemental forecasting task.

3.1.1 The Tool

For the purpose of this Thesis the Tool will be deemed to be related to both the *media* and the related *technology*.

In essence, the *media*, which may also be referred to as the *presentation and working medium* relates to the physical medium on which data representations are displayed to the user – or to the subjects undertaking the experiment. In turn, the t*echnology* refers to the related *manipulation device* for each *media* (pencil and eraser; slider, mouse and keyboard; and mouse).

The potential impact of the *media* and the related *technology* on task performance is of major interest to this Thesis. Each of the *components* of the Tool will be discussed in the following sections.

3.1.1.1 The Tool – Presentation and Working Media

As identified in section 2.6, prior research in the area of the influence of presentation and working medium on task performance has tended to concentrate on a very narrow range of tasks. Namely, two major tasks, reading and proof-reading. Findings from these studies have also produced inconclusive and mixed results (for a full summary and discussion of results please refer to section 2.6.5). However, some general conclusions can be made from the reported summary. The indication being, that when modern screen technology is used, a predominant difference between the reading and proof-reading tasks occurs. Findings indicate that reading accuracy does not deteriorate when screen media is used compared to paper; whereas proof-reading accuracy can often be compromised by the use of screen media, with proof-reading being more accurate when done with paper.

This body of research has used the general theoretical framework for task characteristics as developed by Campbell (1988) in order to categorise elements of reading and proof-reading. In addition, empirical media studies involving both reading and proof-reading have been reviewed separately.

This Thesis will seek to extend research in the *media* area – paper versus screen-based – by expanding the number of experimental tasks to include the judgemental extrapolation of time series data.

3.1.1.2 The Tool – The Technology

The general view of *technology* that the current study will adopt is that proposed by Goodhue and Thompson (1995 p. 216 – and discussed at section 2.2). For the purpose of this Thesis, *technology* is viewed as the tools that are used by individuals whilst carrying out their designated tasks. In the context of IS research, this generally refers to the computer systems as a whole and typically, this includes the hardware, software and the manipulation mechanism which allows the user to articulate their needs – all serving to assist users with the performance of their task/s.

More specifically, this Thesis will focus on the *manipulation mechanism* or *device* used by task-doers in order to articulate their needs and perform the experimental task. Three different types of manipulation devices were employed as part of the experimental design.

Tool 1 – slider, mouse and keyboard (screen-based Media)
Tool 2 – pencil and eraser (paper-based Media)
Tool 3 – mouse (screen-based Media).

Subjects participating in the experiment were only assigned to one of the above treatment conditions.

3.1.2 Task Complexity

As discussed in sections 2.3 and 2.5, the judgemental extrapolation of time series data has been selected as the experimental task for the current study. In addition, Campbell's (1988) taxonomy of tasks has been used in order to determine the complexity of the judgemental forecasting task.

3.1.2.1 The Task – Judgemental Extrapolation of time series data

Initially introduced in section 2.3, the basic task of forecasting can be described as "the gathering and analysing of repeated observations" (Makridakis and Wheelwright 1989, p.9) and the prediction of future events. Many researchers have identified that in practice, forecasting is typically performed through the exercise of human judgement (Dalrymple, 1987 and Goodwin, 1998). In addition, the methods used to record and analyse observations and derive predictions determine the nature of the forecasting task. For example, if observational information is recorded directly in human memory and then used intuitively for prediction purposes, the forecast is referred to as judgemental (Makridakis and Wheelwright 1989, p.9).

Judgemental forecasters may use a number of techniques in order to try to predict future events. Of interest to the current study, is the use of historical patterns and relationships as a basis for predicting future values, this is commonly referred to as extrapolation. Armstrong (1985, p.152) describes that extrapolation, in its judgemental form, is known as "eye-balling" and relies on the assumption that the future events will conform to the available historical data.

3.1.2.2 Complexity Characteristics

In terms of determining the *complexity* of the judgemental forecasting task, this current research has looked to the work undertaken by Campbell (1988) and has applied his proposed taxonomy of tasks (for a full review please refer to section 2.5).

We previously identified the judgemental extrapolation of time series data as a "judgement" task, one which focuses on finding the best way to incorporate available task information in order to achieve the desired outcome. We also noted that the task is often complicated by uncertainties involved in determining a path's specific relationship to an outcome. In order to fully appreciate the complexity characteristics of the experimental task, this study sought to identify the individual components of time series forecasting. This was able to be achieved by decomposing the time series into several unique components or sub-patterns. In particular, trend and seasonality were of interest to the current study.

It is generally accepted that decomposing a task into several smaller components or subsets, allows the task-doer to concentrate on one thing at a time and thus enables a complex task to be reduced into smaller and "cognitively more manageable parts" (MacGregor and Armstrong, 1994 p. 32). Thus by definition, decomposition may be seen to be decreasing the complexity of that task. Findings from Armstrong *et al.'s* 1975 study showed that in general, the use of a decomposed approach led to improvements in people's judgement decisions; although as MacGregor (2001) reports, it will not always be beneficial.

The basic premise behind decomposition, the notion of *structure*, has also been the subject of many research studies. Findings from MacGregor *et al.'s*1988 study showed that when greater structure was imposed on the decision aid, accuracy increased. Overall, the notion that judgemental decomposition adds structure to the task, has been supported by findings as reported in Webby and O'Connor's 1996 review of the literature and MacGregor's 2001 study, which deemed decomposition to be "... an effective strategy for improving the quality of judgemental forecasts" (p. 121), particularly for tasks with high uncertainty.

In terms of decomposition and its effect on performance outcome, studies in this area continue to provide inconclusive and mixed results. This was further supported by findings in Webby and O'Connor's 1996 review. They concluded that more investigation was required into the impact of decomposition on the relationship between task complexity and decision quality.

This Thesis will decompose time series representations and review the impact of structure on the accuracy of predictions in a judgemental forecasting task – in essence, the relationship between task complexity and decision quality. It will also seek to review the influence of the *mode* of representation (i.e. the media) and thus extend the focus of prior research conducted in the area of task presentation.

For the purpose of the current study, the experimental task will be presented as either:

- Unstructured, with no decomposition and thus deemed to be a more complex task
- Structured, decomposed – in terms of trend and seasonality – and thus deemed to be a less complex task.

3.1.3 People

For the purpose of this Thesis, the "People" variable will be viewed in terms of education and age. People will be classified as either Undergraduate or Postgraduate subjects. This will take into account the differing levels of education as well as age and maturity between the two groups. This distinction is to support and acknowledge the inclusion of "Individual Characteristics" as an additional variable in Goodhue and Thompson's TPC model (1995). As the authors note, the inclusion of this variable serves to highlight the fact that it is the individuals who are using the technology to assist them in carrying out their tasks; "characteristics of the individual (training, computer experience, motivation) could affect how easily or well he or she will utilize the technology" (p. 216).

The categorisation of "People" in terms of education and age in the proposed research model acknowledges that it is the "experience", which has been identified as a "people" characteristic in Goodhue and Thompson's model of TTF (1995). However, what must be noted, is that the focus has moved away from the traditional notion of "experience" which has generally been linked to the task itself. This is required in order to recognise that the people, or subjects, undertaking the task under the different treatment conditions – the *media* and related *technologies* – have to also be familiar with the different forms of *media* and the *technologies* which they will be exposed to. With this is mind, we will concentrate on another individual characteristic which we believe to be a more appropriate indicator for the experimental studies undertaken as part of the current research. It is believed that the "age/maturity" characteristic of the people undertaking the task will provide a better indication of "experience" – particularly in relation to the *media* and the related *technologies*. For the purpose of the current study, this emphasis provides the basis for using the undergraduate and postgraduate categorisation, with postgraduate subjects deemed to have a greater age, level of maturity and education than undergraduate subjects.

3.1.4 Time Horizon

As previously noted in section 2.3.3.3 there is a tendency for judgemental forecasting accuracy to decrease as the forecast horizon increases (Dalrymple and King, 1981; Armstrong, 1985; Lawrence *et al.*, 1985; O'Connor, Remus and Griggs, 1993).
One of the aims of this Thesis is to determine the extent, if any, of differences between the accuracy of predictions across the time horizon. In particular, across two "sets" of time horizon groupings; periods 1-6 and periods 7-12. We are interested not in the

influence of time horizon *per se*: we are only interested in horizon as it interacts with the accuracy of various tools and task complexity.

3.2 Research Design

To investigate the impact of presentation and working medium – *media* and the related *technology* – on the accuracy of the time series extrapolation decision, a series of laboratory experiments or studies (four in total), was conducted. Full details of each of the experimental designs will be provided in Sections 4 - 7 below. The basic structure of the experimental studies undertaken as part of this Thesis is presented in Table 3.2.1 below:

Table 3.2.1 Structure of Experimental Studies

Exp ID	Subjects	Tool ID	Media	Manipulation Device	Task Complexity
1	U/G	Tool 1	Screen	slider, mouse & keyboard	• Unstructured • no decomposition
1	U/G	Tool 2	Paper	pencil & eraser	• Unstructured • no decomposition
2	P/G	Tool 1	Screen	slider, mouse & keyboard	• Unstructured • no decomposition
2	P/G	Tool 2	Paper	pencil & eraser	• Unstructured • no decomposition
3	P/G	Tool 1	Screen	slider, mouse & keyboard	• Unstructured • no decomposition
3	P/G	Tool 3	Screen	Mouse	• Unstructured • no decomposition
4	P/G	Tool 3	Screen	Mouse	• Structured • decomposition
4	P/G	Tool 2	Paper	pencil & eraser	• Structured • decomposition

Each of the experiments sought to build upon the other, and to further investigate the relationship between judgemental forecast accuracy and the Tool – the *media* and the related *technology*. The overall aim of this Thesis is to add to the body of research in the area of Task-Technology Fit (Goodhue and Thompson, 1995). The basic premise behind the theoretical model of TTF is that it is the interaction between the task, the technology and the individual which is crucial to the framework of TTF (Goodhue 1995). This Thesis will seek to address one of the major limitations of Goodhue and Thompson's 1995 study, the fact that given the nature of their study, they were not able to obtain an *objective* measure of performance in the field context. They themselves admit that no objective measure would have been compatible across such a large number of individuals with many diverse task portfolios. Instead, they had to measure performance *subjectively* – in terms of "perceived" performance impacts as rated by task-doers i.e. – user evaluations of TTF. This Thesis will seek to extend the area by conducting a series of controlled laboratory experiments across different *media* and related *technology*. These experiments will be conducted using a task that will; provide an objective measure of performance; is widely used in a business environment, namely – the judgemental extrapolation of time series data.

3.3 Design of the System

The screen-based tool was developed using Microsoft Visual Basic on a Personal Computer (PC) (a copy of the Tool has been provides as Appendix 7). This tool enabled a comparative study to be undertaken to determine the impact if any, of the variation of presentation and working medium and related technology. In addition, the screen-based tool utilised a graphical user interface (GUI) in order to minimise the instrument learning, required by task-doers.

The paper-based treatment condition was identical to that presented to task-doers in the screen-based treatment condition and was basically a "screen dump" with all environmental attributes or factors held constant (please refer to Appendix 4 for a detailed review of these environmental attributes).

3.4 The Experimental Task

To experimentally investigate the effects of the Tool – *media* and related *technology* – on forecasting accuracy, an experimental task of roughly similar scale and complexity to many of those used in the reading and proof-reading studies was required (please refer to section 2.6 for a more detailed discussion). In addition, time series data for which actual values were available was required. The origin of the time series data will be discussed next, followed by a discussion of the experimental task to be undertaken by subjects.

3.4.1 Origin of the Time Series data

Time series data for the forecasting task were selected from the M-Competition database of 111 real time series (Makridakis *et al.,* 1982). This database has been used as a source of real time series in other forecasting research (Davey and Flores 1993; Edmundson and Terry 1986; Lawrence *et al.,* 1985; Makridakis and Winkler 1983). The database contains both quarterly and monthly time series with 50 to 100 observations per series, facilitating the evaluation of forecast accuracy.

The series in the M-Competition database were identified as being either seasonal or non-seasonal on the basis of autocorrelation analysis. In total, to satisfy all four experimental studies, five randomly selected monthly series were used for the task – three of the series chosen were seasonal, and two were non-seasonal, all had the number of monthly observations available.

3.4.2 The Judgemental Extrapolation of time series data

The forecasting task presented historical data graphically as a time series (i.e. a series of linked points or values for previous periods). The graphical presentation of historical data facilitating task performance, especially when the task is undertaken by novice forecasters (Lawrence, Edmundson and O'Connor, 1985).

The task was relatively straight-forward and quick to undertake, as well as easily explained to non-expert task-doers. Task-doers were asked to predict point forecasts for future periods rather than probability, distribution or prediction interval methods of forecasts (Goodwin, 1998). As previously advised, time series data for the experimental task were selected from the M-Competition database of 111 real time series – primarily to facilitate evaluation of forecast accuracy.

For each of the time series presented, task-doers were asked to predict the next twelve (12) periods (i.e. months) ahead, based on the previous thirty six (36) periods (i.e. three years) of historical values provided to them. No feedback was provided about the accuracy of predictions. This is in keeping with previous media studies.

The volume of historical data and time horizon for predictions meant that all the task information for one time series could be provided on one page (i.e. one computer screen-full or one A4 landscape piece of paper). This enabled task-doers to concentrate on completing one time series at a time before moving onto the next.

In an effort to minimise any between-subject accuracy variations, no contextual information was provided (for a more detailed discussion please refer to section 2.3.3.4). For the forecasting task in this study, subjects were provided with only the historical time series and no information about what the series measured or the environment in which it was used. This choice both simplifies the task and enhances its similarity to the proof-reading and reading media studies where no efforts were made to either match or not match the task-doers to the material being proofed or read.

To facilitate evaluation of the effects of presentation and working *media* and related *technology*, groups of task-doers were randomly divided into one of the four treatment conditions as presented in Table 3.4.2.1 below:

Table 3.4.2.1 Treatment Conditions

Experiment ID	Tool_ID	Treatment Condition	Manipulation Device	Task Complexity
Exp 1 Exp 2	Tool 1	Screen-based	combination of slider, mouse and keyboard	Unstructured More complex
Exp 1 Exp 2	Tool 2	Paper-based	pencil and eraser.	Unstructured More complex
Exp 3 Exp 4	Tool 3	Screen-based	mouse	Unstructured More complex
Exp 4	Tool 3	Screen-based	mouse	Structured less complex

Graphical presentations depicting how the time series were presented to task-doers in each of the respective treatment conditions will be presented in the following sections. Each task-doer was presented with all the task information for one time series on one page/screen. The representations of time series data under each of the treatment conditions was identical in terms of the following – scale, orientation, aspect ratio, image polarity, display characteristics and contrast (for a detailed review of these environmental attributes, please refer to Appendix 4).

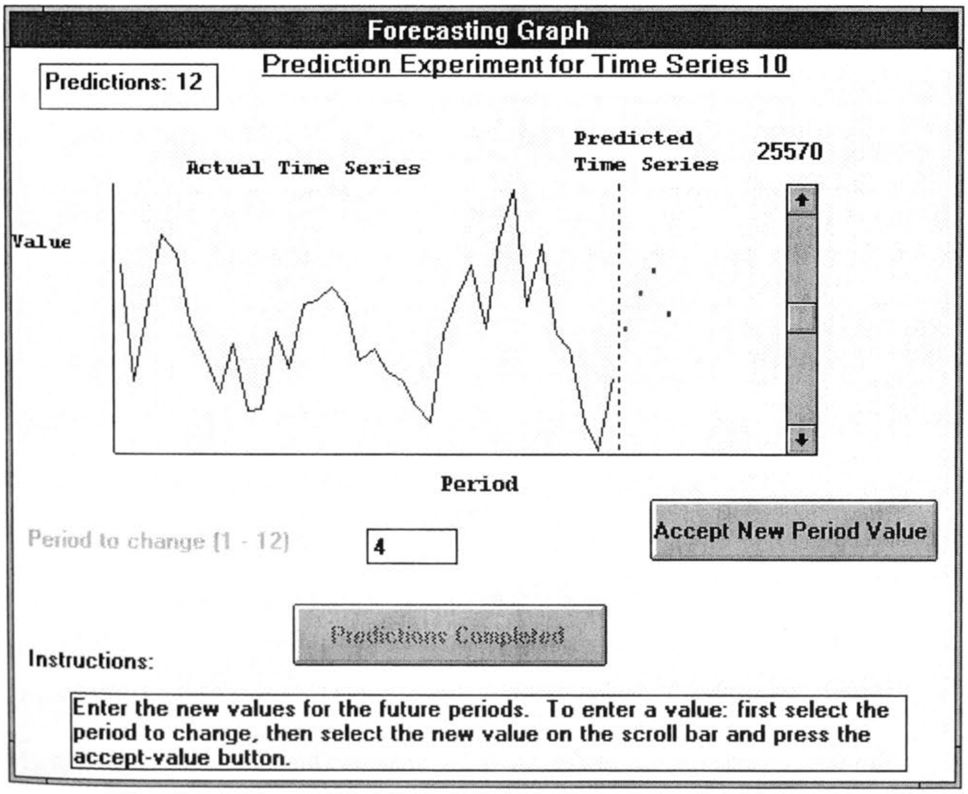
In each treatment condition, task-doers were given a standard introduction to the experiment and a prediction demonstration using a sample time series (these are presented as Appendix 5 and 6e). During the introduction, task-doers were advised that the results were purely for the basis of experimentation rather than any contribution to course credits.

In the screen-based treatment, task-doers were encouraged to familiarise themselves with the software tool and related *technology* before commencing with the main experiment. To do this, they utilised the software tool and related *technology* to record example predictions for the same sample time series as used in the demonstration. Similar familiarisation with the medium was not considered necessary for task-doers undertaking the paper-based experiment – as it was considered "natural" for task-doers to interact with paper and the related *technology,* namely a pencil and an eraser.

3.4.3 Tool 1: ***Media*** – screen-based; ***Technology*** – slider, mouse and keyboard ***Task Complexity*** – unstructured, more complex

A representation of the screen-based medium is presented as Figure 3.4.3.1. To record predictions, task-doers were required to use the keyboard to enter the period number and then the mouse to select the predicted value for that period using the slider (next to the time series). To complete the marking of the value on the graph, task-doers clicked on the "Accept New Period Value" button. This process was required for each of the twelve (12) period values to be predicted and for any subsequent prediction adjustments. After all twelve (12) periods had been predicted and accepted, the "Predictions Completed" button was enabled, once this button was clicked, task-doers moved onto the next time series. Once all period values for the final time series were entered and accepted, task-doers were thanked for their participation in the experiment. The software was disabled and the task completed.

Figure 3.4.3.1 Time Series Presentation[1] in the Computer-supported Treatment Tool 1 – Experiments 1 and 2



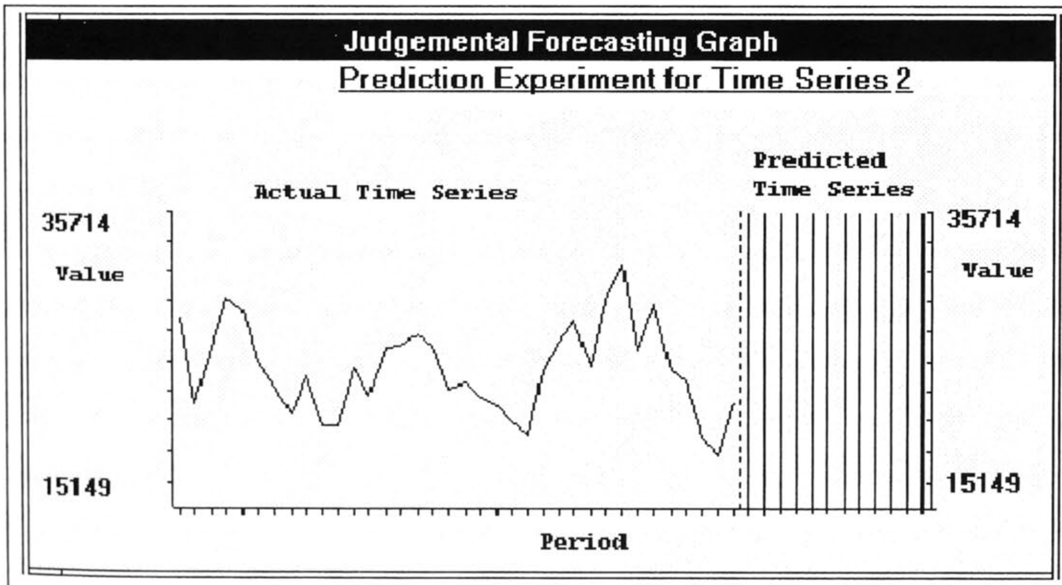
[1] Picture is smaller than actual size presented to task-doers.

3.4.4 Tool 2: *Media* – paper-based; T*echnology* – pencil and eraser
Task Complexity – unstructured, more complex

Prediction recording in the paper-based treatment was much less elaborate than in the screen-based medium. A representation of the paper-based medium is presented as Figure 2. Task-doers marked their predicted values on the graph image using a pencil, and, after carrying out any adjustments (using an eraser), moved onto the next time series.

Once all period values for the final time series were entered and accepted, task-doers were asked to raise their hand so that their paper-based submissions could be collected. They were advised that the task had been completed and were thanked for their participation in the experiment.

Figure 3.4.4.1 Time Series Presentation[2] in the Paper-based Treatment Tool 2 – Experiments 1 and 2:



You are required to plot the values for the next 12 periods on the graph above.
You may change your selected period values if you wish, erase previous values & plot new ones.
Once you are satisfied with each selection, please move on to the next graph and do not make any futher changes to this graph.

[2] Picture is smaller than actual size presented to task-doers.

3.4.5 Tool 3: *Media* – screen-based; *Technology* – mouse

Task Complexity – (1) unstructured, more complex

(2) structured, less complex

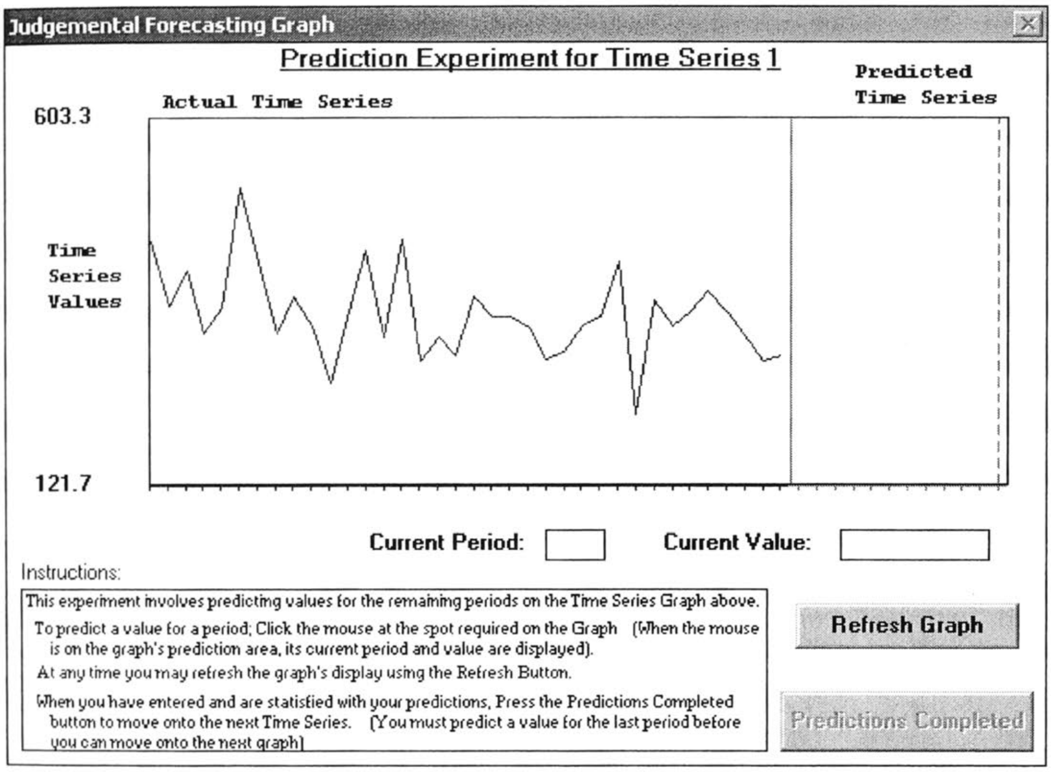
Representations of Tool 3, the screen-based *media,* utilizing mouse *technology*, are presented as Figures 3.4.5.1 and 3.4.5.2. It is necessary to show the two graphical depictions of the Tool because, although both the *media* and the related *technology* in both figures is the same, the task complexity has been varied. This has been achieved by the provision of additional features to enable the time series data to be decomposed or structured and thus display individual characteristics of the time series, such as seasonality and/or trend.

This is essentially a change in the display format – which in turn, has a direct impact upon task complexity – rather than the Tool itself, and will be discussed in conjunction with the graphical presentation in Figure 3.4.5.1 (for a more detailed discussion of the impact of structure and in particular, decomposition, please refer to section 2.3.4).

Using Tool 3 in order to complete the time series extrapolation task, subjects were required to use the mouse to "point and click" in order to register their predictions. Task-doers were instructed that in order "To predict a value for a period; Click the mouse at the spot required on the Graph" – these instructions were provided in the text box displayed at the bottom of the screen. They were also instructed that "When the mouse is on the graph's prediction area, its current period and value is displayed" – these two text boxes were located directly above the instructions box. Task-doers were also able to click on a displayed value, or data point, and simply use the mouse to "drag" that data point either up or down. Once all the predicted values for the time series had been entered, a graphical depiction of time series predictions was displayed. Task-doers were also instructed that "At any time you may refresh the graph's display using the Refresh Button" and that once they had entered all their predictions and were satisfied with them, they would need to press the "Predictions Completed" button to move onto the next time series.

Once all period values for the final time series had been entered and accepted, task-doers were thanked for their participation in the experiment. The software was disabled and the task completed.

Figure 3.4.5.1 Time Series Presentation[3] in the Computer-supported Treatment
Tool 3 – Experiment 3, *Task Complexity* – unstructured, more complex



[3] Picture is smaller than actual size presented to task-doers.

The major difference between the two screen-based Tools (Tool 1 and Tool 3) is the manipulation device or the direct level of interaction between the task-doer and the manipulation device. Tool 1 employed the use of the slider, mouse and keyboard. In order to record predictions, task-doers were required to use the keyboard to enter the period number and then the mouse to select the predicted value for that period using the slider (next to the time series). This required what could be described as a fairly *high* level of interaction with the manipulation device/s. In contrast, Tool 3 employed the use of the mouse. Subjects were required to use the mouse to "point and click" in order to register their predictions. In addition, task-doers were also able to use the mouse to "drag" a data point either up or down. Essentially, this required what could be described as a fairly *low* level of interaction with the manipulation device.

Looking to the results of empirical studies in the area of Media – as reviewed in section 2.6 – it can be seen that different results in terms of accuracy, were obtained for the two experimental tasks, namely, proof-reading and reading for comprehension. It was further speculated that perhaps the impact on task performance could somehow be related to the level of *interaction* with the manipulation device. In order to find support for this proposed impact, we looked to the work of Speier, Valacich and Vessey (1997). Their work concentrated on the effects of task interruptions and information presentation on task performance. Using Corragio's 1990 definition of interruption – an "externally-generated …event that breaks the continuity of cognitive focus on a primary task" (p. 19), they advised that, interruptions can create both capacity and structural interference (as reported by Kahneman, 1973). Findings from Speier *et al.'s* study further identified that decision-makers performing tasks of a complex nature have little or no excess cognitive capacity. In turn, interruptions tended to lead to a deterioration in task performance.

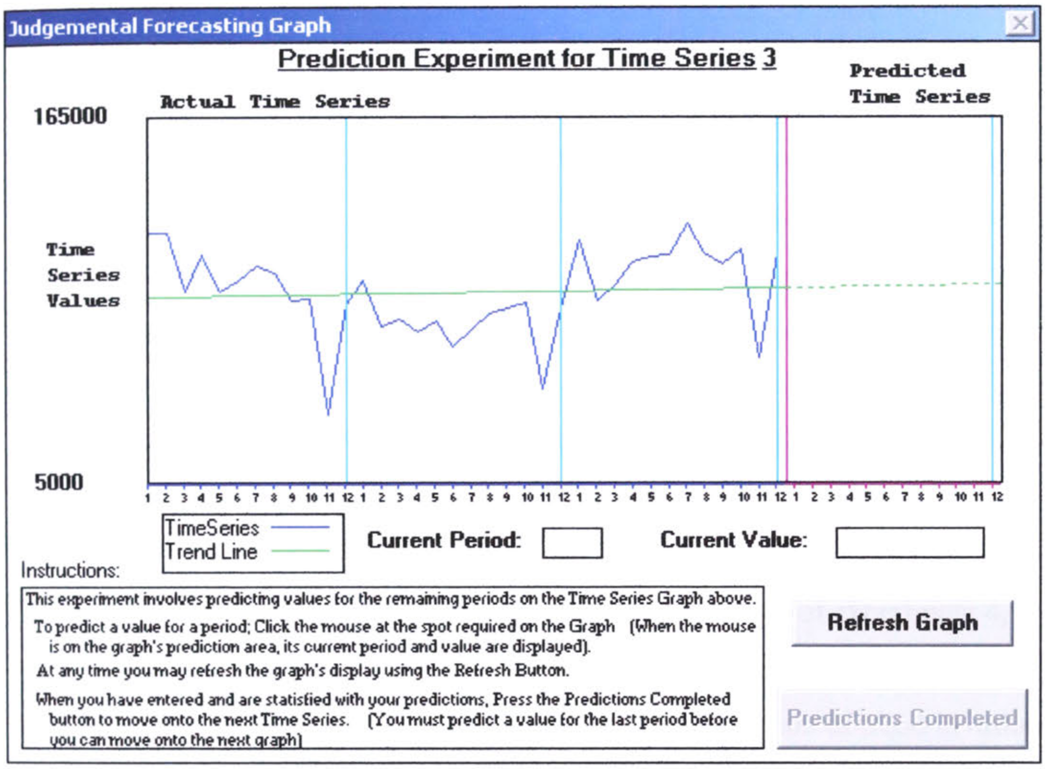
Citing the work of Baecker *et al.,* 1995, Speier *et al.* further proposed that in today's general work environment, computer-based tasks often involve "high cognitive loads", and that in essence this might make the task more prone to interference from interruptions (p. 21).

By simplifying the *degree* or *level* of interaction with the manipulation mechanism, experiment 3 sought to further examine Speier *et al.'s* propositions and also sought to support results as reported in the Media literature and as presented in section 2.6.

3.4.6 Tool 3: *Media* – screen-based; *Technology* – mouse

Task Complexity – (2) structured, less complex

Figure 3.4.6.1 Time Series Presentation[4] in the Computer-supported Treatment Tool 3 – Experiment 4; *Task Complexity* – structured, less complex



[4] Picture is smaller than actual size presented to task-doers.

As previously advised, experiment 4 sought to vary the last remaining variable presented in the Research Model (section 2.4) – namely that of Task Complexity. Using Tool 3 – screen-based *media* and related *technology* i.e. mouse – Time series data was presented in a *structured* mode i.e. time series data was decomposed. This was achieved by presenting the additional "features" of the time series data, such as Trend and/or Seasonality.

As can be seen in Figure 3.4.6.1 above, data for this time series (which is in fact M6) shows that the time series is Seasonal with a fairly flat Trend.

In the graphical depiction of time series data, the notion of seasonality, is shown by the inclusion of distinct separations between each set of twelve (12) period values. The inclusion of seasonality as a factor in the time series graphs, was used to add additional structure to the time series data representations and thus to the task itself. The addition of structure was also deemed to be aiding the "decomposition" of the task and in turn, decreasing overall task complexity.

The provision of additional information to the task-doer highlights the fact that this time series is seasonal, and that in general, history will tend to repeat itself over the next twelve period values – the periods for which they must register their predictions. This assumption can be supported by looking to Davey and Flores' 1993 study in which they define seasonality as a "structured pattern of changes within a year". The generally accepted notion of seasonality being, that it is a pattern which repeats itself over a series of fixed intervals in time (for a more detailed discussion on the impact of seasonality please refer to section 2.3.3.2).

Another additional feature included in Tool 3 for the purpose of experiment 4, was that of Trend. This factor was used to add additional structure to the time series data representations and thus to the task itself. In turn, the inclusion of trend in order to impose additional structure was also deemed to be aiding the "decomposition" of the task and decreasing overall task complexity.

As has previously been noted, in a time series, the trend factor represents the long-term behaviour of the data. In general, a trend pattern can therefore be identified as an increase or decrease in the value of the measured variable over a period of time, and is most commonly presented as a linear relationship between the data points. The trend line may also be projected into the future and used to aid forecast predictions. The use of trend in order aid task-doers is in keeping with the general principle of judgemental forecast accuracy which maintains that "judgemental forecasts based on trended series presented graphically are much less biased" (Harvey 2001, p. 64). (For a more detailed discussion on the impact and use of Trend in time series extrapolation please refer to section 2.3.3.1.)

Based on the above, our proposition is that for experiment 4, the judgemental time series extrapolation task undertaken by subjects using Tool 3 will be a structured, less complex task – essentially one that has been *decomposed.* In terms of the time series forecasting literature, decomposition is the term used to denote the separation of the data into several unique components or subpatterns – such as trend and seasonality – which are of primary interest to the current Thesis. As has previously been noted, the basic principle behind decomposition is that of "divide and conquer" (Raiffa, 1968). By decomposing a task into several smaller components or subsets, it allows the subject to concentrate on one thing at a time. Findings from MacGregor and Armstrong's 1994 study support this notion, they reported that as a decision-analytic tool, decomposition enables a complex task to be reduced into smaller and "cognitively more manageable parts" (p. 32). Thus by definition, decomposition may be seen to be decreasing the complexity of that task and in turn decreasing overall cognitive load. This was also suggested by Webby and O'Connor in their 1996 review of the literature (p. 106), and further supported by MacGregor's (2001) summary which deemed decomposition to be "... an effective strategy for improving the quality of judgemental forecasts" (p. 121), particularly for tasks with high uncertainty.

The notion of structure, and *the* division of tasks into "cognitively more manageable parts", was also one of the major focus points for Vessey's 1991 research in which she sought to explain performance outcome in a decision-making environment in terms of "cognitive fit". Citing Newell and Simon's 1970's work in the area of human information processing, Vessey purported that "since humans are limited information processors, more effective problem solving will result when the complexity in the task environment is reduced" (p. 220). The primary aim of Vessey's study was to find support for her theory of "cognitive fit" and in essence support the notion that task complexity can be reduced by both decomposing a task into several sub-components; *and* finding the best "mode of representation" for that particular task (p. 220). Vessey's work concentrated on decision-making tasks, identifying tasks involving judgement or inference as being "complex tasks", ones which can, in turn, be further decomposed into several subtasks (p. 225).

For the series of experimental studies undertaken as part of this Thesis, the *best* "mode of representation" for the time series extrapolation task has been deemed to be the graphical representation of time series data. For a more detailed discussion on the impact and use of Decomposition in time series extrapolation please refer to section 2.3.4.

3.4.7 General Instructions

Under all treatment conditions, task-doers were advised, that once they had completed predictions for one time series and moved on to the next, they were not to go back and revise predictions for any previously completed time series. This was able to be enforced in the screen-based treatment (via the software) and was requested of the task-doers in the paper-based treatment (please refer to Figures 2 and 3).

General instructions were provided to all task-doers under each of the treatment conditions, a copy of these instructions is presented as Appendix 5).

3.4.8 Subjects and Experimental Procedure

Students undertaking both undergraduate and postgraduate Degree Programmes at the University of New South Wales were used as subjects in the experimental studies. They were asked to either complete the task using the paper-based treatment (Tool 2), or one of the screen-based treatments (Tool 1, or Tool 3). In addition, subjects were allocated to undertake either the structured or the unstructured task.

Task-doers were randomly allocated to treatment groups, to the extent that they were randomly allocated to classes at the commencement of the university session.
Subjects participated voluntarily in all experiments and received no course credit or monetary rewards. Each task-doer participated in the experiment during their normal class times in one sitting, and typically took about twenty five (25) minutes to complete the task.

All *students* were undertaking courses in Information Systems at UNSW and were (arguably) highly proficient in the use of computers and related technology.

Student task-doers were used in all of the experimental studies, enabling the question of familiarity with the medium to be controlled. Findings from Lawrence *et al.*'s (1985) study, suggest that, at least in terms of accuracy, the results obtained from non-expert forecast subjects could be expected to be consistent with that of experts.

3.5 Analysis Methodology

The metric used to assess forecast accuracy and to compare treatments was the Mean Absolute Percentage Error (MAPE) calculated as follows:

$$\text{MAPE} = \frac{\Sigma\left[\,|y_t - \hat{y}_t| / y_t\right] \times 100}{n}$$

where y_t is the actual value of Y observed at time t, and $\hat{y}_t$ is the forecast value of Y for time t, and n is the period number.

The MAPE metric was employed as it is a common measure used to assess relative forecasting accuracy (Carbone and Armstrong, 1982; Armstrong, 1985; Lawrence *et al.*, 1985; Makridakis *et al.*, 1993). MAPE has advantages over squared error measures which are not unit-free and overly emphasise extreme errors (Chatfield, 1988).

Webby (1993) indicates that percentage error measures such as MAPE are generally preferred over squared because they control for scale. However, it has been noted by Armstrong and Collopy (1992) that MAPE penalises forecasts exceeding the actual, in relation to forecasts less than the actual – MAPE is bounded on the low side by 0%, but is unbounded on the high side.

To facilitate analysis of data in terms of time horizon, MAPE values were derived for "groups" of forecast periods; periods 1-6; periods 7-2; and All 12 periods. Given that the major emphasis of the current study is to further review the notion of TTF – "the degree to which a technology assists an individual in performing his or her portfolio of tasks" (Goodhue and Thompson, 1995 p. 216), and that task complexity is deemed to be one of the independent variables, data analysis will be performed separately for each individual time series. This will be done in order to ensure that each of the time series is considered a separate task, each with its own characteristics. Once again, this is in keeping with Goodhue and Thompson's definition of *task* – "actions carried out by individuals in turning inputs into outputs" (p. 216). In terms of the current study, the *output* is deemed to be the accuracy of predictions in a judgemental forecasting task.

In summary, the overall Research Question for this Thesis is to explore whether the interaction between the Tool –the *media* and the related *technology*, the Task – in terms of task complexity – and the People, has an effect on task performance – measured in terms of accuracy of predictions in a judgemental time series extrapolation task.

Each of the experimental studies will now be presented and discussed in detail in Sections 4 - 7.

4. Experiment One

4. Experiment 1

4.1 Introduction

The following section will describe the first experimental study which seeks to review the influence of *media* and the related *technology* on the accuracy of predictions in a judgemental extrapolation task.

The judgemental extrapolation of time series data was chosen as the experimental task because of a number of key factors. Firstly, it was deemed to be an essential component of many business decisions. Secondly, it was considered relevant in a decision making context, and thirdly, it was able to provide an objective measure of task performance (accuracy of forecast predictions).

This experimental study uses the general framework of task characteristics as developed by Campbell (1988), to categorise elements of the experimental task/s and determine the fundamental complexity of the task/s. In addition, relevant literature in relation to Task-Technology Fit; Task; Media; and the experimental task itself; the judgemental extrapolation of time series data – is reviewed and presented in section 4.2.

As previously noted in section 3, there are two major objectives for this Thesis. The first is to review the impact of *media* and related *technology* on task performance. The second is to seek to extend the area by conducting such research on a task that is deemed to be relevant in the decision-making context. As covered in section 2.4, the majority of research in the area of *media* and related *technology* – or more specifically, Task-Technology Fit, as it has become known since the mid 1990's – has tended to concentrate on two experimental tasks, namely reading for comprehension and proof-reading.

The current study seeks to extend this research area by introducing the judgemental extrapolation of time series data as a new experimental task.

Using Campbell's (1988) framework, we have previously classified our three tasks – reading, proof-reading and judgemental extrapolation (presented in sections 2.4.1 and 2.4.2). Experiment 1 will seek to provide some empirical evidence to support our hypothesis that there will be no difference in task performance across the two different *media* and related *technologies*. It will provide an objective measure of task performance – the accuracy of predictions as entered by subjects undertaking a judgemental extrapolation task – against which to test our hypothesis.

As discussed in section 2.4, Campbell's framework provides a classification for task complexity according to the presence or absence of specific characteristics. This has been adapted in order to classify the fundamental complexity of the three experimental tasks – reading; proof-reading; and the *object of interest* for the current study – judgemental extrapolation. A summary is presented in Table 4.1.1 below.

Table 4.1.1 The Fundamental Complexity* of the three experimental tasks: Judgemental extrapolation; Reading and Proof-Reading

Task	Outcomes Source 2	Interdependencies Source 3	Uncertainty Source 4	Level of Manipulation device Interaction
Judgemental Extrapolation	—	Moderate-High	Moderate-High	Moderate-High
Reading for Comprehension	Low	Moderate-High	Moderate-High	Low
Proof-Reading	Low	Low	Low	Moderate-High

*Source 1 (paths) has been omitted from Table 4.1, it was not present in any of the three experimental tasks

As can be seen from the above, our experimental task shares many "commonalities" with both reading and proof-reading. Our review of the literature in relation to *media* has indicated that reading accuracy does not deteriorate when screen-media is used compared to paper. However, the same cannot be said for proof-reading accuracy. General findings indicate that proof-reading accuracy is often compromised by the use of screen-media. Perhaps the answer lies in degree of TTF, Task-Technology Fit (Goodhue and Thompson 1995). TTF can best be described as the extent to which the technology helps the individual in performing their task, or more specifically as "the correspondence between task requirements, individual abilities, and the functionality of the technology." (p. 218).

When compared to reading, the complexity of the judgemental extrapolation task is very similar in relation to the level of uncertainty and the presence of conflicting interdependencies (for a more detailed discussion please refer to section 2.4.1.4). However, in terms of its relationship to the *technology*, i.e. the level of interaction with the manipulation device, our experimental task is more similar to that of proof-reading. Given the similarities with reading, one could speculate that the use of screen-media and the related technology should have no impact on the accuracy of predictions in our judgemental forecasting task. However, if we look to the level of interaction with the manipulation device, we could speculate that the use of screen-media and the related technology will have a negative impact on the accuracy of predictions in our judgemental forecasting task. Experiment 1 will seek to explore these issues further.

4.2 Research Design for Experiment 1

Experiment 1 used a 2 x (2) research design and sought to perform an analysis over time. The first factor (a between subjects factor) was the Tool – *media* and the related *technology* – and the second, the time horizon (a within subjects factor). The Analysis of Variance (ANOVA) statistical procedure was employed to undertake data analysis.

Experiment 1 sought to compare the accuracy of judgemental forecasting predictions produced using black and white paper-based time series representations, with the accuracy of those produced using black and white screen-supported time series display. An additional aim of experiment 1 was to determine the extent, if any, of differences between the accuracy of predictions across the time horizon. It is widely reported, and generally accepted that there is a tendency for forecasting accuracy to decrease as the forecast horizon increases (Dalrymple and King, 1981; Armstrong, 1985). It would be interesting to see if this tendency would be found to occur in experiment 1.

In terms of the interaction with the *technology* or the manipulation device, task-doers in the paper-based treatment condition were required to use a pencil and eraser in order to articulate their needs and register their forecast predictions. In the screen-supported treatment condition, task-doers were required to use a combination of slider, mouse and keyboard in order to articulate their needs and register their forecast predictions.

In terms of task complexity, this was held constant for experiment 1. Subjects were presented with an unstructured judgemental extrapolation task, no decomposition was provided and the task was deemed to be a "more" complex task.

In terms of people, or individual characteristics as referred to by Goodhue and Thompson (1995), this was also held constant for experiment 1 – only undergraduate task doers undertook the experiment.

A summary of the characteristics of the task and the various tools employed – *media* (treatment condition) and the related *technology* (manipulation device) – for experiment 1 is presented in Table 4.2.1 below:

Table 4.2.1 Experimental Tool/s and Task – Experiment 1

Tool_ID	Treatment Condition	Manipulation Device	Task Complexity
Tool 1*	Screen-based Black & white	combination of slider, mouse and keyboard	Unstructured more complex
Tool 2 *	Paper-based Black & white	pencil and eraser.	Unstructured more complex

* Graphic presentation of both Tools is provided at sections 4.4.3 and 4.4.4 respectively.

Given the many "commonalities" which our experimental task shares with both of the tasks previously undertaken in *media* studies, it is hypothesised that there will be no difference in task performance across the different *media* and related *technologies*. Hypothesis one states:

> **H1: There will be no difference in the accuracy of predictions in a judgemental forecasting task between paper-based and screen-based *media* and related *technologies*.**

4.3 Subjects and Experimental Procedure

Experiment 1 employed the use of one hundred and forty (140) undergraduate students. Eighty one (81) of the subjects completed the task using Tool 1 – screen-based *media* and related *technology* – and the remaining fifty nine (59) subjects completed the task using Tool 2 – paper-based *media* and related *technology*. The task was the judgemental extrapolation of time series data. There was no decomposition of time series representations and the task was deemed to be an "unstructured task" as well as being a "more complex" task (please refer to section 2.3.4 for a more detailed discussion).

All task information for one time series was provided on one page (i.e. one computer screen-full or one A4 landscape piece of paper). This enabled task-doers to concentrate on completing one time series at a time before moving onto the next.

In keeping with tasks undertaken in previous media studies, the experimental task was relatively straight-forward, quick to undertake and easily explained to non-expert task-doers. Historical data was graphically presented as a time series (i.e. a series of linked points or values for previous periods). This was in keeping with findings that novice forecasters are often as accurate as experts when this form of presentation is used (Lawrence, Edmundson and O'Connor's 1985).

In an effort to minimise any between-subject accuracy variations, no contextual information was provided (for a more detailed discussion please refer to section 2.3.3.4). Subjects were provided with only the historical time series and no information about what the series measured or the environment in which it was used. This choice both simplifies the task and enhances its similarity to the proof-reading and reading media studies where no effort was made to either match or not match the task-doers to the material being proofed or read.

Task-doers under each treatment condition were given a standard introduction to the experiment and a prediction demonstration using a sample time series (these are presented as Appendix 5 and 6e). During the introduction, task-doers were advised that the results were purely for the basis of experimentation rather than any contribution to course credits. In the screen-based treatment, task-doers were encouraged to familiarise themselves with the software tool and related *technology* before commencing with the main experiment. To do this, they utilised the software tool and related *technology* to record example predictions for the same sample time series as used in the demonstration. Similar familiarisation with the medium was not considered necessary for task-doers undertaking the paper-based experiment – as it was considered "natural" for task-doers to interact with paper and the related *technology,* namely a pencil and an eraser (Hansen and Haas, 1988, p.9).

Graphical depictions of how the time series were presented to task-doers in each of the respective treatment conditions are provided in the following sections.

4.3.1 Tool 1: *Media* – screen-based; *Technology* – slider, mouse and keyboard *Task Complexity* – unstructured, more complex

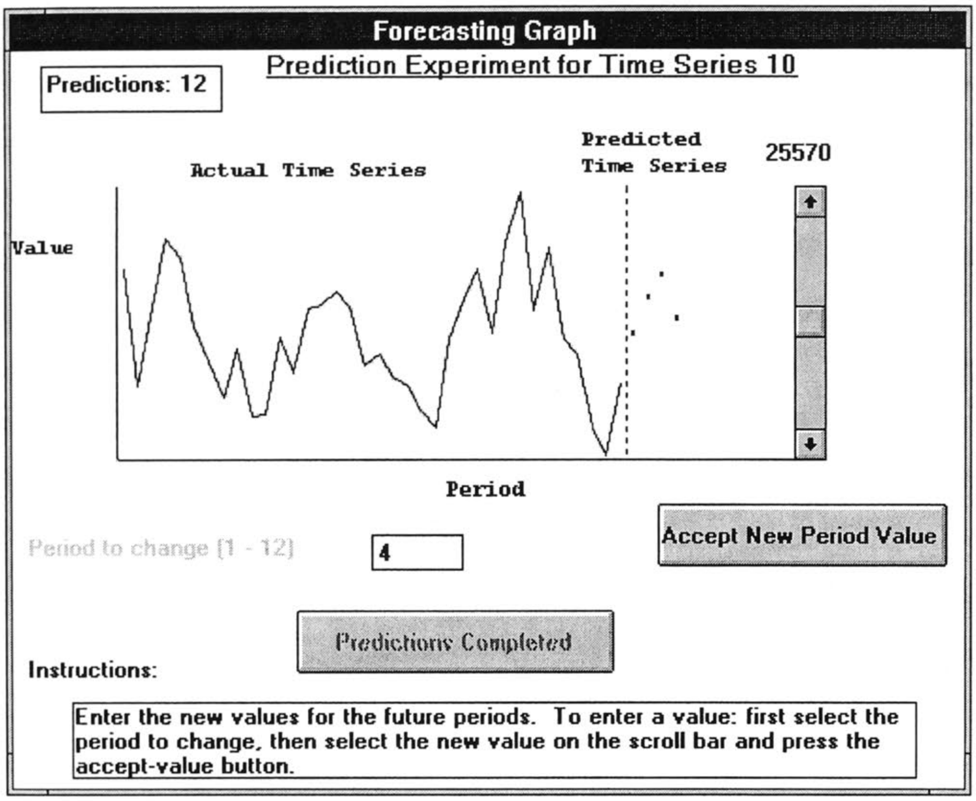
A representation of the screen-based medium is presented as Figure 4.3.1. To record predictions, task-doers were required to use the keyboard to enter the period number and then the mouse to select the predicted value for that period using the slider (next to the time series).

To complete the marking of the value on the graph, task-doers clicked on the "Accept New Period Value" button. This process was required for each of the twelve (12) period values to be predicted and for any subsequent prediction adjustments. After all twelve (12) periods had been predicted and accepted, the "Predictions Completed" button was enabled, once this button was clicked, task-doers moved onto the next time series. Once all period values for the final time series were entered and accepted, task-doers were thanked for their participation in the experiment. The software was disabled and the task completed.

Figure 4.3.1.1

Time Series Presentation[1] in the Computer-supported Treatment.

Tool 1 – Experiment 1



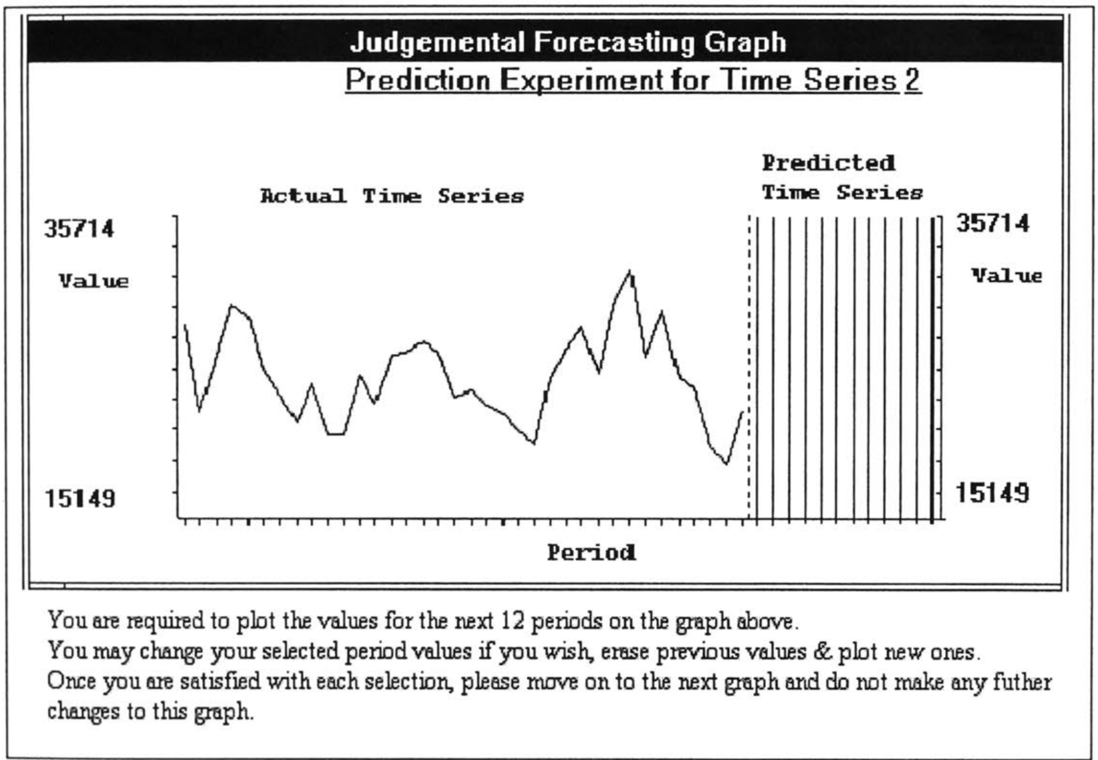
[1] Picture is smaller than actual size presented to task-doers.

4.3.2 Tool 2: *Media* – paper-based; *Technology* – pencil and eraser *Task Complexity* – unstructured, more complex

Prediction recording in the paper-based treatment was much less elaborate than in the screen-based medium. A representation of the paper-based medium is presented as Figure 4.3.2.1. Task-doers marked their predicted values on the graph image using a pencil, and, after carrying out any adjustments (using an eraser), moved onto the next time series.

Once all period values for the final time series were entered and accepted, task-doers were asked to raise their hand so that their paper-based submissions could be collected. They were advised that the task had been completed and were thanked for their participation in the experiment.

Figure 4.3.2.1 Time Series Presentation[2] in the Paper-based Treatment Tool 2 – Experiment 1



[2] Picture is smaller than actual size presented to task-doers.

4.4 Analysis Methodology

The metric used to assess forecast accuracy and to compare treatments was the Mean Absolute Percentage Error (MAPE). The MAPE metric was employed as it is a common measure used to assess relative forecasting accuracy (Carbone and Armstrong, 1982; Armstrong, 1985; Lawrence *et al.*, 1985; Makridakis *et al.*, 1993). The Analysis of Variance (ANOVA) statistical procedure was employed to facilitate data analysis. To facilitate analysis of data in terms of time horizon, MAPE values were derived for "groups" of forecast periods; periods 1-6; periods 7-12; and All 12 periods.

4.5 Results

4.5.1 Overall Results

The data collection process recorded predictions made by one hundred and forty (140) subjects. Eighty one (81) of the subjects completed the task using Tool 1 – screen-based *media* and related *technology* – and the remaining fifty nine (59) subjects completed the task using Tool 2 – paper-based *media* and related *technology*. In accordance with the hypotheses presented, data was collected for each treatment condition and analysed across time-horizon. A summary of overall results for all four (4) time series and across both treatment conditions is provided below in Table 4.5.1.1.

Table 4.5.1.1 Overall Results Summary – A Comparison of Tool 1 & Tool 2 Average MAPEs

Time Series	Time Horizon	Tool 1 Screen	Tool 2 Paper	p	Accuracy
1	First 6 periods	23.66	20.46	**0.0019**	**P > S**
B17	Last 6 periods	32.21	27.91	0.0606[1]	p > s
	All 12 periods	27.93	24.18	**0.0033**	**P > S**
2	First 6 periods	11.70	9.72	**0.0126**	**P > S**
B26	Last 6 periods	18.96	17.13	0.1475	**p ≅ s**
	All 12 periods	15.33	13.42	**0.0113**	**P > S**
3	First 6 periods	119.03	129.44	0.2831	**p ≅ s**
B8	Last 6 periods	68.26	74.57	0.2288	**p ≅ s**
	All 12 periods	93.65	102.01	0.1752	**p ≅ s**
4	First 6 periods	21.89	14.64	**0.0000**	**P > S**
M6	Last 6 periods	32.61	28.92	**0.0280**	**P > S**
	All 12 periods	27.25	21.78	**0.0000**	**P > S**

Note: Significance Values: $p < 0.05$ are emboldened.

[1] The paper treatment sample was non-normal; however a Kruskal-Wallis test confirmed the original Anova result that media effects were non-significant over this time horizon.

LEGEND: P = Paper; S = Screen;
P > S indicates paper is significantly better than screen
p ≅ s indicates paper and screen overall equivalent
p > s indicates paper is better than screen but not significantly so

One of the major aims of the current study is to further review the extent to which the technology helps the individual in performing their task, or the degree of TTF as described by Goodhue and Thompson (1995). For this experiment, the degree of TTF is measured by the accuracy of predictions and task complexity has been deemed to be one of the independent variables. In order to gather as much data as possible against which to measure task performance or the "degree" of TTF, data analysis was performed separately for each individual time series. Although not generally in keeping with analysis of data commonly undertaken in previous studies in judgemental extrapolation of time series data, this was done in order to ensure that each of the time series was considered a separate task, each with its own characteristics. Once again, this is in keeping with Goodhue and Thompson's definition of *task* – "actions carried out by individuals in turning inputs into outputs" (p. 216). In terms of the current study, the *output* is deemed to be the accuracy of predictions in a judgemental forecasting task.

However, in using MAPE, a metric commonly used to assess forecast *accuracy* across time series, as our accuracy measure, it would appear unusual to the reader for experiment 1 to analyse and compare results separately for each of the time series. It must be noted that given the magnitude of the MAPE error reported for the third time series (B8), the pooling of all time series data would have been dominated by B8. In effect, the huge error for B8 would have masked any *media* effects which may have been reported across the other three time series.

4.5.2 H1: No Difference in Accuracy

Our Hypothesis for experiment 1 predicted that would be no difference in the accuracy of predictions in a judgemental forecasting task between paper-based and screen-based *media* and related *technologies*.

As the results in Table 4.5.1 show, H1 is not supported. The *media* and the related *technology* clearly does have an effect on the accuracy of predictions in judgemental extrapolation. Clearly accuracy for task-doers using Tool 2 – the paper-based *media* and related *technologies* – was superior to those using Tool 1 – the screen-based *media* and related *technologies*. Findings support the notion that task-doers are better at predicting the accuracy of a forecast using pencil and paper than they are using screen-supported time series displays. This is in direct contrast to the widespread belief that the use of computers can potentially improve forecasting accuracy (Dalrymple 1987).

With the exception of the third time series presented, B8, all of the results indicate superiority of Tool 2 (paper-based) over Tool 1 (screen-based) across the 1-6 periods time horizon grouping. In addition, for time series four (M6), the predictions for Tool 2 (paper-based) were significantly superior to Tool 1 (screen-based) over both of the time horizon groupings (please refer to table 4.5.1).

In terms of reviewing the accuracy of predictions across time horizon groupings, the results support those widely reported in the forecasting literature – that forecasting accuracy decreases as the time horizon decreases. Generally, accuracy across periods 7-12 was found to be inferior compared to accuracy for periods 1-6, although not significantly so. This was found to occur across both treatment conditions (Tool 1 and Tool 2) and for all of the time series expect the third – B8. Something peculiar appears to be happening with the accuracy of predictions across both treatment conditions with time series B8. It is the only time series which reported a significant accuracy gains across the 7-12 period time horizon grouping for both paper-based and screen-based media. A review of the attributes of B8 indicate that it is a seasonal time series with extreme peaks and troughs (a graphical presentation of the time series can be found at Appendix 6).

The results for experiment 1 indicate that task-doers experienced considerable difficulty in forecasting this time series, under both treatment conditions. In addition, it appears that inaccurate predictions by task-doers may have completely overwhelmed any media effects for this series. A factor that has been identified in a number of studies as contributing to forecasting difficulties, is the presence of instability in the time series, or more specifically, the amount and direction of *ramping* in a time series. As presented in O'Connor *et al.*'s (1993) study, *ramping* refers to a consistent trend upward or downward over a number of consecutive periods (i.e. the presence of marked peaks and throughs within the time series). Goodwin (1998) reports that several studies have indicated that forecasters tend to underestimate upward trends.

In contrast, O'Connor, Remus and Griggs (1997) describe that people have significant difficulties in forecasting downward sloping series because they tend to anticipate that the downward trend will reverse itself. In hindsight, given the nature of the time series and the related attributes, especially the marked peaks and throughs, perhaps it was not an ideal choice as an experimental task. In an effort to minimise the potential contribution of confounding variables, it was decided to exclude the third time series, B8 from further experimental studies conducted as part of the current research.

4.6 Discussion

4.6.1 Main Findings

The experimental results indicate, counter to what was hypothesised, that for three (3) of the four (4) time series presented, the use of a paper-based medium and related technology (Tool 2) for judgemental extrapolation, gave superior accuracy over the screen-medium and related technology (Tool 1). In fact, paper was significantly superior across at least one of the time horizon groupings for each of the (4) series. In addition, as expected, results show that forecasting accuracy reduces in the last six (6) periods (7-12) compared to the first six (6) periods (1-6). This was found to hold true for three (3) of the four (4) time series across both treatment conditions.

As was previously advised, the people – individual characteristics (Goodhue and Thompson 1995) – independent variable was held constant for experiment 1. Only undergraduate task-doers undertook the experiment. Given the nature of the experiment 1, utilizing two different Tools – *media* and the related *technology* – in order to determine forecast accuracy, findings from Whitecotton's (1996) study may provide some insight. Essentially, they indicated that when using a decision aid, experience had a positive impact on forecast accuracy.

Given that we used undergraduate subjects in experiment 1, perhaps it could be said that the nature of the subjects and their limited experience may have influenced the accuracy of their predictions. It would be interesting to see whether the use of more experienced or postgraduate subjects would have produced a different result. A question which needs to be posed is whether *media* and the related *technology* – i.e. presentation and working medium – will have an impact on the accuracy of predictions made by postgraduate subjects in the judgemental extrapolation task.

As previously noted, four monthly time series were used for the task in experiment 1. The time series were randomly selected from the M-Competition database comprising 111 real time series (Makridakis *et al.,* 1982). This database has been used as a source of real time series in other forecasting research (Davey and Flores 1993; Edmundson and Terry 1986; Lawrence *et al.,* 1985; Makridakis and Winkler 1983). It was our desire to have each group of subjects predict the accuracy of forecasts for all four time series under each of the treatment conditions. In order to minimise any possible task fatigue effects, it was decided to only select and display a total of four time series to subjects as part of the experimental task.

A comparative analysis of the results across all of the experimental tasks is presented in Table 4.6.1.1 below.

Table 4.6.1.1 A Comparison of Accuracy Results and Fundamental Complexity of the three experimental Tasks – Judgemental Extrapolation; Reading and Proof-reading.

Task	Fundamental Task Complexity Campbell's Model	Level of Manipulation Device Interaction	Results Accuracy
Judgemental Extrapolation	moderate-high uncertainty and interdependencies Unstructured – "more complex"	Moderate-High	P > S
Reading for Comprehension	Moderate-high uncertainty and interdependencies	Low	p ≅ s
Proof-Reading	Low uncertainty and interdependencies	Moderate-High	P > S

LEGEND: P = Paper; S = Screen;
P > S indicates paper is usually better than screen and frequently significantly better
p ≅ s indicates paper and screen overall equivalent.

It is interesting to note that in terms of task accuracy, the results of judgemental extrapolation are similar to those of proof-reading. Perhaps the level of interaction with the manipulation device played a major role or had a major impact on task performance. Additional experimentation is required in order to review this further.

It is also interesting to note that in terms of "fundamental task complexity" – using Campbell's framework – the task of judgemental extrapolation is more closely linked to that of reading for comprehension. Given this association, we might have expected accuracy results to be similar across both tasks. However, perhaps this proposed association was adversely affected by the nature of the judgemental extrapolation task – deemed to be Unstructured and thus "more complex". The question needs to be posed;

> Would the results have been different if the judgemental extrapolation task had been "Structured" and deemed to be "less complex"?

Once again, additional experimentation is required in order to review this further.

The results for experiment 1 show that task-doers experienced considerable difficulty in forecasting the third time series (B8), under both treatment conditions. It appears as though inaccurate predictions by task-doers may have completely overwhelmed any media effects for this series. As previously discussed, perhaps the very nature of the series – especially the marked peaks and troughs – may have contributed to this confusion. The B8 time series will therefore be excluded from further experimental studies undertaken as part of the current research.

4.7 Limitations

A number of limitations within the experimental design for experiment 1 must be acknowledged and possible implications discussed.

4.7.1 Non-Randomisation of Time Series

Examining the design of experiment 1 with the perspective of hindsight, it can be seen that the presentation of time series in the same order to each task-doer represents a potentially confounding factor. In mitigation, media effect results are largely consistent among time series except for B8, the third time series, which was also the middle one predicted by task-doers each of the treatment conditions. This supports the notion that the media effects found, are unlikely to be a consequence of presentation order. However, the influence of non-randomisation should certainly be clarified by similar future experiments by presenting time series to task-doers in a random order.

4.7.2 Time Series Characteristics

A review of the characteristics of the time series used in experiment 1 reveals that one of the series is non-seasonal (B17), whilst the other three (B26, B8 and M6) are all seasonal. Once again, with the benefit of hindsight, it can be seen that in order to manipulate Task Complexity (an independent variable) in future experimental tasks, it would be beneficial to have had an even number of both seasonal and non-seasonal series (please refer to section 2.4 for a review of the proposed Research Model). This will be accommodated in future experimental studies by the omission of the third time series, B8 (for reasons as discussed in section 4.5.2 above). It will be replaced by a non-seasonal time series to facilitate future experimental studies conducted as part of this research, allowing an equal number of both seasonal and non-seasonal time series to be presented to task-doers.

4.7.3 External Validity

The use of Tools – screen-supported and paper-based – in order to simulate what would generally be a task undertaken in a real-life environment, under real-time conditions, may have had an adverse effect on the study's external validity. In general, decision making is not undertaken under such strict conditions. However, in defence, the time series which were used in experiment 1 were all "real" time series with actual values for the periods to be predicted. In addition, the use of a controlled experiment allows for other external and potentially confounding factors to be controlled, thus helping to maximise the generalisation of results.

4.7.4 Undergraduate Subjects

Task-doers who participated in experiment 1 were undergraduate students undertaking Information Systems courses at UNSW. There may be inherent differences in the decision strategies involved in the judgemental extrapolation task of this select group, compared to both the general population and also other student populations e.g. postgraduate students. The potential influence of this factor should certainly be reviewed by conducting future similar experiments using postgraduate task-doers.

4.8 Future Research

The findings in experiment 1 and the limitations identified in section 4.7 above lead us to a number of questions which could be investigated in future research. Given that forecasting accuracy in the screen-supported medium (Tool 1) was inferior in comparison to forecasting accuracy in the paper-based medium (Tool 2), it would be interesting to undertake further research in this particular area. The question which must be asked is whether forecasting accuracy is still compromised in the screen-supported medium when the task is undertaken by more experienced task-doers. As already mentioned, Lawrence *et al.*'s (1985) study indicates that novice forecasters, in the paper medium, are largely as accurate as experts. The study's findings also suggest that results obtained from non-expert forecast subjects could be expected to be consistent with that of experts – at least in terms of forecast accuracy for the paper-based medium. It would be interesting to see if the same were to hold true for postgraduate, or more experienced task-doers, this will be addressed in the next experiment.

In terms of task performance in relation to the use of a decision aid, we can look to findings from Whitecotton's 1996 study. Results indicated that when using a decision aid, experience had a positive impact on forecast accuracy. Again, it would be interesting to see if a similar situation existed when postgraduate task-doers undertook the task utilizing Tool 1 – the screen-supported *media* and related *technology*. This is also one of the key components of this Thesis. As outlined in the Research Model presented in section 2.4, the characteristics of the people, or the individuals undertaking the task, are an integral component of this Thesis. Having identified people as being one of the independent variables of the current study, it would be interesting to review the results of further experimentation, using more experienced subjects, and holding all other factors constant. This is also in keeping with Goodhue and Thompson's (1995) theoretical model of TTF – another crucial component of the current study.

Perhaps future research should investigate whether the use of postgraduate subjects would have produced a different result. A question which needs to be posed is:

> Will the *media* and the related *technology* – i.e. presentation and working medium – have an impact on the accuracy of predictions made by postgraduate subjects in the judgemental extrapolation task?

4.9 Conclusion

Experiment 1 has explored differences in the accuracy of predictions in a judgemental extrapolation task when that task is presented across two different presentation and working media; namely paper-based and screen-supported. In addition, experiment 1 sought to determine whether the accuracy of predictions in a judgemental extrapolation task differed across the time horizon groupings (periods 1-6 and periods 7-12) for the presented series.

The main finding of experiment 1 is that judgemental extrapolation, in common with proof-reading, is a task which is often more accurately undertaken in the paper-based medium rather than the screen-supported medium. Experiment 1 has also shown that forecast accuracy decreases over time horizon under each of the treatment conditions and that the decrease in accuracy is generally more pronounced in the screen-supported medium.

These findings are, of course, qualified by the novice forecasters and simplified version of the forecasting task used in experiment 1. Findings from experiment 1 and limitations as identified by the analysis of data and experimental procedures, suggest that the use of postgraduate subjects is an area that could be addressed by future research efforts. This is the basis for undertaking experiment 2 which will be discussed in section 5.

5. Experiment Two

5. Experiment 2

5.1 Introduction

The following section will describe the second experimental study which seeks to review the influence of *media* and the related *technology* on the accuracy of predictions in a judgemental extrapolation task using postgraduate subjects.

The general results of experiment 1 showed that (undergraduate) subjects undertaking a judgemental extrapolation task in the paper-based *media* and related *technology* (Tool 2) produced more accurate forecast predictions than subjects utilizing the screen-medium and related technology (Tool 1). This was contrary to our proposed hypothesis that there would be no difference in task performance across the two different *media* and related *technologies.* Following on from those general results, and with particular reference to the proposed Research Model for this Thesis (presented in section 2.4), we were interested in testing whether the same would hold true if the independent variable – *people* – were changed for experiment 2. With this in mind, experiment 2 sought to replicate all aspects of experiment 1, with the exception of the "people" variable – experiment 2 utilized postgraduate subjects.

Essentially, it is the "experience", which has been identified as a "people" characteristic in Goodhue and Thompson's model of TTF (1995). This has been incorporated into the proposed Research Model for the current Thesis (section 2.4). However, the focus has moved away from the traditional notion of "experience" which has generally been linked to the task itself. This is required in order to recognise that subjects undertaking the task under the different treatment conditions – the *media* and related *technologies* – have to also be familiar with the different forms of *media* and the *technologies* which they will be exposed to. With this is mind, we decided to concentrate on another individual characteristic which we believe to be more appropriate indicator for the experimental studies undertaken as part of the current research. It is believed that the "age/maturity" characteristic of the people undertaking the task will provide a better indication of "experience" – particularly in relation to the *media* and the related *technologies*. For the purpose of the current study, this emphasis provides the basis for conducting experiment 2 using postgraduate subjects who generally have a greater age and level of maturity than undergraduate subjects.

This is also in keeping with previous studies in the area of presentation *media* and related *format* (i.e. graphs versus tables), identified as another area of interest in section 2.3. As has generally been reported, experience with the media – in particular the screen-based media – has been found to have an impact on overall task performance. More specifically, Lucas' 1981 study, which reported that a lack of overall experience with the screen-based media had an adverse effect on overall task performance. This has also been reported by Kagan and Pietron (1987), prior experience with computers or knowledge of computers was found to be an important predictor of overall task performance. Kasper and Cerveny's (1985) study of user-characteristics, in relation to decision-making performance, having previously identified computer familiarity as an important contributor capable of enhancing the decision-making process.

In addition, as identified by Webby and O'Connor's (1996) review of the judgemental forecasting literature, the general area of subjects' "experience" or expertise with the task, and forecast accuracy, has tended to produce non-specific results. The authors advise that evidence from the literature is not easy to interpret, and that is difficult to determine or identify the impact on forecast accuracy that can be directly attributed to the subjects' level of experience or expertise. Referencing Remus' 1990 study, they propose that, similar to the evidence gathered from a variety of disciplines, such as psychology and finance,

> "...which generally indicates that expertise beyond a certain minimal level has little incremental value ...perhaps this also holds true in time-series forecasting – maybe total novices (e.g. undergraduates) will be inept at the task, but subjects with a limited amount of training (e.g. graduates) will be just as good as the experts, e.g. managers." (p. 97).

Citing the work undertaken by Edmundson *et al.,* 1988 and Sanders and Ritzman, 1992, Webby and O'Connor further maintain that "... it appears that experience ... has no effect on forecast accuracy." (p. 97) – the authors qualify this statement, adding that "...those two studies varied experience within a time series only task." (p. 97).

Looking to Whitecotton's (1996) study, subjects were required to forecast earnings for 16 firms under two treatment conditions – one with, and one without access to a decision aid. Whitecotton reported a positive result in terms of the association between experience and accuracy. Using subjects with differing levels of "task experience" – financial analysts, postgraduate students (MBA) and undergraduate students – Whitecotton reported a higher accuracy for postgraduate subjects and financial analysts, compared to the undergraduates. In addition, the use of the decision aid improved accuracy across all groups. However, as previously noted, although the result of Whitecotton's study are of interest to the current research, her use of "experience" or "expertise" was directly related to the task being undertaken. For the purpose of this Thesis, the distinction between postgraduate and undergraduate subjects will be more closely related to their "age/maturity". In turn, this can be directly linked to their level of experience with the *media* and the related *technologies* – with specific emphasis on the screen-supported media and related technologies.

As previously noted in section 3.4.8, the use of student subjects enabled the question of familiarity with the medium to be controlled. In addition, the use of postgraduate student subjects in experiment 2 sought to both test a specific area of the proposed research Model for the current Thesis – individual characteristics – as well as specifically address one of the limitations of experiment 1.

5.2 Research Design for Experiment 2

Experiment 2 used a 2 x (2) research design and sought to perform an analysis over time. The first factor (a between subjects factor) was the Tool – *media* and the related *technology* – and the second, the time horizon (a within subjects factor).

Postgraduate subjects undertook the task under each treatment condition. As for experiment 1, the Analysis of Variance (ANOVA) statistical procedure was employed to undertake data analysis.

Using postgraduate subjects, experiment 2 sought to compare the accuracy of judgemental forecasting predictions produced using black and white paper-based time series representations, with the accuracy of those produced using black and white screen-supported time series display. In addition, it was our wish to further examine any differences in the reported results compared to those previously reported for experiment 1. A direct comparison of results would help to further strengthen our resolve to extend the area in relation to TTF. Our aim is to achieve this by conducting a series of controlled laboratory experiments across different *media* and related *technology* and on a task that will; provide an objective measure of performance; and is widely used in a business environment. This task is the judgemental extrapolation of time series data.

An additional aim of experiment 2 was to determine the extent, if any, of differences between the accuracy of predictions across the time horizon. Although strictly speaking, our motivation was to review the accuracy of predictions by time horizon *across* the two different treatment conditions, rather than within any one specific treatment condition. As previously reported in section 3.1.3, we are interested not in the influence of time horizon *per se*: we are only interested in horizon as it interacts with the accuracy of various tools – the *media* and the related *technologies*. As has been widely reported, and generally accepted, there is a tendency for forecasting accuracy to decrease as the forecast horizon increases (Dalrymple and King, 1981; Armstrong, 1985). It would be interesting to see if this tendency would be found to occur in experiment 2 (as it was in experiment 1) and to identify the treatment condition/s under which this was found to be more prominent.

In terms of the interaction with the *technology* or the manipulation device, task-doers in the paper-based treatment condition, were required to use a pencil and eraser in order to articulate their needs and register their forecast predictions. In the screen-supported treatment condition, task-doers were required to use a combination of slider, mouse and keyboard in order to articulate their needs and register their forecast predictions.

In terms of task complexity, this was held constant for experiment 2. Subjects were presented with an unstructured judgemental extrapolation task, no decomposition was provided and the task was deemed to be a "more" complex task.

As previously noted, the focus of experiment 2 was also the people – in particular the age/maturity of the subjects undertaking the task. Given that the proposed focus of the current research is to extend the area in relation to TTF, it is believed that the "age/maturity" characteristic of the people undertaking the task will provide a better indication of "experience" – particularly in relation to the *media* and the related *technologies*. In addition, given that generally, subjects' "experience" has been linked directly to the task itself, the use of age/maturity in order to identify an essential element of the subjects' characteristics, we believe, eliminates any potential confusion. Therefore, in terms of people, or individual characteristics as referred to by Goodhue and Thompson (1995), this was also held constant for experiment 2 – only postgraduate task doers undertook the experiment.

In summary, the characteristics of the task and the various tools employed – *media* (treatment condition) and the related *technology* (manipulation device), for Experiment 2 are presented in Table 5.2.1 below:

Table 5.2.1 Experimental Tool/s and Task
Experiment 2 (Postgraduate Subjects)

Tool_ID	Treatment Condition	Manipulation Device	Task Complexity
Tool 1*	Screen-based Black & white	combination of slider, mouse and keyboard	Unstructured more complex
Tool 2 *	Paper-based Black & white	pencil and eraser.	Unstructured more complex

* Graphic presentation of both Tools is provided at Sections 4.4.3 and 4.4.4 respectively.

Given the many "commonalities" which our experimental task shares with both of the tasks previously undertaken in *media* studies, it is hypothesised that there will be no difference in task performance across the different *media* and related *technologies* when postgraduate subjects undertake the task. Hypothesis one states:

> **H1: There will be no difference in the accuracy of predictions in a judgemental forecasting task between paper-based and screen-based *media* and related *technologies*, when the task is undertaken by postgraduate subjects.**

5.3 Subjects and Experimental Procedure

Experiment 2 employed the use of one hundred and one (101) postgraduate students undertaking Degree Programmes at the University of New South Wales. Forty three (43) subjects completed the task using Tool 1 – screen-based *media* and related *technology* – and remaining fifty eight (58) subjects completed the task using Tool 2 – paper-based *media* and related *technology*. As for experiment 1, the task was the judgemental extrapolation of time series data. There was no decomposition of time series representations for experiment 2. Subjects participating in the experiment were presented with the "unstructured task" – which was also deemed to be a "more complex" task (for a more detailed discussion please refer to section 2.3.4).

Experiment 2 sought to address some of the limitations within the experimental design for experiment 1. One such limitation being, any effects which may have been attributable to the non-randomisation of time series data (as discussed in section 4.7.1). In order to minimise any potential effects, time series in experiment 2 were presented to task-doers in a random order.

In addition, as discussed in sections 4.5.2 and 4.6.1, task-doers experienced considerable difficulty in forecasting time series B8 under both treatment conditions (Tool 1 and Tool 2). With the benefit of hindsight, the inclusion of time series B8 – with its marked peaks and troughs – may not have been a good choice as an experimental task. It was therefore decided that B8 would be excluded from any further experimental studies. However, in order to ensure a balanced experiment, there was a need to replace B8 with another randomly selected time series from the M-Competition database. Similar to the other three time series, F3 was a monthly time series which had forty eight (48) actual period values available. Thirty six (36) of these were displayed to subjects as historical data. As for all other time series, task-doers were asked to predict the next twelve periods. It was hoped that the omission of B8 and inclusion of F3 would not prove to be problematic for task-doers.

All other experimental procedures for experiment 2 were identical to those for experiment 1 – as detailed in section 4.3. This also included the Tools used in the experiment, thus the only differences between the two experimental studies being;

- the "people" – use of postgraduate subjects;
- randomisation of presentation order of time series to subjects; and
- the exclusion of time series B8 and the inclusion of time series F3.

For a detailed discussion of experimental procedures, please refer to section 4.3.

5.4 Analysis Methodology

As for experiment 1, the metric used to assess forecast accuracy and to compare treatments was the Mean Absolute Percentage Error (MAPE), employed as it is a common measure used to assess relative forecasting accuracy (Carbone and Armstrong, 1982; Armstrong, 1985; Lawrence *et al.*, 1985; Makridakis *et al.*, 1993). The Analysis of Variance (ANOVA) statistical procedure was employed to compare the means of the various MAPE samples derived in experiment 2. To facilitate analysis of data in terms of time horizon, MAPE values were derived for "groups" of forecast periods; 1-6; 7-12; and All 12 periods (periods 1-12).

5.5 Results

5.5.1 Overall Results

The data collection process recorded predictions made by one hundred and one (101) postgraduate subjects. Forty three (43) of the postgraduate subjects completed the task using Tool 1 – screen-based *media* and related *technology*, and the remaining fifty eight (58) postgraduate subjects completed the task using Tool 2 – paper-based *media* and related *technology*. In accordance with the hypotheses presented, data was collected for each treatment condition and across both time-horizons. A summary of overall results for all four (4) time series and across both treatment conditions is provided below in Table 5.5.1.1

Table 5.5.1.1 Overall Results Summary – for Postgraduate Subjects A Comparison of Tool 1 and Tool 2 Average MAPEs

Time Series	Time Horizon	Tool 1 Screen	Tool 2 Paper	p	Accuracy
B17	First 6 periods	21.80	20.25	0.0747	p ≅ s
	Last 6 periods	26.56	25.30	0.5411	p ≅ s
	All 12 periods	24.18	22.77	0.2146	p ≅ s
B26	First 6 periods	9.83	9.86	0.9816	p ≅ s
	Last 6 periods	15.83	19.21	**0.0218**	**S > P**
	All 12 periods	12.83	14.53	**0.0321**	**S > P**
M6	First 6 periods	13.55	11.96	0.2773	p ≅ s
	Last 6 periods	23.36	29.77	**0.0015**	**S > P**
	All 12 periods	18.46	20.86	0.0673	s > p
F3	First 6 periods	2.83	2.32	0.1257	p ≅ s
	Last 6 periods	4.13	4.02	0.8460	p ≅ s
	All 12 periods	3.48	3.17	0.4126	p ≅ s

Note: Significance Values: $p < 0.05$ are emboldened.

5.5.2 H1: No Difference in Accuracy for Postgraduate subjects

Hypothesis 1 for experiment 2 predicted that there would be no difference in the accuracy of predictions in a judgemental forecasting task between paper-based and screen-based *media* and related *technologies* when the task is undertaken by postgraduate subjects.

As the results in Table 5.5.1.1 show, H1 is not supported. The *media* and the related *technology* clearly does have an effect on the accuracy of predictions in judgemental extrapolation. Clearly for two of the time series presented, B26 and M6, accuracy for subjects using Tool 1 – the screen-based *media* and related *technology* – was superior to those using Tool 2 – the paper-based *media* and related *technology*. Whilst for the other two time series presented, B17 and F3, there was no reported difference between the accuracy of predictions for each treatment condition (Tool 1 ≅ Tool 2). Therefore, findings for two of the four time series presented, (B26 and M6), support the notion that postgraduate subjects produced more accurate forecast predictions using the screen-supported time series displays than they did using pencil and paper. This is in keeping with the widespread belief that the use of computers can potentially improve forecasting accuracy (Dalrymple 1987) and is a totally different result than that produced by experiment 1 – where the results for B26 and M6 indicated clear superiority of Tool 2 – the paper-based media.

However, the results for experiment 2 appear equally divided, with two of the time series – B26 and M6 – indicating significant superiority of Tool 1 (screen-based) over Tool 2 (paper-based) across the 7-12 period time horizon grouping. Whilst the other two time series – B17 and F3 – reported no significant difference in terms of accuracy across the two treatment conditions for the 7-12 period time grouping horizon. It is interesting to note, that for time series B26, the significance level (for the 7-12 period grouping) was so great, that it rendered the accuracy of predictions for Tool 1 (screen-based) to be significantly superior to Tool 2 (paper-based) over the All12 periods, time horizon grouping (i.e. periods 1-12). Although this was not replicated for time series M6, the accuracy of predictions for Tool 1 was superior to those of Tool 2 over the All 12 periods time horizon grouping – although not significantly so. For a full review of reported statistics, please refer to Table 5.5.1.1.

In reference to the accuracy of predictions over the 1-6 period time horizon grouping, results for experiment 2 are uniform across all of the four time series. As indicated in Table 5.5.1.1, there is no reported significant difference between the accuracy of predictions by postgraduate subjects using Tool 1 (screen-based) over the accuracy of predictions by postgraduate subjects using Tool 2 (paper-based) across the 1-6 period time horizon grouping. This is reported for all four time series – B17, B26, M6 and F3 (a graphical presentation of all the time series can be found at Appendix 6).

In terms of reviewing the accuracy of predictions across time horizon groupings, for the 7-12 period grouping, results are mixed. For two of the four time series – B26 and M6 – accuracy across the 7-12 period time horizon grouping was found to be significantly superior in the screen-based treatment condition, compared to the paper-based treatment condition. However, perhaps this can largely be attributed to the individual characteristics of the time series itself rather than any perceived interaction effect between the time series and the Tool.

However, the question remains, why are the results between experiment 1 and experiment 2 so markedly different? In addition, there are further questions which arise, such as; Whether the difference in results can be solely attributed to the difference in the "individual characteristics" variable – i.e. undergraduate students versus postgraduate students? Or, whether the difference in results can be attributed to some other variable, such as the Tool? In order to seek to provide some plausible explanations for the vastly different results, it was decided to conduct some further post-hoc testing. Details of this additional analysis are provided in the following sections below.

5.5.2.1 Postgraduate subjects versus Undergraduate subjects

Further data analysis was conducted to determine whether the reported differences between Tool 1 and Tool 2 (experiment 2) could be directly attributed to "individual characteristics"– people and their age/maturity, or level of experience with the *media* and the related *technologies*. This analysis involved a direct comparison of data collected from postgraduate subjects undertaking a judgemental forecasting task, against data collected from undergraduate subjects undertaking the same task.

Data collected from one hundred and one (101) postgraduate subjects who completed the judgemental extrapolation task was compared to that collected from one hundred and forty (140) undergraduate students also completing the judgemental extrapolation task. Both groups of subjects used both Tool 1 – screen-based *media* and related *technology*, and Tool 2 – paper-based *media* and related *technology* in order to complete the task.

Data was collected for postgraduate and undergraduate subjects across both time horizons and pooled across Tools – *media* and related *technologies*. Due to the fact that time series B8 was excluded from the experimental design for experiment 2, and that F3 was introduced as the fourth experimental task to be undertaken in experiment 2, the analysis of data can only be presented across three (3) of the four (4) time series. A summary of overall results for three (3) of the four (4) time series is provided below in Table 5.5.2.1.1.

Table 5.5.2.1.1 Overall Results Summary – Tool 1 and Tool 2 data pooled

A Comparison of P/G and U/G Average MAPEs

Time Series	Time Horizon	Tool 1 & Tool 2 PG	Tool 1 & Tool 2 UG	p	Accuracy
B17	First 6 periods	26.84	30.25	**0.0414**	**PG > UG**
	Last 6 periods	21.16	22.20	0.1550	pg ≅ ug
	All 12 periods	23.99	26.22	**0.0180**	**PG > UG**
B26	First 6 periods	9.61	10.80	0.0654	pg > ug
	Last 6 periods	18.03	18.12	0.9276	pg ≅ ug
	All 12 periods	13.82	14.46	0.2522	pg ≅ ug
M6	First 6 periods	12.44	18.58	**0.0000**	**PG > UG**
	Last 6 periods	26.70	30.93	**0.0012**	**PG > UG**
	All 12 periods	19.57	24.76	**0.0000**	**PG > UG**

Note: Significance Values: $p < 0.05$ are emboldened.

As can be seen from the above results, the accuracy of predictions of postgraduate subjects in a judgemental forecasting task is clearly superior to the accuracy of predictions for undergraduate subjects – regardless of the *media* and related *technologies*. Postgraduate subjects undertaking a judgemental forecasting task produced more accurate results than undergraduate subjects undertaking the same experimental task. Clearly, for two of the time series presented, B17 and M6, accuracy for postgraduate subjects was superior to undergraduate subjects (**PG > UG**). Whilst for the other time series presented, B26, there was no reported difference between the accuracy of predictions made by postgraduate subjects compared to the accuracy of predictions made by undergraduate subjects (pg ≅ ug). Therefore, findings for two of the time series presented – B17 and M6 – support the notion that postgraduate subjects produce more accurate results than undergraduate subjects in a time series extrapolation task. This is also in keeping with findings from Whitecotton's (1996) study which reported a higher accuracy for postgraduate subjects and financial analysts, compared to undergraduates.

A more detailed review of the results presented in Table 5.5.2.1.1, identifies that for two of the time series – B17 and M6 – postgraduate subjects produced significantly superior predictions than undergraduate subjects across both the 1-6 period time horizon grouping, as well as the All 12 period (periods 1-12), time horizon grouping. More specifically, results reported for time series M6 showed that postgraduate subjects were significantly superior across all of the three time horizon groupings; periods 1-6; periods 7-12; as well as All 12 periods (periods 1-12). It is also interesting to note that for time series B17, the significance level for the 1-6 period, time horizon grouping was so great, that it rendered the accuracy of predictions for postgraduate subjects to be significantly superior to undergraduate subjects overall – over the All 12 periods time horizon grouping. For a full review of reported statistics, please refer to Table 5.5.2.1.1.

Given that the general results indicate that the accuracy of forecasts as predicted by postgraduate subjects were superior to those predicted by undergraduate-subjects, it would be interesting to determine whether this superiority could be further linked to any specific Tool used in the judgemental extrapolation task. In order to examine this further, additional post-hoc testing was performed.

5.5.2.2 Tool 1 Comparison – Postgraduate subjects versus Undergraduate subjects

In order to determine whether the reported differences in forecast accuracy between postgraduate subjects and undergraduate subjects as presented in Table 5.5.2.1.1 could be directly attributed to the Tool – the *media* and the related *technology* – further post-hoc testing was conducted. The analysis involved a direct comparison of data collected from both postgraduate, and undergraduate subjects, undertaking a judgemental forecasting task where both groups of subjects used Tool 1 in order register their forecast predictions (screen-based *media* and related *technologies*). More specifically, using Tool 1, both groups of subjects were required to use a combination of slider, mouse and keyboard in order to articulate their needs and register their forecast predictions.

Data collected from forty three (43) postgraduate subjects who completed the judgemental extrapolation task utilizing Tool 1 was compared to that collected from eighty one (81) undergraduate subjects who also completed the judgemental extrapolation task utilizing Tool 1. As previously noted, due to the fact time series B8 was excluded from the experimental design for experiment 2, and the fact that F3 was introduced as the fourth experimental task to be undertaken in experiment 2, the analysis of data can only be presented across three (3) of the four (4) time series. A summary of overall results for three (3) of the four (4) time series is provided below in Table 5.5.2.2.1.

Table 5.5.2.2.1 Overall Results Summary – Tool 1
A Comparison of Postgraduate and Undergraduate Average MAPEs

Time Series	Time Horizon	Tool 1 PG	Tool 1 U/G	p	Accuracy
B17	First 6 periods	21.80	23.66	0.1433	pg > ug
	Last 6 periods	26.56	32.21	**0.0318**	**PG > UG**
	All 12 periods	24.18	27.93	**0.0136**	**PG > UG**
B26	First 6 periods	9.83	11.70	**0.0422**	**PG > UG**
	Last 6 periods	15.83	18.96	**0.0297**	**PG > UG**
	All 12 periods	12.83	15.33	**0.0040**	**PG > UG**
M6	First 6 periods	13.55	21.89	**0.0000**	**PG > UG**
	Last 6 periods	23.36	32.62	**0.0000**	**PG > UG**
	All 12 periods	18.46	27.25	**0.0000**	**PG > UG**

Note: Significance Values: $p < 0.05$ are emboldened.

The results in Table 5.5.2.2.1 are overwhelming, showing very clearly that when utilizing Tool 1 (the screen-based *media* and related *technologies*) in order to articulate their needs and register their forecast predictions, postgraduate subjects undertaking a judgemental forecasting task produced significantly more accurate results than undergraduate subjects undertaking the same experimental task. Clearly, for all three of the time series presented, B17, B26 and M6, accuracy for postgraduate subjects was significantly superior to undergraduate subjects (PG > UG).

Findings from this additional post-hoc testing further support the notion that postgraduate subjects produce more accurate results than undergraduate subjects in a time series extrapolation task. This is also in keeping with the generally accepted notion that, experience with the media – in particular the screen-based media – has been found to have a positive impact on overall task performance (Lucas, 1981; Kasper and Cerveny, 1985; Kagan and Pietron, 1987). In addition, this finding also seems to indicate, that for postgraduate subjects, the use of computers can potentially improve forecasting accuracy (Dalrymple 1987).

A more detailed review of the results presented in Table 5.5.2.2.1, identifies that for two of the time series – B26 and M6 – postgraduate subjects produced significantly superior predictions than undergraduate subjects across all of the three time horizon groupings; periods 1-6; periods 7-12; as well as All 12 periods (periods 1-12). In addition, the results identify that for all of the three time series – B17, B26 and M6 – postgraduate subjects produced significantly superior predictions than undergraduate subjects across two of the three time horizon groupings; periods 7-12 and, All 12 periods (periods 1-12).

It is also interesting to note that for time series B17, the accuracy of predictions for postgraduate subjects in the period 1-6 time horizon grouping, was superior to the accuracy of the undergraduate subjects, although not significantly so. This was the only non-significant difference reported for any of the time series and across all of the three time horizon groupings – using Tool 1 and comparing postgraduate subjects to undergraduate subjects. For a full review of reported statistics, please refer to Table 5.5.2.2.1.

Given the overwhelming superiority of the accuracy of forecasts as predicted by postgraduate subjects, it appears that potentially, these results can largely be attributed to the nature of the subjects who undertook the experimental task, rather than simply the Tool itself. There were differing results for B26 when all data was pooled across both Tools, than when data for Tool 1 only was compared. Results for postgraduate subjects were equivalent to undergraduate subjects when data was pooled across both Tools (pg ≅ ug). However, when data for Tool 1 only was compared, postgraduate subjects produced significantly better results than undergraduate subjects (**PG > UG**). It would be interesting to conduct further post-hoc testing to determine the extent of the difference between postgraduate and undergraduate subjects when using Tool 2 (paper *media* and related *technologies).* It would be interesting to determine whether this superiority between postgraduate and undergraduate subjects could also be found to exist when using Tool 2 in the judgemental extrapolation task. In order to examine this further, additional analysis was undertaken.

5.5.2.3 Tool 2 Comparison – Postgraduate subjects versus Undergraduate subjects

In order to review whether the reported differences in forecast accuracy between postgraduate subjects and undergraduate subjects as presented in Table 5.5.2.1.1 can be directly attributed to the Tool – the *media* and the related *technology* – some post-hoc testing was undertaken. This involved a direct comparison of data collected from both postgraduate, and undergraduate subjects, undertaking a judgemental forecasting task, where both groups of subjects used Tool 2 in order register their forecast predictions. More specifically, using the paper-based *media* and related *technologies*, both groups of subjects were required to use a pencil and an eraser in order to articulate their needs and register their forecast predictions.

Data collected from fifty eight (58) postgraduate subjects who completed the judgemental extrapolation task utilizing Tool 2 was compared to that collected from fifty nine (59) undergraduate subjects who also completed the judgemental extrapolation task utilizing Tool 2. As previously noted, due to the fact time series B8 was excluded from the experimental design for experiment 2, and the fact that F3 was introduced as the fourth experimental task to be undertaken in experiment 2, the analysis of data can only be presented across three (3) of the four (4) time series. A summary of overall results for three (3) of the four (4) time series is provided below in Table 5.5.2.3.1.

Table 5.5.2.3.1 Overall Results Summary – Tool 2
A Comparison of Postgraduate and Undergraduate Average MAPEs

Time Series	Time Horizon	Tool 2 P/G	Tool 2 U/G	p	Accuracy
B17	First 6 periods	20.25	20.45	0.7687	pg ≅ ug
	Last 6 periods	25.30	27.92	0.6838	pg ≅ ug
	All 12 periods	22.77	24.18	0.7688	pg ≅ ug
B26	First 6 periods	9.86	9.72	0.7533	pg ≅ ug
	Last 6 periods	19.21	17.13	**0.0442**	**UG > PG**
	All 12 periods	14.53	13.42	0.1009	pg ≅ ug
M6	First 6 periods	11.96	14.64	**0.0205**	**PG > UG**
	Last 6 periods	29.77	28.92	0.8407	pg ≅ ug
	All 12 periods	20.86	21.78	0.2396	pg ≅ ug

Note: Significance Values: $p < 0.05$ are emboldened.

The results in Table 5.5.1.3.1 clearly show that when utilizing Tool 2 (the paper-based *media* and related *technologies*) in order to articulate their needs and register their forecast predictions, postgraduate subjects undertaking a judgemental forecasting task produced results that were equivalent to undergraduate subjects undertaking the same experimental task (in terms of accuracy). Clearly, for all three of the time series presented; B17, B26 and M6, accuracy for postgraduate subjects was equivalent (pg ≅ ug) across at least two of the time horizon groupings. More specifically, for all three of the time series presented, B17, B26 and M6, accuracy for postgraduate subjects was equivalent to undergraduate subjects (pg ≅ ug) for the All 12 periods (periods 1-12) time horizon grouping. In addition, for time series B17, the accuracy of postgraduate subjects was equivalent to undergraduate subjects across all of the three time horizon groupings; periods 1-6; periods 7-12; as well as All 12 periods (periods 1-12). Whilst for time series B26, the accuracy of postgraduate subjects was equivalent to undergraduate subjects for two of the three time horizon groupings; periods 1-6 and the All 12 periods (periods 1-12). There were some unusual results reported for the 7-12 period time horizon grouping for B26. Undergraduate subjects produced significantly more accurate results than postgraduate subjects. However, in the main, there was no significant difference reported across the All 12 periods (periods 1-12) time horizon grouping for B26, indicating no significant difference between subjects. Results for M6 indicate that the accuracy of postgraduate subjects was equivalent to undergraduate subjects for two of the three time horizon groupings; periods 7-12 and the All 12 periods (periods 1-12). There were some unusual results reported for M6 in the 1-6 periods time horizon grouping, with postgraduate subjects producing significantly more accurate results than undergraduate subjects. However, in the main, there was no significant difference reported across the All 12 periods (periods 1-12), time horizon grouping, indicating no significant difference between undergraduate and postgraduate subjects.

In general, findings from this additional post-hoc testing further support the notion that utilizing Tool 2 – paper-based *media* and related *technology* – postgraduate subjects do not produce more accurate results than undergraduate subjects in a time series extrapolation task. There appears to be a level of interaction or a relationship between postgraduate subjects and the Tool, especially Tool 1 the screen-supported *media* and related *technology*. Given the findings presented in the previous section, 5.5.2.2, this appears to add further strength to the belief that for postgraduate subjects, the use of computers can potentially improve forecasting accuracy (Dalrymple 1987). Perhaps this needs to be investigated further.

Given the above findings, it would be interesting to review whether the *degree* of TTF – measured in terms of forecast accuracy – could be further improved by varying some of other independent variables in the proposed Research Model for the current Thesis – in particular, the Tool.

5.6 Discussion

5.6.1 Main Findings

We used four monthly time series for the task in experiment 2. As for previous experiments, the time series were randomly selected from the M-Competition database comprising 111 real time series (Makridakis *et al.,* 1982). As discussed in section 5.3, time series B8 was eliminated from the experimental design and replaced with time series F3. In keeping with previous experiments, each group of subjects predicted the accuracy of forecasts for all four time series under each of the treatment conditions. In order to minimise any possible task fatigue effects, it was decided to only select and display a total of four time series to subjects as part of the experimental task.

The experimental results indicate, counter to what was hypothesised, that for two of the four time series presented to postgraduate subjects, the use of a screen-based medium and related technology (Tool 1) for judgemental extrapolation, gave superior accuracy over the paper-based medium and related technology (Tool 2). Further post-hoc testing revealed that in the main, postgraduate subjects exhibited superior forecast accuracy over undergraduate subjects in the screen-supported medium.

Our motivation for using postgraduate students as subjects in experiment 2 was primarily to determine whether or not the results which had been obtained from experiment 1 – using undergraduate students as subjects – could be replicated. In addition, the use of postgraduate students enabled us to test our proposed Research Model presented in section 2.4. This was achieved by altering one of the independent variables – namely "people" and holding all other factors constant, and reviewing the overall effect on "task performance" – measured in terms of forecasting accuracy.

Findings from experiment 2 highlighted the fact that the results of experiment 1 could largely be attributed to the nature of the subjects undertaking the judgemental extrapolation task. Undergraduate students were not able to perform as well when using Tool 1 – the screen based *media* and related *technologies* – in order to articulate their needs and register their forecast predictions. In addition, findings from experiment 2 add strength to the general notion that "experience" with the media – in particular the screen based media – has often been found to have a positive impact on task performance (Lucas, 1981; Kasper and Cerveny, 1985; Kagan and Pietron, 1987). As has been outlined in section 5.1, we chose to use "age/maturity" as an indicator of experience. This was primarily due to the fact that we elected not to use the traditional notion of "experience" which has generally been linked to the task itself, but chose to concentrate on the subjects' experience with the *media* and the related *technologies.* With this goal in mind, it was felt that "age/maturity" would prove to be a better indicator especially given the fact that postgraduate students participating in the experimental study were undertaking courses in Information Systems at UNSW at a Masters level and were (arguably) highly proficient in the use of computers and related technology. In addition, findings from experiment 2 also seem to indicate, that for postgraduate subjects, the use of computers can potentially improve forecasting accuracy (Dalrymple 1987).

It is interesting to note that results of post-hoc testing identify that when the task is undertaken using Tool 2 – paper-based *media* and related *technologies* – no significant difference is reported between the accuracy of undergraduate and postgraduate subjects. That is, when using the paper-based Tool, and a pencil and an eraser, in order to articulate their needs and register their forecast predictions, undergraduate and postgraduate subjects produce results that are deemed to be equivalent. As previously stated, there appears to be a level of interaction or a relationship between postgraduate subjects and the Tool – especially Tool 1 – screen-supported *media* and related *technologies* – something which needs to be investigated further.

In addition, as previously mentioned, it would also be interesting to review whether the *degree* of TTF – measured in terms of forecast accuracy – could be further improved by varying some of other independent variables in the proposed Research Model for the current Thesis – in particular, the Tool.

Comparing the results for experiment 2, in terms of Campbell's Task Complexity Characteristics (please refer to section 4.6.1), it is interesting to note that in terms of task accuracy, the results of judgemental extrapolation are similar to those produced by the reading for comprehension studies – an entirely different result to that of experiment 1. Perhaps the level of interaction with the Tool, in particular the manipulation device (*related* technologies) in *combination* with the "age/maturity" of the individual (the postgraduate subject) played a major role, or had a major impact on task performance. Additional experimentation is required in order to review this further.

5.7 Future Research

The findings in experiment 2 lead us to a number of questions which could be investigated in future research.

Given that forecasting accuracy in the screen-supported medium (Tool 1) was superior in comparison to forecasting accuracy in the paper-based medium (Tool 2), it would be interesting to undertake further research in this particular area. The question which must be asked is;

> Could forecasting accuracy in the screen-supported medium be further improved by enhancing the Tool? That is, would a more sophisticated Tool, one with a decreased level of interaction with the manipulation device, or one with a more advanced *method* of interaction, improve the accuracy of predictions for postgraduate subjects?

This is the question at the heart of Goodhue and Thompson's TTF model (1995), the notion of whether the *degree* of fit or, TTF, can be improved by manipulating the independent variables presented in their theoretical model. This is also one of the key components of this Thesis. As outlined in the Research Model presented in section 2.4, the characteristics of Tool – the *media* and the related *technology* – are an integral component of this Thesis. Having identified the Tool as being one of the independent variables of the current Thesis, and given the results of experiment 2, it would be interesting to review the results of further experimentation, using a more sophisticated screen-based Tool, and holding all other factors constant. As previously noted, this is also in keeping with Goodhue and Thompson's' (1995) theoretical model of TTF – a crucial component of this Thesis.

Perhaps future research should investigate whether the use of a more sophisticated screen-based Tool would have produced a more accurate result. A question which needs to be posed is:

> Would a more advanced Tool – screen-based *media* and the related *technologies* – have an impact on the accuracy of predictions made by postgraduate subjects in the judgemental extrapolation task?

5.8 Conclusion

Experiment 2 has explored differences in the accuracy of predictions in a judgemental extrapolation task when that task is undertaken by postgraduate subjects, and is presented across two different presentation and working media; namely paper-based and screen-supported. In addition, post-hoc testing was conducted to try to determine the exact nature or extent of any reported differences. In addition, experiment 2 sought to determine whether the accuracy of predictions in a judgemental extrapolation task differed between Tools, across the time horizon groupings (periods 1-6 and periods 7-12), for the presented series.

The main finding of experiment 2 is that when undertaken by postgraduate subjects, judgemental extrapolation is a task which is often more accurately undertaken in the screen-supported medium rather than the paper-based medium. This is a very different finding to that reported for experiment 1, where the opposite was found to occur. We may speculate that these findings indicate that the *degree* of TTF is somewhat improved when more individuals with a greater "age/maturity" level utilize a Tool with the following characteristics – screen-based *media* and related *technologies*. These findings are, of course, qualified by the simplified version of the forecasting task used in experiment 2.

However, given the above, it would be interesting to review whether the *degree* of TTF – measured in terms of forecast accuracy – could be further improved by varying some of other independent variables in the proposed Research Model for the current Thesis. In particular it would be interesting to undertake further experimental studies in order to seek to provide answers to the following question:

> Would a more sophisticated Tool – perhaps one with a decreased level of interaction with the manipulation device – *improve* the degree of TTF, and in turn the accuracy of forecast predictions in a judgemental extrapolation task?

This is the basis for undertaking experiment 3 which will be discussed in section 6.

6. Experiment Three

6. Experiment 3

6.1 Introduction

The following section will describe the third experimental study which seeks to further review the influence of *media* and the related *technology* on the accuracy of predictions in a judgemental extrapolation task by introducing a more sophisticated screen-supported Tool.

Prior experiments undertaken as part of this Thesis (experiments 1 and 2) both used two different treatment conditions in order to review the influence of *media* and the related *technology* on task accuracy. Both series of experiments compared the accuracy of forecasts using a screen-based Tool (Tool 1) to those produced using a paper-based Tool (Tool 2). The focus of these two prior studies was the differences *between* the media. In experiment 3, we will concentrate on the differences *within* the *media*. The primary focus will be the difference *between* the related *technologies* – the use of slider, mouse and keyboard (Tool 1); and mouse only (Tool 3). Separate groups of subjects will be required to use either Tool 1, or Tool 3, in order to articulate their needs and register their forecast predictions.

In experiment 3, the *technology*, in terms of user-interaction with the manipulation device will be improved. The screen-based tool used for experiment 3 was the same as Tool 1 (used in both experiments 1 and 2) in terms of the display format. However, the major difference between the Tools was the fact that task-doers utilizing Tool 3 were required to use only the mouse in order to articulate their needs and register their forecast predictions. Therefore subjects undertaking experiment 3, and utilizing Tool 3, can be said to have a decreased level of interaction with the manipulation or articulation device. Given that the level of interaction with Tool 3 is limited to subjects having to use the mouse to predict period values, it can be said that the use of Tool 3 presents a decrease in both the *level* and *depth* of interaction with the manipulation mechanism, and so can be seen to be decreasing the amount of "interruptions" to the user's cognitive processes whilst undertaking the task at hand.

Looking to the literature in relation to the issue of interaction with the manipulation device, in terms of "interrupting" the cognitive processes of the task-doer, we can see that the work of Speier, Valacich and Vessey (1997) is of interest. Their work was first referenced in section 2.6. The aim of their study was to conduct further analysis into the effects of task interruptions and information presentation on task performance. Speier *et al.,* 1997 used Corragio's 1990 definition of interruption – an "externally-generated …event that breaks the continuity of cognitive focus on a primary task" (p. 19), and further advised that interruptions can create both capacity and structural interference (as reported by Kahneman, 1973). In addition, citing the work of Baecker *et al.,* 1995, Speier *et al.* mention that in today's general work environment, computer-based tasks often involve "high cognitive loads", and that in turn, this might make the task more prone to interference from interruptions (p. 21).

This issue is a motivating factor for conducting experiment 3. Looking to the results of previous experimental studies, in particular experiment 2, which appeared to indicate that perhaps there, was an interaction between the Tool and the subjects – especially when postgraduate students utilized screen-supported *media* and related *technologies* – it was considered logical for this study to proceed onto another of the "variables" in the proposed Research Model (section 2.4). In particular, the motivation behind experiment 3 was to review the potential impact of modifying the actual Tool – or more specifically the articulation mechanism used by subjects in order to register their forecast predictions. Instead of having to use a combination of slider, mouse and keyboard in order to articulate their needs and register their forecast predictions, subjects utilizing Tool 3 were required to use only the mouse. As previously noted, the use of Tool 3 can be said to present a "decrease" in both the *level* and *depth* of interaction with the manipulation mechanism, and so can be seen to be decreasing the amount of "interruptions" to the user's cognitive processes whilst undertaking the task at hand.

This line of thought also complements the work undertaken by Jacko and Ward (1996). As previously mentioned in section 2.5.2, the primary aim of their 1996 study, was to validate a short-term memory link between Campbell's (1988) task complexity framework and the information processing model for computer-based psychomotor tasks as developed by Salvendy and Knight (1982). Their validated information processing model sought to provide some insight into task performance, given a certain level of task complexity, but their object of interest was psychomotor tasks. Essentially, psychomotor tasks are those which require the person to use controlled movements involving their body, or some instrument or device as an extension of their body – such as a mouse – in order to accomplish the task. Their findings identified that an increase in information load could challenge task-doer's short-term memory due to the additional burden requiring them to attend to more and more things. Jacko and Ward propose that this increased information load could be directly translated to represent an increase in task complexity (p. 535).

In turn, this line of thought also complements Goodhue and Thompson's (1995) model of TTF. Whilst providing a review of the background literature for their proposed model of Task-Technology Fit, Goodhue and Thompson acknowledge that much of the previous research in the area of "fit" component of TTF has centered around the graphs versus tables literature. Citing the work of Vessey 1991 (and others p. 214 – many of whom have been referenced in sections 2.3 and 2.6) and her theory of "cognitive fit" they acknowledge her valuable contribution to the research area. The essence of Vessey's theory being – that mismatches between technology characteristics and tasks have an adverse effect on the decision-making process, and in turn, affect performance. In his 1995 study, Goodhue also argued that Vessey's theory of "cognitive fit" could be seen as a more general case of the TTF perspective (p. 1831). He maintains that whilst research in the area of "cognitive fit" "... tends to focus narrowly on cognitive task processes, TTF applies to a more macro task domain including not only cognitive but also other task processes such as the mechanical process of accessing data ..." (p. 1830). This is also the focus of experiment 3 – to review the potential impact on TTF by altering the "mechanical process of accessing data" i.e. the interaction with the manipulation device. The extent of this impact will be measured in terms of task performance, or more specifically, in terms of forecast accuracy.

Therefore, given the previous research as cited above, and the general results from earlier experimental studies undertaken as part of this Thesis, it was decided to conduct a further experiment as part of the current study. Of particular interest, were the results of experiment 2 which indicated that postgraduate subjects produced more accurate results when using the screen-supported media. In an effort to investigate why this was found to occur, it was decided to conduct further experimentation, this time *within* the *media*, and across different *technologies*. That is, by varying the *technology* but keeping the *media* constant. Experiment 3 employed the use of a more sophisticated screen-based Tool, one with a decreased level of interaction with the manipulation device – mouse only. Or, to put it another way, using a more advanced screen-based Tool, with a more advanced *method* of interaction with the manipulation device – mouse only. Our motivation in varying the Tool was to determine whether the different *level* and *extent* of the interaction with the manipulation device would have any impact upon task performance – measured in terms of the accuracy of forecast predictions for postgraduate subjects.

6.2 Research Design for Experiment 3

Experiment 3 used a 2 x (2) research design and sought to perform an analysis over time. The first factor (a between subjects factor) being the Tool – *media* and the related *technology* – and the second, the time horizon (a within subjects factor). Postgraduate subjects undertook the task under each treatment condition. As for experiments 1 and 2, the Analysis of Variance (ANOVA) statistical procedure was employed to undertake data analysis.

Using postgraduate subjects, experiment 3 sought to compare the accuracy of judgemental forecasting predictions produced using black and white, screen-supported time series representations presented on Tool 1, with the accuracy of those produced using colour, screen-supported time series displays presented on Tool 3. In addition, as for previous experiments undertaken as part of the current study, experiment 3 sought to determine the extent, if any, of differences between the accuracy of predictions by time horizon *across* the two different treatment conditions. As previously reported in section 3.1.3, we are interested not in the influence of time horizon *per se*: we are only interested in horizon as it interacts with the accuracy of various tools – the *media* and the related *technologies*. As has been widely reported, and generally accepted, there is a tendency for forecasting accuracy to decrease as the forecast horizon increases (Dalrymple and King, 1981; Armstrong, 1985). It would be interesting to see if this tendency would be found to occur in experiment 3 (as it was in experiments 1 and 2) and to identify the treatment condition/s under which this was found to be more prominent.

Therefore, in summary, the focus of experiment 3 is the Tool – specifically, the level of engagement or interaction with the manipulation device.

In terms of the interaction with the *technology* or the manipulation device, task-doers utilizing Tool 1 were required to use the slider, mouse and keyboard in order to articulate their needs and register their forecast predictions (this was no different to experiments 1 and 2). However, task-doers utilizing Tool 3, were required to use a new more advanced Tool, using only the mouse, in order to articulate their needs and register their forecast predictions.

In terms of task complexity, this was held constant for experiment 3. Subjects were presented with an unstructured judgemental extrapolation task, no decomposition was provided and as such the task was deemed to be a "more" complex task.

In terms of people, or individual characteristics as referred to by Goodhue and Thompson (1995), this was also held constant for experiment 3 – only postgraduate subjects undertook the experiment.

In summary, the characteristics of the task and the various tools employed – *media* (treatment condition) and the related *technology* (manipulation device), for experiment 3 are presented in Table 6.2.1 below:

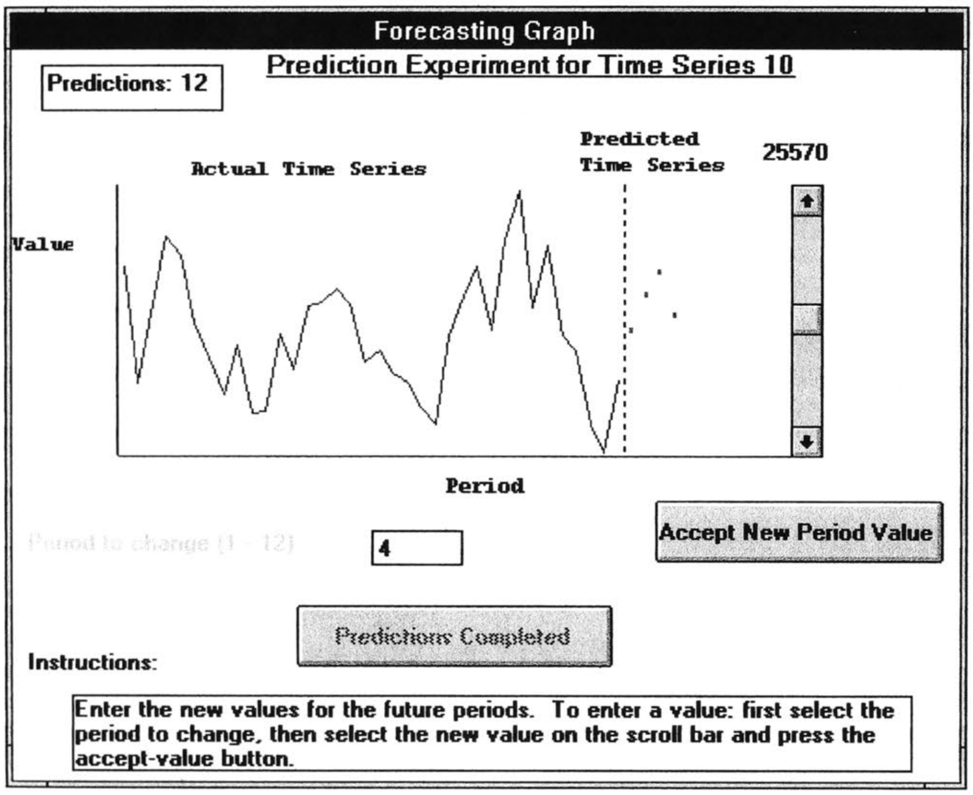
Table 6.2.1 Experimental Tool/s and Task Experiment 3

Tool_ID	Treatment Condition	Manipulation Device	Task Complexity
Tool 1	Screen-based	combination of slider, mouse and keyboard	Unstructured more complex
Tool 3	Screen-based	mouse	Unstructured more complex

In order to facilitate readability of the following material, a graphic presentation of the Tools is provided below in sections 6.2.1.1 and 6.2.1.2 respectively.

Figure 6.2.1.1 Time Series Presentation[1] for Tool 1
Computer-supported Treatment

Tool 1: ***Media*** – screen-based; ***Technology*** – slider, mouse and keyboard
Task Complexity – unstructured, more complex



[1] Picture is smaller than actual size presented to task-doers.

As previously advised in section 3.4.3, in order to record predictions, task-doers were required to use the keyboard to enter the period number and then the mouse to select the predicted value for that period using the slider (next to the time series).

To complete the marking of the value on the graph, task-doers clicked on the "Accept New Period Value" button. This process was required for each of the twelve (12) period values to be predicted and for any subsequent prediction adjustments. After all twelve (12) periods had been predicted and accepted, the "Predictions Completed" button was enabled, once this button was clicked, task-doers moved onto the next time series.

Once all period values for the final time series were entered and accepted, task-doers were thanked for their participation in the experiment. The software was disabled and the task completed.

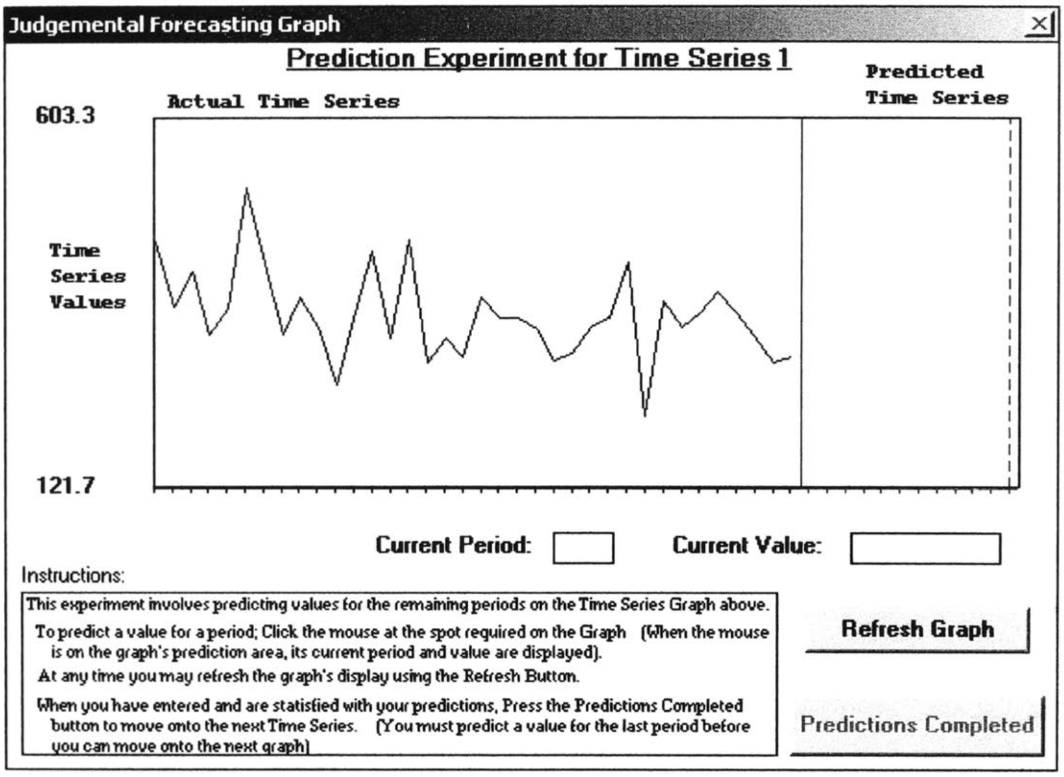
As can be seen from the above, the use of Tool 1 commanded a fairly high degree of direct interaction with the manipulation device – or, to use Speier *et al.'s* terminology, subjected the task-doer to a fairly high level or degree of "interruptions".

Figure 6.2.1.2 Time Series Presentation[2] for Tool 3

Computer-supported Treatment

Tool 3: ***Media*** – screen-based; ***Technology*** – mouse

Task Complexity – unstructured, more complex



[2] Picture is smaller than actual size presented to task-doers.

Given the results of experiment 2 which indicated that postgraduate subjects produced more accurate results when utilizing Tool 1 – the screen-based *media* and related *technology* – and our proposition that a more advanced Tool – in terms of the *level* and *depth* of interaction with the manipulation device – will facilitate task performance, it is hypothesised that the accuracy of predictions will be superior when the task is undertaken using a more advanced screen-based Tool. Hypothesis one states:

H1: There will be a superior difference in the accuracy of predictions in a judgemental forecasting task when the task is undertaken using a more advanced Tool – screen-based *media* and related *technology.*

6.3 Subjects and Experimental Procedure

Experiment 3 employed the use of forty three (43) postgraduate students undertaking Degree Programmes at the University of New South Wales. Twenty two (22) subjects completed the task using Tool 1 and the remaining twenty one (21) subjects completed the task using Tool 3. The media for both tools was screen-based *media* and related *technologies*; the difference between the Tools was in the *technology*. Task-doers utilizing Tool 1, were required to use a combination of the slider, mouse and keyboard in order to articulate their needs and register their forecast predictions. Whilst task-doers completing the task using Tool 3, were required to use only the mouse in order to articulate their needs and register their forecast predictions.

To facilitate evaluation of the effects of the Tool – the *media* and the related *technologies* – and in particular, the *technology*, groups of postgraduate subjects were randomly divided into one of the two treatment conditions as outlined in Table 6.2.1 above.

As for experiments 1 and 2, the task was the judgemental extrapolation of time series data. There was no decomposition of time series representations for experiment 3. Subjects participating in the experiment were presented with the "unstructured task" – which was also deemed to be a "more complex" task (please refer to section 2.3.4 for a more detailed discussion).

Both groups of students undertook the task in the screen-based medium, with the primary difference between the two treatment conditions being the manipulation device i.e. the related *technology*. Or more precisely, the *level* or *degree* of interaction between task-doers and the manipulation device.

Postgraduate subjects were provided with the same four randomly presented time series (B17, B26, M6, and F3) and thirty six (36) periods of historical data for each of the time series. Subjects were asked to forecast the next twelve (12) periods ahead for each time series and no feedback was provided.

As for previous experiments all postgraduate subjects were undertaking courses in Information Systems at UNSW and were (arguably) highly proficient in the use of computers and related technology.

6.4 Analysis Methodology

As for previous experiments, the metric used to assess forecast accuracy and to compare treatments was the Mean Absolute Percentage Error (MAPE); with the Analysis of Variance (ANOVA) statistical procedure employed to compare the means of the various MAPE samples derived in experiment 3. To facilitate the analysis of data in terms of time horizon, MAPE values were derived for "groups" of forecast periods; periods 1-6; periods 7-12; and All 12 periods (periods 1-12).

6.5 Results

6.5.1 Overall Results

The data collection process recorded predictions made by forty three (43) postgraduate subjects under two separate screen-based treatment conditions.
The first group of twenty two (22) subjects completed the task utilizing Tool 1, with a combination of slider, mouse and keyboard as the manipulation device. The second group of twenty one (21) subjects completed the task utilizing Tool 3, with the manipulation device being a mouse. In accordance with the hypothesis

presented, data was collected for each treatment condition and across both time-horizons. A summary of overall results for all four (4) time series and across both treatment conditions is provided below in Table 6.5.1.1.

Table 6.5.1.1 Overall Results Summary – across Screen-based Media
A Comparison of Tool 3 and Tool 1 Average MAPEs
For Postgraduate Subjects via Tool / Manipulation Device

Time Series	**Time Horizon**	**Tool 1 slider, mouse & keyboard**	**Tool 3 Mouse**	**p**	**Accuracy**
B17	First 6 periods	22.23	31.17	**0.0007**	**T1 > T3**
	Last 6 periods	29.30	19.20	**0.0006**	**T3 > T1**
	All 12 periods	25.76	25.19	0.7532	T1 ≅ T3
B26	First 6 periods	9.99	11.40	0.2779	t1 > t3
	Last 6 periods	15.61	18.52	0.2491	t1 > t3
	All 12 periods	12.80	14.96	0.1264	T1 ≅ T3
M6	First 6 periods	14.38	26.39	**0.0000**	**T1 > T3**
	Last 6 periods	24.58	15.70	**0.0006**	**T3 > T1**
	All 12 periods	19.48	21.05	0.3613	T1 ≅ T3
F3	First 6 periods	2.78	3.27	0.3529	t1 > t3
	Last 6 periods	4.61	3.38	0.0910	t3 > t1
	All 12 periods	3.69	3.32	0.5139	T1 ≅ T3

Note: Significance Values: $p < 0.05$ are emboldened.

LEGEND: T1 = Tool 1 (slider, mouse and keyboard) T3 = Tool 3 (mouse);
T1 > T3 indicates Tool 1 is significantly better than Tool 3
T1 ≅ T3 indicates paper and screen overall equivalent
t1 > t3 indicates tool 1 is better than tool 3 but not significantly so

6.5.2 H1: Difference in Accuracy between Tools, better accuracy with Tool 3.

Hypothesis 1 for experiment 3 predicted that there would be a difference in the accuracy of predictions in a judgemental forecasting task between Tool 1 and Tool 3 – both screen-based *media* but with different related *technologies*. Hypothesis 1 further predicted that accuracy of forecast predictions in a judgemental time series extrapolation task would be superior for postgraduate subjects utilizing Tool 3.

As the mixed results in Table 6.5.1.1 show, H1 is not supported across the All 12 period time horizon grouping. The overall results indicate that there is no significant difference in the accuracy of forecasts between postgraduate subjects utilizing Tool 1 and those utilizing Tool 3 across the All 12 period time horizon grouping. However, on closer examination, there is a reported difference between the Tools. This difference indicates superiority of Tool 1 over Tool 3 in the 1-6 period time horizon grouping, and the exact opposite occurring for the 7-12 period time horizon grouping. The overall results are therefore not conclusive for three of the four time series. There is no clear indication provided of the effect of the Tool, in particular, the interaction with the manipulation device on task accuracy, or more specifically, the *degree* of TTF. A more detailed examination of results, in terms of forecast horizon grouping is therefore required.

Whilst attempting to interpret results across the 1-6 time period horizon grouping, it is clear that for all four time series presented (B17, B26, M6 and F3); a definite pattern in relation to forecast accuracy has been established. The accuracy for subjects using Tool 1 – the screen-based *media* and related mouse, slider and keyboard *technology* – was superior in the 1-6 period time horizon grouping for all four time series. In fact, for two of the four time series – B17 and M6 – this reported difference was found to be significant. There was no reported significant difference between postgraduate subjects using Tool 1, and those using Tool 3 for the remaining two time series – B26 and F3. However, subjects utilizing Tool 1 produced marginally more accurate forecasts that those using Tool 3 in the 1-6 period time horizon grouping. These general results tend to indicate that for *up-close* forecast predictions, subjects undertaking a judgmental time series extrapolation task were not adversely affected by the *level* or *degree* of interaction with the manipulation device, or, to use Speier *et al.'s* (1997) terminology, by the extent of "interruptions".

A further review of the results across the 7-12 period time horizon grouping, as presented in Table 6.5.1.1, indicates that for three of the four time series presented (B17, M6 and F3), a definite pattern in relation to forecast accuracy has also been established. The accuracy for subjects using Tool 3 – the screen-based *media* and related mouse *technology* – was superior in the 7-12 period time horizon grouping for all three series. For two of the four time series (B17 and M6) this reported difference was found to be significant. This result is in keeping with that reported for the 1-6 period time horizon grouping for these two time series (B17 and M6). There was no reported significant difference between postgraduate subjects using Tool 1, and those using Tool 3 for the remaining two time series (B26 and F3). However, subjects utilizing Tool 1 produced marginally more accurate forecasts than those using Tool 3, in the 1-6 period time horizon grouping for time series F3. These general results tend to indicate that for forecast predictions further along the time horizon, subjects undertaking a judgmental time series extrapolation task were adversely affected by the *level* or *degree* of interaction with the manipulation device, or, to use Speier *et al.'s* (1997) terminology, by the extent of "interruptions".

These findings for the 7-12 period time horizon grouping are also in keeping with the notion that there is a tendency for forecasting accuracy to decrease as the forecast horizon increases. This is widely reported in the forecasting literature (Dalrymple and King, 1981; Armstrong, 1985) and has been shown to occur in judgemental forecasting (Lawrence *et al.*, 1985; O'Connor, Remus and Griggs, 1993). Perhaps this is an indication that essentially what is happening is as a result of the interaction between the Tool and the Task itself. As previously reported, the primary aim of experiment 3 was to review whether the *degree* of TTF – measured in terms of forecast accuracy – could be further improved by varying the Tool – one of the independent variables in the proposed Research Model for this Thesis presented in section 2.4.

For a full review of reported statistics, please refer to Table 6.5.1.1.

6.6. Discussion

6.6.1 Main Findings

We used four monthly time series for the task in experiment 3 – the same time series that were used in experiment 2. As for previous experiments, the time series were randomly selected from the M-Competition database comprising 111 real time series (Makridakis *et al.,* 1982). In keeping with previous experiments, each group of subjects predicted the accuracy of forecasts for all four time series under each of the treatment conditions. In order to minimise any possible task fatigue effects, it was decided to only select and display a total of four time series to subjects as part of the experimental task.

The experimental results indicate, counter to what was hypothesised, that superior accuracy of Tool 3 – screen-based *media* and related mouse *technology* – was only reported for three of the four time series presented to postgraduate subjects. In turn, this superior accuracy was only reported across the 7-12 period time horizon grouping for those three time series. Experimental results further indicated that there was no reported superior accuracy of Tool 3 over Tool 1 in the 1-6 period, time horizon grouping. In fact, the results indicated superior accuracy of Tool 1 in this time horizon grouping for all four time series presented to postgraduate subjects. For two of the time series presented to postgraduate subjects (B17 and M6), this superiority was deemed to be significant. Whilst there was no reported significant difference between the accuracy of predictions reported for the remaining two time series (B26 and F3).

Our motivation for employing the use of a more sophisticated screen-based Tool, one with a decreased level of interaction with the manipulation device – mouse only – was to examine the impact on the *degree* of TTF, in terms of task performance, or forecast accuracy. To put it another way, using a more advanced screen-based Tool, one with a *simpler* method of interaction with the manipulation device, we sought to determine whether the different *level* and *extent* of the interaction with the manipulation device would have any impact upon the accuracy of forecast predictions for postgraduate subjects.

In addition, our motivation was to further examine the interaction between the variables as presented in our proposed Research Model (section 2.4). Having used postgraduate subjects in experiment 2, as opposed to undergraduate subjects in experiment 1 – essentially altering the "People" variable – it was our desire to further review the impact on task performance by varying the "Tool" variable in experiment 3. It was hypothesised that task accuracy for postgraduate subjects undertaking the task using Tool 3 – using the mouse – in order to articulate their needs and register their forecast predictions, would be superior to those undertaking the task using Tool 1 – using a combination of slider mouse and keyboard in order to complete the task.

Findings from experiment 2 highlighted the fact that when using screen-supported *media* and related *technologies*, there appeared to be a level of interaction or a relationship between postgraduate subjects and the Tool which in turn had an effect on task performance. This was deemed to be an area of further interest for the current Thesis. Results from experiment 3 seem to support the notion that a relationship does exist between the Tool and task performance, or, to put it another way, that the Tool has an impact upon the *degree* or *level* of TTF.

It is interesting to review the results for experiment 3, in terms of Speier *et al.'s* 1997 study. Their study sought to support the notion that "cognitive fit facilitates decision making because the specific process used to act on the problem representation is the same as that needed to solve the problem" (p. 23). Findings from their study further identified that decision-makers performing tasks of a complex nature have little or no excess cognitive capacity. In turn, interruptions tended to lead to a deterioration in task performance. This deterioration was found to be more marked for tasks which were deemed to be more "complex" tasks. If we look to the findings of Speier *et al.* (1997), we can see that the judgemental extrapolation task in experiment 3, identified as a complex task, tends to *increase* in terms of complexity as subjects try to forecast the values for periods further along the time horizon grouping. That is, the task of forecasting period values for close-up data points in the time series – periods 1-6 – can be considered to be *less* complex than the task of forecasting period values for the *further-out* data points – periods 7-12. Applying findings from Speier *et al.'s* study which found that interruptions inhibited performance on more complex tasks, we can propose the following. Perhaps the use of Tool 1 – using the slider, mouse and keyboard related *technology* – with its increased *level* of interruptions, did not have such an adverse effect on the accuracy of forecasts for the close-up periods (1-6).

However, as subjects were required to predict values for further-out data points (periods 7-12), the task became more complex and in turn, the *level* of interruptions had an adverse effect on the task accuracy. We could further propose that the use of the more sophisticated tool, Tool 3 – using the mouse *technology* – with its decreased *level* of interruptions, aided the task-doer in predicting period values for further-out data points (periods 7-12), because there was no excess load on their cognitive capacity. The decrease in the *degree* or *level* of interaction with the manipulation device providing a good "cognitive fit" for the task at hand and in turn, facilitating task performance.

Goodhue and Thompson's (1995) model of TTF refers to TTF as "the degree to which a technology assists an individual in performing his or her portfolio of tasks". More specifically, TTF is the correspondence between task requirements, individual abilities, and the functionality of the technology." (p. 218). It is interesting to note that the results of experiment 3 indicate that using Tool 1, TTF tends to decrease as the task becomes more complex – predicting period values for further-out data points, 7-12 period time horizon grouping. Whilst the opposite is seen to occur when using Tool 3; TTF tends to increase as the task becomes more complex. Goodhue and Thompson (1995) propose that "as the gap between the requirements of a task and the functionalities of a technology widens, TTF is reduced" (p.218). They further maintain that it is the interaction between the task, the technology and the individual which is crucial to the framework of TTF.

Addressing the results of experiment 3 from the TTF perspective, perhaps the *level* of interaction with the Tool, or the manipulation device in *combination* with the "complexity" of the task (unstructured, more complex) played a major role, or had a major impact on task performance. Additional experimentation is required in order to review this further.

6.7 Future Research

The findings in experiment 3 lead us to a number of questions which could be investigated in future research.

Given that experiment 2 concentrated on the "People" variable as presented in the proposed Research Model (section 2.4), and experiment 3 concentrated on the effects of Tool upon TTF, it would be interesting to conduct further research on the last remaining variable, namely that of Task Complexity.

It is interesting to note that forecasting accuracy using Tool 3 was superior in comparison to forecasting accuracy using Tool 1 for the 7-12 period time horizon grouping (for three of the four time series). These general results tend to indicate that the use of the more sophisticated tool – Tool 3 – with its minimal interaction required with the manipulation device – the mouse – led to a greater degree of overall TTF. It would be interesting to undertake further research in this particular area. The question which must be asked is whether forecasting accuracy in the screen-supported medium could be further improved by decreasing task complexity? That is, would the addition of structure, or essentially the decomposition of the judgemental time series extrapolation task, coupled with the use of a more sophisticated Tool, one with a decreased level of interaction with the manipulation device, improve the accuracy of predictions for postgraduate subjects? As previously mentioned, this is the question at the heart of Goodhue and Thompson's TTF model (1995), the notion of whether the *degree* of fit or, TTF, can be improved by manipulating the independent variables presented in their theoretical model. This is also one of the key components of this Thesis.

The characteristics of the Task are an integral component of this Thesis – in particular Task Complexity (as outlined in the Research Model at section 2.4).

Having identified Task Complexity as being one of the independent variables of the current study, and given the results of experiment 3, it would be interesting to review the results of further experimentation, using the addition of structure in order to decompose the task, effectively decreasing task complexity, and holding all other factors constant. As previously noted, this is also in keeping with Goodhue and Thompson's (1995) theoretical model of TTF – a crucial component of this Thesis.

Perhaps future research should investigate whether the use of decomposition to add structure to the task and in turn, decrease overall task complexity, would have produced a more accurate result. A question which needs to be posed is:

> Would the use of a more advanced Tool – screen-based *media* and mouse *technology* – in addition to the decomposition of the task, or essentially the addition of structure, have an impact on the accuracy of predictions made by postgraduate subjects in the judgemental extrapolation task?

6.8 Conclusion

Experiment 3 has explored differences in the accuracy of predictions in a judgemental extrapolation task when that task is undertaken by postgraduate subjects using two different types of screen-based tools. We were interested in determining the impact upon task performance for subjects using a more sophisticated screen-based Tool, one with a decreased level of interaction with the manipulation device – mouse only. Our motivation in varying the Tool was to determine whether the different *level* and *extent* of the interaction with the manipulation device would have any impact upon task performance – measured in terms of the accuracy of forecast predictions for postgraduate subjects.

In addition, experiment 3 sought to determine whether the accuracy of predictions in a judgemental extrapolation task differed between Tools, across the time horizon groupings (periods 1-6 and periods 7-12), for the presented series.

There are essentially two main findings from experiment 3. The first being, that general results indicate that for *up-close* forecast predictions (i.e. periods 1-6), subjects undertaking a judgmental time series extrapolation task were not adversely affected by the *level* or *degree* of interaction with the manipulation device, or, to use Speier *et al.'s* (1997) terminology, by the extent of "interruptions". The second being an indication that for forecast predictions further along the time horizon, subjects undertaking a judgmental time series extrapolation task were adversely affected by the *level* or *degree* of interaction with the manipulation device, or, to use Speier *et al.'s* (1997) terminology, by the extent of "interruptions". To put it another way, the more advanced Tool – in terms of the *level* and *depth* of interaction with the manipulation device – facilitated task performance for the *further-out* data points, those within the 7-12 period time horizon grouping.

We may speculate that these findings indicate that the *degree* of TTF is somewhat improved when the *level* or *degree* of interruptions – in terms of direct interaction with the manipulation device – is reduced as task complexity increases. These findings are, of course, qualified by the simplified version of the forecasting task used in experiment 3.

However, given the above, it would be interesting to review whether the *degree* of TTF – measured in terms of forecast accuracy – could be further improved by varying some of the other independent variables in the proposed Research Model for the current Thesis. In particular it would be interesting to undertake further experimental studies in order to seek to provide answers to the following questions:

> Could forecasting accuracy in the screen-supported medium be further improved by decreasing task complexity?
>
> Would the decomposition of the judgemental time series extrapolation task – achieved by the addition of structure, coupled with the use of a more sophisticated Tool, one with a decreased *level* of interaction with the manipulation device, improve the *degree* of TTF and in turn, task performance?

This is the basis for undertaking experiment 4 which will be discussed in section 7.

7. Experiment Four

7. Experiment 4

7.1 Introduction

The following section will describe the fourth experimental study which seeks to review the influence of task complexity on *media* and the related *technology* in relation to the accuracy of predictions in a judgemental extrapolation task. This will be achieved by decomposing the task – adding structure – and thus varying the Task Complexity.

Experiment 4 will explore the interaction of the last remaining variable in the proposed Research Model (section 2.4), namely that of Task Complexity. We are interested in providing an answer to the following questions:

> Could forecasting accuracy in the screen-supported medium be further improved by decreasing task complexity? That is, would the addition of structure or essentially the decomposition of the judgemental time series extrapolation task, coupled with the use of a more sophisticated Tool – one with a decreased level of interaction with the manipulation device – improve the accuracy of predictions for postgraduate subjects?

This is also the question at the heart of Goodhue and Thompson's TTF model (1995) – whether the *degree* of fit, or TTF, can be improved by manipulating the independent variables presented in their theoretical model. This is also one of the key components of this Thesis.

Of particular interest to this Thesis, and as outlined in the Research Model presented in section 2.4, are the characteristics of the Task – in particular Task Complexity. Having identified Task Complexity as being one of the independent variables of this Thesis and given the results of both experiments 2 and 3, we believe it is essential to conduct further experimentation in this area. Our aim is further the area of TTF by identifying the optimum task/technology alignments required in order to maximise TTF, and in turn, maximise task performance.

The results of experiment 2 indicated that postgraduate subjects produced more accurate results when using the screen-supported medium. Experiment 3 then tried to further identify whether the *level* and *depth* of user interaction with the manipulation device in the screen-supported medium also had an effect on task performance. Subjects undertaking experiment 3 utilized Tool 3 – screen-based *media* and mouse *technology*. Subjects were said to have a decreased level of interaction with the manipulation or articulation device. Given that the level of interaction with Tool 3 was limited to subjects having to use only the mouse to predict period values, it was proposed that the use of Tool 3 presented a decrease in both the *level* and *depth* of interaction with the manipulation mechanism. However, the results from experiment 3 were not conclusive, indicating that perhaps other factors may have influenced task performance. Perhaps it may even have been an interaction between the complexity of the task and the *level* and *depth* of user interaction with the manipulation device.

In order to review this further, experiment 4 will concentrate on the potential impact of Task Complexity on overall performance. Holding all other factors constant, we will add structure in order to decompose the task, thus effectively decreasing overall task complexity.

Prior to reviewing the elements of a task that have an effect on its *complexity*, it is important to firstly make a distinction between task complexity and task difficulty. It is often noted that the two terms are used interchangeably, when in fact they actually refer to two different things (Jacko and Ward, 1996). Some researchers have defined the difficulty of a task in terms of the amount of effort required to complete the task (Shaw, 1976). Some acknowledge that a task may be deemed to be difficult, without necessarily being complex, because of the sheer amount of effort required by the person in order to undertake and complete the task (Jacko and Ward, 1996). Whilst others propose that a task may be perceived to be difficult or complex, because the task actually requires the person to possess a high level of skill or knowledge in order to undertake and complete the task (Campbell, 1988; Locke, Shaw and Saari and Latham, 1981). Wood (1986) notes the failure of researchers to make the distinction between attributes of the task and attributes of the individual performing the task. Campbell (1988) proposes that task difficulty and complexity are in fact two different things which can be represented and measured objectively. He proposes the notion that task *difficulty* is an attribute that represents a *person x task interaction*, whereas task *complexity* can be viewed as an *independent measure*. Campbell concludes that within the literature, the notion of task complexity is not well understood. In trying to provide a definition of task complexity, Campbell looks to specific task attributes or qualities, and maintains that:

> "Task complexity can be directly related to the task attributes that increase information load, diversity or rate of change. Thus complexity, can be defined objectively and it can be determined independently of any particular task-doer" (p. 43).

Campbell introduces a "typology of complex tasks", used to categorise a task by identifying the fundamental task characteristics which contribute to its complexity (for a more detailed discussion please refer to section 2.5.3). Vessey (1994) also acknowledges that there is no generally accepted comprehensive theory of tasks and references the works of Campbell (1988), Fleishman (1982) and Wood (1986), praising them for their contribution to the area. Of specific interest to Vessey, is the identification of the *abstract* characteristics of a task which are influenced by different types of representation, or *modes*, of that task (eg; graphical versus tabular). This was also referred to by Campbell (1988) when he referenced the work of Hammond (1986):

> "Hammond pointed out how the complexity of essentially similar tasks varies as a function of the task's *mode* of representation" (p. 41, emphasis added).

Campbell's proposition, that task complexity may be affected by not only the *mode* of representation, but also by the *physical form*, is in essence, one of the components of the primary focus of this Thesis – the *media* – the second being, the related *technology*.

Having conducted three previous experiments, the first two which primarily addressed the media, or *physical form*, as described by Campbell, and the third concentrating on the related *technology*, this fourth experiment will concentrate on Task Complexity. Of particular interest to experiment 4 is task complexity in relation the experimental task – the judgemental extrapolation of time series data. As previously noted in section 2.3, the complexity of the judgemental extrapolation task can be varied by the addition of structure, essentially, by decomposing the task into several unique components or sub-patterns.
Of interest to the current experimental study, are the trend and seasonality sub-patterns which may be present in the time series selected for the experimental task. In summary, Experiment 4 will seek to decompose time series representations and review the impact of structure on the accuracy of predictions in a judgemental forecasting task – in essence, the relationship between task complexity and task performance.

7.2 Research Design for Experiment 4

Experiment 4 used a 2 x (2) research design and sought to perform an analysis over time. The first factor (a between subjects factor) being Task Complexity – structured versus unstructured, and the second, the time horizon (a within subjects factor). Postgraduate subjects undertook the task under each treatment condition. As for previous experiments, the Analysis of Variance (ANOVA) statistical procedure was employed to undertake data analysis.

Using postgraduate subjects, experiment 4 sought to compare the accuracy of judgemental forecast predictions produced using structured or decomposed time series representations, with the accuracy of those produced using unstructured time series representations. Both groups of subjects used the same *media* and related *technology* – Tool 3 – screen-based media and mouse. As for previous experiments undertaken as part of this Thesis, experiment 4 sought to determine the extent, if any, of differences between the accuracy of predictions by time horizon *across* the two different treatment conditions. As previously reported in section 3.1.3, we are interested not in the influence of time horizon *per se*: we are only interested in horizon as it interacts with the accuracy of the Tool and various task representations – structured versus unstructured. As has been widely reported, and generally accepted, there is a tendency for forecasting accuracy to decrease as the forecast horizon increases (Dalrymple and King, 1981; Armstrong, 1985). It would be interesting to see if this tendency would be found to occur in experiment 4 (as it was in previous experiments) and to identify the treatment condition/s under which this was found to be more prominent.

The focus of experiment 4 is task complexity, more specifically, the effect of task complexity – manipulated by the addition of structure or decomposition – on task performance, as measured by judgemental forecast accuracy. The addition of structure was achieved by the provision of extra features within Tool 3 to enable the time series data to be decomposed or structured and thus display individual characteristics of the time series, such as trend and/or seasonality. This is essentially a change in the display format, which in turn, has a direct impact upon task complexity, rather than the Tool itself (for a more detailed discussion of the impact of structure and in particular, decomposition, please refer to section 2.3.4).

Subjects were presented with the same four time series as used for experiment 3 – B17, B26, M6 and F3. In order to decompose the time series data representations, by sub-patterns or individual characteristics in terms of trend and/or seasonality, we looked to the origin of the data. As previously noted in section 3.4.1, time series data were obtained from the M-Competition database. All the series within this database had previously been identified as being either seasonal or non-seasonal on the basis of autocorrelation analysis. The task characteristics of each of the four individual time series used in experiment 4 is presented in Table 7.2.1 below:

Table 7.2.1 Time Series Characteristics – Trend and/or Seasonality

		Seasonality	
		Y	N
Trend	Y	M6	F3
	N	B26	B17

In order to accommodate the inclusion of Trend and Seasonality in the time series representations for those subjects in the "structured" or decomposed treatment condition, the experiment included a line to denote trend and/or distinct lines inserted between each "set" of twelve (12) periods to denote seasonality. Seasonality has been defined as a structured pattern of changes within a year (Davey and Flores, 1993). In general, it is defined as a pattern which repeats itself over a series of fixed intervals in time. By providing subjects with a set of distinct lines denoting seasonality – in the structured treatment condition – we provided them with an additional aid to assist them with the forecasting task. (In order to view this graphically, please refer to Figure 7.2.2 below).

In terms of the interaction with the *technology* or the manipulation device, this was held constant for experiment 4. Both groups of subjects utilized Tool 3, using only the mouse, in order to articulate their needs and register their forecast predictions.

In terms of people, or individual characteristics as referred to by Goodhue and Thompson (1995), this was also held constant for experiment 4 – only postgraduate subjects undertook the experiment.

In summary, the characteristics of the task and the various tools employed for experiment 4 are presented in Table 7.2.2 below:

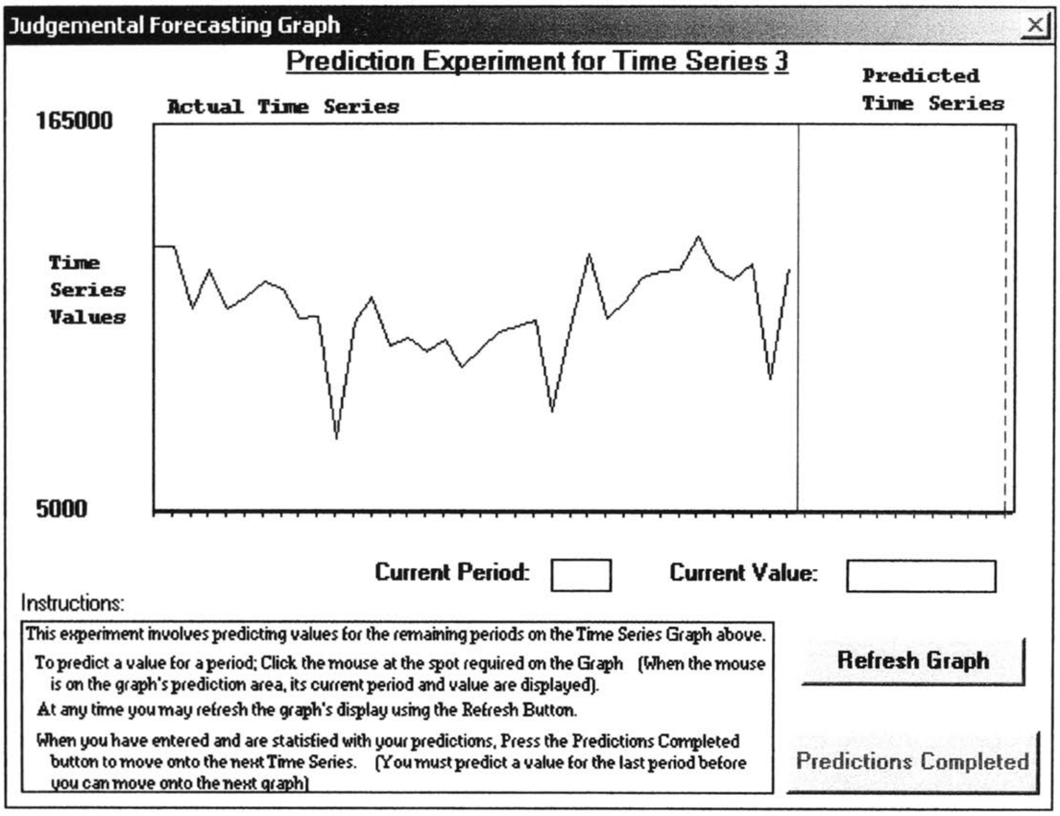
Table 7.2.2 Experimental Tool/s and Task Experiment 4

Tool_ID	Treatment Condition Task Complexity	Manipulation Device	Media
Tool 3*	Unstructured – more complex	mouse	screen-based
Tool 3S*	Structured – less complex	mouse	screen-based

* In order to facilitate readability of the following material, a graphic presentation of the Tools is provided below in Figures 7.2.1 and 7.2.2 respectively. The M6 time series has been chosen as the series to present as it has both trend and seasonality included in the structured presentation format.

Figure 7.2.1 Time Series Presentation[1] for Tool 3
Computer-supported Treatment

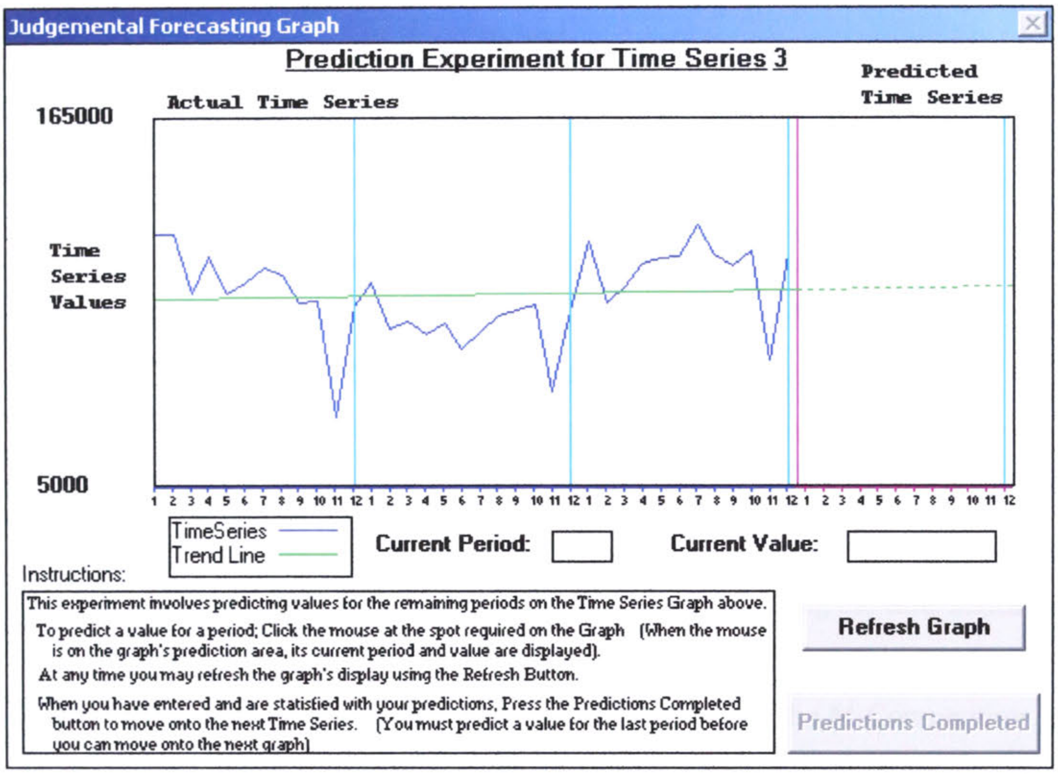
Tool 3: ***Media*** – screen-based; ***Technology*** – mouse
Task Complexity – unstructured, more complex



[1] Picture is smaller than actual size presented to task-doers.

Figure 7.2.2 Time Series Presentation[2] for Tool 3S
Computer-supported Treatment

Tool 3S: ***Media*** – screen-based; ***Technology*** – mouse
Task Complexity – structured, less complex



[2] Picture is smaller than actual size presented to task-doers.

Both groups of subjects completed the task using Tool 3, screen-based media. Subjects utilized the mouse in order to articulate their needs and register their forecast predictions. The major distinction between the two groups of subjects was the complexity of the task.

Given the results of previous experimental studies conducted as part of this Thesis, we propose that forecasting accuracy in the screen-supported medium could be further improved by decreasing task-complexity. It is hypothesised that the accuracy of predictions will be superior when the task is presented in a structured or decomposed mode. Hypothesis one states:

> **H1: The forecast accuracy of subjects presented with a structured or decomposed task will be higher than the forecast accuracy of subjects presented with a task which has not been structured or decomposed.**

7.3 Subjects and Experimental Procedure

Experiment 4 employed the use of forty six (46) postgraduate students who undertook the task using Tool 3 – screen-based media and related mouse *technology*. Twenty five (25) of the subjects completed the structured or decomposed task – deemed to be a less complex task, and the remaining twenty one (21) subjects completed the unstructured task – deemed to be more complex. The task was the judgemental extrapolation of time series data. Task complexity for both treatment conditions was manipulated by decomposing the time series representations to show trend and/or seasonality – where applicable. This was determined by the nature of the time series as identified via the M-Competition database. As previously noted in section 7.2 above, in order to accommodate the inclusion of Trend and Seasonality in the time series representations for those subjects in the "structured" or decomposed treatment condition, the experiment included a line to denote trend and/or distinct lines inserted between each "set" of twelve (12) periods to denote seasonality.

In keeping with Campbell's task classification, the decomposed judgemental forecasting task was deemed to be a "structured task" as well as being a "less complex" task.

Time series representations for experiment 4 remained the same as for experiments 2 and 3 (B17, B26, M6, F3). Postgraduate subjects were provided with the same four randomly presented time series and thirty six (36) periods of historical data for each of the time series. Subjects were asked to forecast the next twelve (12) periods ahead for each time series and no feedback was provided.

As for previous experiments all postgraduate subjects were undertaking courses in Information Systems at UNSW and were (arguably) highly proficient in the use of computers and related technology.

As previously advised, both groups of students undertook the structured task utilizing screen-based media. The primary difference between the two treatment conditions is Task Complexity – manipulated through the decomposition of time series data.

7.4 Analysis Methodology

As for previous experiments, the metric used to assess forecast accuracy and to compare treatments was the Mean Absolute Percentage Error (MAPE); with the Analysis of Variance (ANOVA) statistical procedure employed to compare the means of the various MAPE samples derived in experiment 4. To further facilitate analysis of data in terms of time horizon, MAPE values were derived for "groups" of forecast periods; periods 1 to 6; periods 7 to 12; and All 12 periods.

7.5 Results

7.5.1 Overall Results

The data collection process recorded predictions made by forty six (46) postgraduate subjects under two separate treatment conditions. All subjects used the same *media* and related *technology* – Tool 3 – screen-based media and mouse. The first group of twenty five (25) subjects completed the structured or decomposed task – deemed to be a less complex task, and the remaining twenty one (21) subjects completed the unstructured task – deemed to be more complex. In accordance with the hypothesis presented, data was collected for each treatment condition and across both time-horizons. A summary of overall results for all four (4) time series and across both treatment conditions is provided below in Table 7.5.1.1.

Table 7.5.1.1
Overall Results Summary – across Screen-based Media & Mouse Technology: A Comparison of Average MAPEs for Postgraduate Subjects via Task Complexity

Time Series	Series Type	Time Horizon	Task Complexity		p	Accuracy
			Unstructured	Structured		
B17	non-seasonal	First 6 periods	29.77	21.19	**0.0005**	S > U
	Trend = N	Last 6 periods	18.19	21.78	0.0526	S ≅ U
	Season = N	All 12 periods	23.98	21.49	0.0713	S ≅ U
B26	seasonal	First 6 periods	15.53	10.55	**0.0080**	S > U
	Trend = N	Last 6 periods	15.11	14.14	0.5034	S ≅ U
	Season = Y	All 12 periods	15.32	12.34	**0.0256**	S > U
M6	seasonal	First 6 periods	26.18	15.76	**0.0000**	S > U
	Trend = Y	Last 6 periods	16.53	22.76	**0.0128**	U > S
	Season = Y	All 12 periods	21.36	19.26	0.2733	S ≅ U
F3	non-seasonal	First 6 periods	3.14	3.53	0.4894	S ≅ U
	Trend = Y	Last 6 periods	3.38	3.75	0.5454	S ≅ U
	Season = N	All 12 periods	3.26	3.64	0.4522	S ≅ U

Note: Significance Values: $p < 0.05$ are emboldened.

LEGEND: S = Structured or decomposed, deemed to be less complex
U = Unstructured, deemed to be more complex
S > U indicates Structured is significantly better than Unstructured
S ≅ U indicates structured and unstructured overall equivalent

7.5.2 H1: Difference in Accuracy between Tasks – better accuracy with structured or decomposed Task.

Hypothesis 1 for experiment 4 predicted that there would be a difference in the accuracy of predictions between the structured or decomposed time series representations and the unstructured time series representations, and that a decrease in task complexity will have a positive effect on task performance in the judgemental extrapolation of time series data.

Hypothesis 1 further predicted that accuracy of forecast predictions in a judgemental time series extrapolation task would be superior for postgraduate subjects undertaking a structured or decomposed judgemental forecasting task.

As the mixed results in Table 7.5.1.1 show, H1 is not supported. It was expected that a significant difference in task performance would be reported between the two treatment conditions, especially for the two seasonal time series, B26 and M6. Clearly, this was not the case. It was expected, that the addition of structure, to both B26 and M6 – in terms of the inclusion of seasonality groupings and / or a trend line, would help to improve task performance – in terms of the accuracy of forecast predictions.

The overall results indicate that for the All 12 period, time horizon grouping for B26, there is a significant difference between the accuracy of forecasts between the two treatment conditions. Subjects undertaking the task in the structured or decomposed treatment condition were presented with seasonality groupings for B26. There was no trend line presented for B26. On closer examination the reported difference for the B26, All 12 period, time horizon grouping, can be attributed to the significant difference between the treatment conditions for the 1-6 period time horizon grouping. For B26, subjects undertaking the task in the structured or decomposed treatment condition produced significantly better results than those undertaking the task in the unstructured treatment condition. However, the same was not found to occur for forecast predictions in the 7-12 period time horizon grouping. Here the results indicate that the accuracy of predictions was almost equivalent across both treatment conditions. The addition of structure in terms of seasonality groupings did not assist task doers with predictions in the 7-12 period, essentially the further-out, time horizon grouping.

Therefore, for B26, although H1 is not fully supported across both time horizon groupings, we can report an overall significant difference between the two treatment conditions, with the overall accuracy of predictions in the structured or decomposed treatment condition being significantly superior to the overall accuracy of predictions in the unstructured treatment condition.

The overall results indicate that for the All 12 period time horizon grouping for M6, there is no significant difference in the accuracy of forecasts between postgraduate subjects in the structured treatment condition compared to those undertaking the task in the unstructured treatment condition. Subjects undertaking the task in the structured or decomposed treatment condition were presented with seasonality groupings and a trend line for M6. On closer examination there is a reported significant difference between the treatment conditions for the M6, 1-6 period time horizon grouping. Subjects undertaking the task in the structured or decomposed treatment condition produced significantly better results than those undertaking the task in the unstructured treatment condition. However, the exact opposite was found to occur for forecast predictions in the 7-12 period time horizon grouping. Here, the results indicate that the accuracy of predictions in the structured or decomposed treatment condition was significantly inferior to those in the unstructured treatment condition. The addition of structure in terms of seasonality groupings and trend line did not assist task doers with predictions in the 7-12 period – the *further-out* – time horizon grouping. In fact, the addition of structure seemed to "confound" them.

Therefore, for M6, H1 is not fully supported across both time horizon groupings. The mixed results are very confusing and will be discussed further in section 7.5.3 below. Perhaps the results can largely be attributed to the nature of the time series. Perhaps the downward spike present in the last quarter (periods 10-12) was just too difficult for subjects to predict. Or, perhaps the inclusion of the relatively flat trend line was a "confounding" factor which further inhibited subjects' performance in the task.

In terms of task performance for the two non-seasonal time series, B17 and F3, the results are fairly uniform. It was not expected that a significant difference in task performance would be reported between the two treatment conditions for the two non-seasonal time series. It was expected, that the addition of structure, to F3 only – in terms of the inclusion of a trend line (an upward sloping trend line) may have assisted task performance – in terms of the accuracy of forecast predictions. However, this was clearly not the case.

The overall results indicate that for the All 12 period time horizon grouping for both B17 and F3, there is no significant difference in the accuracy of forecasts between postgraduate subjects in the structured treatment condition compared to those undertaking the task in the unstructured treatment condition. Subjects undertaking the task in the structured or decomposed treatment condition were presented with a trend line (upward sloping) for F3 only, there was no trend line presented for B17.

On closer examination there is a reported difference for the B17, 1-6 period time horizon grouping with subjects in the structured or decomposed treatment condition producing more accurate results than those in the unstructured treatment condition. However, there is no plausible explanation for this. Both groups of subjects were presented with the same task, under essentially the same conditions. No structure was imposed in terms of either trend and/or seasonality, and the Tool was exactly the same, Tool 3 – screen-based with mouse technology. We are unable to offer any plausible explanation for the significant difference between the two treatment conditions in the 1-6 period time horizon grouping.

However, the same was not found to occur for forecast predictions in the 7-12 period time horizon grouping for B17. Here, the results indicate that the accuracy of predictions in the structured or decomposed treatment condition were equivalent to the results in the unstructured treatment condition. Given that there was no significant difference between the two treatment conditions for B17 – a non-seasonal, non-trended time series – in either task complexity, or Tool and related *technology*, this was in keeping with the results we had anticipated.

In terms of the accuracy of predictions for F3, there is no reported difference between the two treatment conditions across any of the time horizon groupings – either, 1-6; 7-12 or for the All 12 periods. The general results indicate that subjects in the structured or decomposed treatment condition produced equivalent results to those in the unstructured treatment condition. However, once again, this is not in keeping with what we had expected. Given that F3 is a non-seasonal time series exhibiting an upward trend, we believed that the inclusion of a trend line would assist task doers.

Our belief was based on general findings in the literature which indicate that forecast accuracy for upward sloping trends has been found to be superior to downward sloping trends (Lawrence and Makridakis, 1989, O'Connor *et al.*, 1993). We believed that the inclusion of a trend line would have assisted subjects with the task and in turn, would have led them to produce more accurate results (for a more detailed discussion on the influence of Trend please refer to section 2.3.3.1). However, as noted, this was not the case. Subjects undertaking the task in the structured or decomposed treatment condition did not produce better results that those undertaking the task in the unstructured treatment condition. There is no plausible explanation for this.

There is no clear indication provided of the effect of task decomposition, essentially representing a decrease in task complexity, on task accuracy, or more specifically the *degree* of TTF. As previously reported, the primary aim of experiment 4 was to review whether the *degree* of TTF – measured in terms of forecast accuracy – could be further improved by varying or more specifically reducing task complexity – one of the independent variables in the proposed Research Model (section 2.4). As has been noted, the results are not in keeping with what was hypothesized. This will be discussed further in the next section. (For a full review of reported statistics, please refer to Table 7.5.1.1).

7.6 Discussion

7.6.1 Main Findings

We used four monthly time series for the task in experiment 4 – the same time series that were used in experiments 2 and 3. As for previous experiments, the time series were randomly selected from the M-Competition database comprising 111 real time series (Makridakis *et al.,* 1982). In keeping with previous experiments, each group of subjects predicted the accuracy of forecasts for all four time series under each of the treatment conditions. In order to minimise any possible task fatigue effects, it was decided to only select and display a total of four time series to subjects as part of the experimental task.

The experimental results indicate that superior accuracy of the structured or decomposed treatment condition could not be proven. This was counter to what was hypothesised. We proposed that the treatment condition which presented a reduced task complexity would aid the task-doer and result in greater forecast accuracy. Superior accuracy could only be shown for B26, a seasonal time series for which no trend was displayed. B26 was one of the three time series presented in a structured or decomposed manner to task-doers (B26, M6 and F3). Superior accuracy for B26 could only be shown for two of the three time horizon groupings; the 1-6, and the All 12 periods, time horizon groupings. Superior accuracy was not found to hold for the 7-12 period, time horizon grouping.

The results for M6, a seasonal time series exhibiting a fairly flat trend line were even more varied. There was no significant difference between treatment conditions reported for the All 12 period, time horizon grouping. This was unexpected. However, the conflicting results for the 1-6 and 7-12 periods, time horizon groupings were also totally unexpected. As was predicted, subjects performed much better in the structured or decomposed treatment condition, and significantly so, but only for the 1-6 period, time horizon grouping. The exact opposite was found to occur in the 7-12 period, time horizon grouping, with subjects performing much better in the unstructured treatment condition. This was totally against what was hypothesised to occur and certainly not in keeping with what was predicted to occur according to the literature.

Goodhue and Thompson (1995) maintain that it is the interaction between the task, the technology and the individual which is crucial to the framework of TTF. In addition, they propose that an increase in task complexity, will lead to a subsequent decrease in TTF (p. 218). Our proposition was therefore, that a decrease in task complexity would lead to a subsequent increase in TTF. By manipulating task complexity in experiment 4, via task decomposition, or the addition of structure, especially for those time series deemed to be seasonal, it was expected that the general *level* or *degree* of TTF would be improved. In turn, this would lead to an overall improvement in task accuracy. However, this was not found to occur.

As has been suggested by Venkatraman and Prescott (1990), the following three steps would provide a good indication of the degree of TTF;

1. identification of distinct task environments
2. specification of ideal task / technological support for the task environment
3. identification of a suitable experiment to test performance effects of task/technology alignments, that is, to test the degree of TTF.

Essentially this is what we have been trying to achieve with the experimental studies which have been conducted as part of this Thesis. Experiment 4 sought to review the "ideal task support" for the task environment. By decomposing the judgemental time series extrapolation task into its unique components or sub-patterns, we sought to add structure to the task and simplify the task as much as possible. It was felt that this would lead to a better "fit" between the task and the technology and in turn, an increased *level* or *degree* of TTF. This was also in keeping with Campbell's notion that a decrease in uncertainty related to the task, could potentially decrease task complexity. As previously noted in section 2.5.4, Campbell reports that "Uncertainty can increase complexity by enlarging the pool of potential paths to a desired outcome" (p. 45). He further advises that a task-doer will experience an increase in information load if they are not able to establish a clear relationship between the information elements provided and the ultimate goal, or desired outcome. It was thought that the addition of structure to the judgemental time series extrapolation task – the decomposition of the task – would facilitate forecast predictions by decreasing the level of uncertainty within the task. However, the results do not provide conclusive evidence of whether or not the addition of structure was able to facilitate forecast predictions in the judgemental time series extrapolation task.

Our motivation for reducing task complexity was to examine the impact on the *degree* of TTF, in terms of task performance, or forecast accuracy. To put it another way, we were interested in determining whether the decomposition of time series data into distinct sub-patterns would facilitate the judgemental extrapolation of time series data and in turn, have a positive effect on task performance.

Our initial motivation was to examine the interaction between the variables presented in our proposed Research Model (section 2.4). In experiment 2, we altered the "People" variable. We used postgraduate subjects in experiment 2, as opposed to undergraduate subjects in experiment 1. In experiment 3 we varied the "Tool" variable and conducted further testing in order to review the impact on task performance. Reviewing the results of experiment 3 from the TTF perspective, we proposed that perhaps the *level* of interaction with the Tool, or the manipulation device in *combination* with the "complexity" of the task played a major role, or had a major impact on task performance. Essentially this was the motivation for experiment 4, to review whether task performance and therefore, the *degree* of TTF could be further improved by reducing task complexity. However, as a whole, the results of experiment 4 do not provide conclusive support for our proposed hypothesis. Additional experimentation is therefore required in order to provide some conclusive answers to the following question which still remains;

> How can we best identify and operationalise the independent variables in our proposed theoretical model of TTF in order to ensure the best *level* or *degree* of "fit", which in turn would ensure optimum task performance?

This question remains unanswered and is beyond the scope of the current Thesis.

7.7 Future Research

Experiment 2 concentrated on the "People" variable as presented in the proposed Research Model (section 2.4), experiment 3 concentrated on the effects of Tool upon TTF, and experiment 4 concentrated on the last remaining variable in the proposed Research Model – namely that of Task Complexity. The findings in experiment 4 lead us to a number of questions which could be investigated in future research.

As previously mentioned, it was always our desire to have each group of subjects predict the accuracy of forecasts for all of the time series presented under each of the treatment conditions. In order to minimise any possible task fatigue effects it was therefore decided to only select and display four time series to subjects as part of each of the experimental studies undertaken. However, in hindsight, the selection of only four time series and our desire to balance the "type" of series selected, meant that we had a small number of time series against which to test our hypothesis in experiment 4. Ideally, it would have been advantageous to have been able to conduct experiment 4 with a larger number of time series in totality. However, the issue of task fatigue would have remained a problem for consideration. This is a potential area for future research in order to further the field of TTF.

Another area of interest is that perhaps the actual time series selected for the experimental studies may not have been optimal. Perhaps the very nature of the time series selected had an impact upon task performance, making it difficult to attribute findings solely to the treatment condition/s. It would be interesting to conduct further research in this area with a different set of time series data.

In terms of reducing task complexity, this was achieved by adding structure to the time series data. We chose to decompose the task by identifying the unique sub-patterns within the time series, such as the presence of trend and/or seasonality. In keeping with previous studies, there was no feedback provided to subjects. Perhaps the task itself could have been further simplified through the provision of feedback. This is also a possible area for future research.

In terms of TTF, the question still remains; how can we best identify the optimum task/technology alignments, that is, the degree of TTF, in order to maximise task performance?

7.8 Conclusion

Experiment 4 has explored differences in the accuracy of predictions in a judgemental extrapolation task by manipulating task complexity. We reduced task complexity by decomposing time series data representations to identify unique trend and/or seasonality sub-patterns. The task was undertaken by postgraduate subjects using a screen-based Tool with mouse *technology*. Task accuracy was compared against results produced by postgraduate subjects undertaking essentially the same task but with no decomposition.

We were interested in determining the impact upon task performance for subjects using a more sophisticated screen-based Tool, one with a decreased level of interaction with the manipulation device – mouse only – as well as an overall decrease in task complexity. Our motivation in varying task complexity was to determine whether the interaction between the Tool and the Task would have an impact upon task performance – measured in terms of the accuracy of forecast predictions. In turn, this would translate into a change in the *level* or *degree* of TTF. Our hypothesis was that the use of a more sophisticated Tool, coupled with a decrease in task complexity, would improve the *level* or *degree* of TTF and in turn task performance, measured in terms of forecast accuracy. Findings from experiment 4 were not able to provide conclusive evidence in order to support our proposed hypothesis. However, these findings are, of course, qualified by the simplified version of the forecasting task used in experiment 4.

In conclusion, the question at the heart of Goodhue and Thompson's TTF model (1995), and essentially, of the proposed Research Model for the current Thesis remains;

> How can we best identify the optimum task / technology alignments, in order to maximise TTF, and in turn, maximise task performance?

8. Conclusion

8. Conclusion

8.1 Introduction

We conducted a series of four experiments in order to seek insight into the following question;

> Are there any differences in judgemental forecasting accuracy that are due to the Tool – i.e. the *media* and related *technology* – that is used in the forecasting process?

This was the overall research question for this Thesis. However, we also explored this question in different task complexity conditions (no decomposition versus decomposition) and with different cohorts of subjects (undergraduate students versus postgraduate students). We were interested in not only the potential effect of the task environment, or the Tool itself, but also the nature of the Task, as well as the nature of the people engaged in the judgemental forecasting task. These components – Tool (media and accompanying technology), people undertaking the task, and task complexity – were presented in our Theoretical Model (section 2.4). The primary purpose of our experimental studies was to identify a relationship between the components. This relationship was measured in terms of task performance, or more precisely, the accuracy of predictions in the judgemental extrapolation of time series data.

Our review of the literature concentrated on four major areas; Task-Technology Fit (TTF), Judgemental Forecasting, Media and Task itself. The primary focus of our proposed Theoretical Model (section 2.4) being Goodhue and Thompson's 1995 model of TTF and Campbell's 1988 Task Complexity Framework. In addition, we were also interested in the results of previous Media studies, primarily in terms of the presentation *mode* – paper versus screen-based.

However, our review of the literature also encompassed presentation *format*, especially in terms of the graphs versus tables literature – which in turn was heavily cited in the TTF literature.

The basic premise behind Goodhue and Thompson's 1995 theoretical model is that:

> "Task-Technology Fit (TTF) is the degree to which a technology assists an individual in performing his or her portfolio of tasks. More specifically, TTF is the correspondence between task requirements, individual abilities, and the functionality of the technology." (p. 218).

This Thesis sought to further extend the area of TTF by conducting a series of experimental studies, varying one essential component each time. In order to determine the complexity of our task, we applied Campbell's 1988 framework to our experimental task – the judgemental extrapolation of time series data. We also applied Campbell's task complexity framework to the other major tasks undertaken in previous Media studies – reading for comprehension and proof-reading. The results were presented in Table 4.1 (as well as sections 2.5.4 and 2.6.7.3) and are reproduced in Table 8.1.1 below (for a more detailed discussion please refer to the individual sections).

Table 8.1.1 The Fundamental Complexity* of the three experimental tasks: Judgemental extrapolation; Reading and Proof-Reading

Task	**Outcomes Source 2**	**Interdependencies Source 3**	**Uncertainty Source 4**	**Level of Manipulation device Interaction**
Judgemental Extrapolation	—	Moderate-High	Moderate-High	Moderate-High
Reading for Comprehension	Low	Moderate-High	Moderate-High	Low
Proof-Reading	Low	Low	Low	Moderate-High

*Source 1 (paths) has been omitted from Table 8.1.1, it was not present in any of the three experimental tasks

We were interested in conducting a series of experimental studies in order to determine whether the Tool – i.e. the *media* and related *technology* – would have an impact upon task performance. Our task, or *object of interest,* was the judgemental extrapolation of time series data and our measure of task performance was the accuracy of forecast predictions. Our experimental task had a lot in common with both of the tasks previously undertaken in Media studies. We noted that it would be interesting to see if previous experimental results – as reported for Media studies (section 2.6) – would be replicated in our own experimental studies. Previous findings indicated that subjects undertaking the reading for comprehension task performed as well in the paper-based medium, as they did in the screen based-medium. However, experimental results for the proof-reading task identified that subjects were significantly more accurate in the paper-based medium compared to the screen-based medium. We were interested in finding out whether subjects undertaking the time series extrapolation task would produce results similar to those of reading for comprehension, or proof-reading, or whether, in fact, they would produce an entirely different set of results. This was the motivation for undertaking a series of four experimental studies. Our summarised findings are presented below.

8.2 Findings from Experimental Studies

Experiment 1 explored differences in the accuracy of predictions in a judgemental extrapolation task when that task was presented across two different presentation and working media; namely paper-based and screen-supported. In terms of our experimental variables, the people and task characteristics were held constant. Only the Tool – *media* and the related *technology* – was varied. In addition, experiment 1 sought to determine whether the accuracy of predictions in a judgemental extrapolation task differed between Tools, across the time horizon groupings (periods 1-6 and periods 7-12) for the presented series.

The main finding of experiment 1 was that the task of judgemental extrapolation was more accurately undertaken in the paper-based medium than the screen-supported medium. This result was in keeping with findings from proof-reading Media studies. We noted that in terms of its relationship to the *technology*, i.e. the level of interaction with the manipulation device, our experimental task was similar to that of proof-reading. Both tasks required subjects to interact with the manipulation device in order to articulate their needs whilst undertaking the task.

However, we had also speculated that given the similarities with the task of reading for comprehension, the use of screen-media and the related technology, should have no impact on the accuracy of predictions in our judgemental forecasting task. Our hypothesis that there would be no difference in the accuracy of predictions between paper-based and screen-based *media* and related *technologies* was clearly not supported by the findings from our first experimental study. Experiment 1 also showed that forecast accuracy decreased over time horizon under each of the treatment conditions and that the decrease in accuracy was generally more pronounced in the screen-supported medium.

Findings and limitations identified by the analysis of data and experimental procedures highlighted experimental subjects – in terms of age and level of maturity – as an area to be addressed by future research efforts. This was the basis for undertaking experiment 2 – essentially a replication of experiment 1 – using postgraduate subjects. For the purposes of our experiment postgraduate subjects were deemed to be both older and more mature than undergraduate subjects. In terms of our experimental variables, task characteristics were held constant. Both the Tool – *media* and the related *technology* – and the people were varied.

In experiment 2 we explored differences in the accuracy of predictions in a judgemental extrapolation task. The task was undertaken by postgraduate subjects, and was presented across two different presentation and working media; paper-based and screen-supported. Post-hoc testing was conducted to try to determine the exact nature or extent of any reported differences. In addition, experiment 2 sought to determine whether the accuracy of predictions in a judgemental extrapolation task differed between Tools, across the time horizon groupings (periods 1-6 and periods 7-12), for the presented series.

The main finding of experiment 2 was that when undertaken by postgraduate subjects, the task of judgemental extrapolation was more accurately undertaken in the screen-supported medium rather than the paper-based medium. Once again, our hypothesis that there would be no difference in accuracy between the media was not supported. Results from experiment 2 were completely the opposite compared to findings reported for experiment 1. In addition, this was also a completely different result to any of the previous Media studies. Neither reading for comprehension, nor proof-reading studies had ever produced results to show superiority of the screen-based medium. We were interested in examining this further.

We speculated that findings indicated that the *degree* of TTF was somewhat improved when individuals who were older and more mature– perhaps one could say, more experienced – utilized a Tool with the following characteristics – screen-based *media* and related *technology*. Given our general results and the results from post-hoc testing, we were interested in reviewing whether the *degree* of TTF – measured in terms of forecast accuracy – was able to be further improved by varying one of the other independent variables in our proposed Research Model, namely, the Tool.

The screen-based Tool used in both experiments 1 and 2 had required subjects to use a combination of slider, mouse and keyboard in order to articulate their needs and register their forecast predictions. We were interested in improving or simplifying the manipulation mechanism in the screen-based Tool. In particular we were interested in undertaking further experimental studies in order to seek to provide answers to the following question:

> Would a more sophisticated Tool – perhaps one with a decreased level of interaction with the manipulation device – *improve* the degree of TTF, and in turn the accuracy of forecast predictions in a judgemental extrapolation task?

This was the basis for undertaking experiment 3 which explored differences in the accuracy of predictions in a judgemental extrapolation task when that task was undertaken by postgraduate subjects using two different types of screen-based Tools.

In experiment 3 we were interested in determining the impact upon task performance for subjects using a more sophisticated screen-based Tool, one with a decreased level of interaction with the manipulation device – mouse only. Our motivation in varying the Tool was to determine whether the different *level* and *extent* of the interaction with the manipulation device would have any impact upon task performance – measured in terms of the accuracy of forecast predictions for postgraduate subjects.

In terms of our experimental variables, the people and task characteristics were held constant. Only the *technology* related to the screen-based Tool was varied – the *media* component was held constant. Experiment 3 was only undertaken in the screen-based medium.

In addition, experiment 3 sought to determine whether the accuracy of predictions in a judgemental extrapolation task differed between Tools, across the time horizon groupings (periods 1-6 and periods 7-12), for the presented series.

There were essentially two main findings from experiment 3. The first being, that for up-close forecast predictions (i.e. periods 1-6), subjects undertaking the judgmental time series extrapolation task were not adversely affected by the *level* or *degree* of interaction with the manipulation device. Simply put, subjects performed just as well using the less sophisticated, more cumbersome Tool (keyboard, mouse and slider) in the 1-6 period time horizon grouping. In fact, results showed that sometimes subjects also performed significantly better using the less sophisticated, Tool in the 1-6 period time horizon grouping.

The second finding was that for forecast predictions further along the time horizon (i.e. periods 7-12), subjects undertaking the judgmental time series extrapolation task were adversely affected by the *level* or *degree* of interaction with the manipulation device. Or, to put it another way, subjects using the more advanced, less cumbersome Tool (mouse only) performed better in the 7-12 period time horizon grouping, sometimes significantly so.

We speculated that findings from experiment 3 indicated that the *degree* of TTF was somewhat improved when the *level* or *degree* of interaction with the manipulation device was reduced, particularly so for forecast predictions further along the time horizon. Looking to the literature in relation to judgemental extrapolation, it is widely reported that forecasting accuracy tends to decrease as the forecast horizon increases (Dalrymple and King, 1981; Armstrong, 1985 – for a more detailed discussion, please refer to section 2.3.3.3). We proposed that this might be linked to an increase in task complexity. We further proposed that the task of predicting the value of future periods became more complex as the time horizon increased. We were therefore interested in somehow “decreasing” task complexity in order to test our proposition.

Having reviewed the impact on task performance by varying two of the three experimental variables, namely the People and the Tool, we were interested in examining the effects of varying the last remaining variable – Task Complexity. We wanted to determine whether forecasting accuracy in the screen-supported medium could be further improved by decreasing task complexity.

Looking to the literature in relation to judgemental forecasting, it has often been said that "decomposition" enables a complex task to be reduced into smaller and "cognitively more manageable parts" (MacGregor and Armstrong 1994, p. 32). By decomposing a task into several smaller components or subsets, we allow the subject to concentrate on one thing at a time. Thus by definition, decomposition may be seen to be decreasing the complexity of the task (Webby and O'Connor, 1996). Our wish was to explore differences in the accuracy of predictions in a judgemental extrapolation task by manipulating task complexity. In particular, we were interested in undertaking further experimental studies in order to seek to provide answers to the following question:

> Would the decomposition of the judgemental time series extrapolation task, coupled with the use of a more sophisticated Tool – one with a decreased *level* of interaction with the manipulation device – improve the *degree* of TTF and in turn, task performance?

This was the basis for undertaking experiment 4.

We reduced task complexity in experiment 4 by decomposing time series data representations to identify unique trend and/or seasonality sub-patterns. The task was undertaken by postgraduate subjects using the screen-based Tool with mouse *technology*. Task accuracy was compared against results produced by postgraduate subjects undertaking essentially the same task using the same screen-based Tool, but with no decomposition.

As for previous experiments, we sought to determine whether the accuracy of predictions in a judgemental extrapolation task differed between Tools, across the time horizon groupings (periods 1-6 and periods 7-12), for the presented series.

Our motivation in varying task complexity was to determine whether the interaction between the Tool and the Task would have an impact upon task performance. In turn, this would translate into a change in the *level* or *degree* of TTF. We sought to identify the "ideal task support" for the task environment. By decomposing the judgemental time series extrapolation task into its unique components or sub-patterns, we added structure and simplified the task as much as possible. It was felt that this would lead to a better "fit" between the task and the technology and in turn, an increased *level* or *degree* of TTF. This was also in keeping with Campbell's notion that a decrease in uncertainty related to the task, could potentially decrease task complexity (for a more detailed discussion please refer to section 2.5.4).

Our hypothesis was that the use of a more sophisticated Tool, coupled with a decrease in task complexity, would improve the *level* or *degree* of TTF and in turn task performance, or forecast accuracy. It was thought that the addition of structure to the judgemental time series extrapolation task – the decomposition of the task – would facilitate forecast predictions by decreasing the level of uncertainty within the task. It was expected that a significant difference in task performance would be reported between the two treatment conditions, especially for the seasonal time series. It was hypothesized that the addition of structure – in terms of the inclusion of seasonality groupings and / or a trend line – would have helped to an improvement in TTF and thus task performance – in terms of the accuracy of forecast predictions. However, the results of experiment 4 were not able to provide conclusive evidence of whether or not the addition of structure was able to facilitate forecast predictions in the judgemental time series extrapolation task.

Findings were mixed and indicate that superior accuracy of the structured or decomposed treatment condition could not be proven. Experiment 4 was not able to provide conclusive evidence in order to support our proposed hypothesis. Given the results from experiment 3, this was unexpected.

A summary of experimental results in presented in Table 8.2.1 below.

Table 8.2.1 Summary of Experimental Results

Exp ID	Subjects	Tool ID	Media	Manipulation Device	Task Complexity	Overall Result
1	U/G	Tool 1	Screen	slider, mouse & keyboard	• Unstructured • no decomposition	
1	U/G	Tool 2	Paper	pencil & eraser	• Unstructured • no decomposition	
						P > S
2	P/G	Tool 1	Screen	slider, mouse & keyboard	• Unstructured • no decomposition	
2	P/G	Tool 2	Paper	pencil & eraser	• Unstructured • no decomposition	s > p
3	P/G	Tool 1	Screen	slider, mouse & keyboard	• Unstructured • no decomposition	
3	P/G	Tool 3	Screen	Mouse	• Unstructured • no decomposition	
						T1 ≅ T3
4	P/G	Tool 3	Screen	Mouse	• Structured • decomposition	
4	P/G	Tool 3	Screen	Mouse	• Unstructured • no decomposition	
						S ≅ U

LEGEND: P = Paper; S = Screen;
P > S indicates paper is usually better than screen and frequently significantly better
s > p indicates screen is usually better than paper but not frequently significantly better
T1 = Tool 1 (slider, mouse and keyboard); T3 = Tool 3 (mouse)
T1 ≅ T3 indicates Tool 1 and Tool 3 overall equivalent
S = Structured or decomposed (less complex); U = Unstructured (more complex)
S ≅ U indicates structured and unstructured overall equivalent

In conclusion, the question at the heart of Goodhue and Thompson's TTF model (1995), and essentially, of the proposed Research Model for the current Thesis remains;

> How can we best identify the optimum task / technology alignments, in order to maximise TTF, and in turn, maximise task performance?

As previously noted this question remains unanswered and is beyond the scope of the current Thesis.

8.3 Future Research

In addition to addressing the limitations of this Thesis, there remain a number of issues which could be investigated in future research. These include the issue of fatigue, the nature and type of time series, the provision of feedback to subjects, as well as the task itself. Perhaps the judgemental extrapolation of time series data was not an optimum task to be used in order to determine or test the degree of TTF? Perhaps the selection of a different task may have facilitated the review of the relationship and interaction between the individual components in the proposed Research Model? These are all areas to be explored and examined further as part of future research in order to extend the area of Task-Technology Fit.

We identified three key variables; People; Tool – *media* and related *technology* and Task Complexity in our proposed Research Model (section 2.4). Of interest to this Thesis was the interaction between the variables and the effect upon task performance. This was identified as the degree of Task-Technology Fit, or TTF. Our overall aim was to review the impact of *media* and the related *technology* on task performance and to extend the area by conducting such research on a task that was deemed to be relevant in a decision making context. We identified that the chosen task must also be considered to be an essential component of many business decisions. The task also had to provide an objective measure of task performance. In attempting to satisfy all of the above requirements, we selected the task of judgemental extrapolation of time series data.

Using Goodhue and Thompson's (1995) model of TTF, we identified that the question at the heart of their model was whether the *degree* of fit or, TTF, could be improved by manipulating the independent variables. Our aim was to further the area of TTF by conducting research on a task that would provide an objective accuracy measure. We wanted to identify the optimum task / technology alignments required in order to maximise TTF, and in turn, maximise task performance. Although this Thesis has provided some conclusive evidence, it has also raised many questions and identified many new areas of future research. It has therefore achieved its goal of extending the area of TTF.

In conclusion, the purpose of the Thesis was to explore Media and Technology effects in the judgemental extrapolation Task. The experimental studies conducted as part of this Thesis have identified that *media* and related *technology* do have an impact upon task performance. In turn, results of the experimental studies have identified possibilities for future research to be conducted in order to further extend the general area of Task-Technology Fit. The ultimate aim is that findings from future research will lead to a refinement of the model of TTF, for future use by researchers and practitioners seeking to measure the effectiveness of organisational information systems.

9. Bibliography

1. Andreassen, P.B. (1988). Explaining the price-volume relationship: The difference between prince changes and changing prices. *Organizational Behavior and Human Decision Processes*, V. 41, pp. 371- 389.

2. Andreassen, P.B. and Kraus, S.J. (1990). Judgemental Extrapolation and the Salience of Change. *Journal of Forecasting,* V. 9, pp. 347 - 372.

3. Armstrong, J.S. (1983). Relative accuracy of judgemental and extrapolative methods in forecasting annual earnings. *Journal of Forecasting,* V. 2, pp. 437 - 447.

4. Armstrong, J.S. *Long Range Forecasting: From Crystal Ball to Computer*, 2nd Ed. New York: John Wiley & Sons, 1985.

5. Armstrong, J.S. (2001). The Forecasting Dictionary, J. S. Armstrong (ed.) *Principles of Forecasting,* Norwell, MA: Kluwer Academic Publishers.

6. Armstrong, J.S. and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting,* 8, pp. 69-80.

7. Ashton, A.H., (1984). A field test of implications of laboratory studies of decision making. *Accounting Review,* V. 59 (3), pp. 361-375.

8. Askwall, S. (1985). Computer Supported Reading Vs Reading Text on Paper: A Comparison of two Reading Situations. *International Journal of Man-Machine Studies,* 22, pp. 425-439.

9. Baecker, R.M., Grudin, J., Buxton, W.A. and Greenberg, S. *Readings in Human-Computer Interaction: Toward the Year 2000,* Morgan Kauffman Publishers, San Francisco, CA, 1995.

10. Baron, R.S. (1986). Distraction-Conflict Theory: Progress and Problems. *Advances in Experimental Social Psychology*, 19, pp.1-39.

11. Bauer, D., Bonacker, M. and Cavonius, C.R. (1983). Frame Repetition rate for Flicker-free Viewing of Bright VDU Screens, *Displays,* Jan. pp. 31-33.

12. Beach, L.R. and Mitchell, T.R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review,* V. 3, pp. 439-449.

13. Bettman, J.R., Johnson, E.J. and Payne, J.W. (1990). A Componential Analysis of Cognitive Effort in Choice. *Organizational Behavior and Human Decision Processes*, V. 45, pp. 111- 139.

14. Black, A. (1990). Visible planning on paper and on screen: The impact of working medium on decision-making by novice graphic designers. *Behaviour & Information Technology*, V. 9 (4), pp. 283- 296.

15. Blalock, H.M., *Social Statistics.* McGraw-Hill, New York NY, 1960.

16. Blocher, E., Moffie, R.P. and Zmud, R.W. (1986) Report Format and Task Complexity; Interaction in risk judgements. *Accounting, Organizations and Society,* V. 11(6), pp. 457-470.

17. Campbell, D.J. And Gingrich, K.F. (1986). The Interactive Effects of Task Complexity and Participation on Task Performance: A Field Experiment. *Organizational Behavior and Human Decision Processes*, V. 38 (2), pp. 162- 180.

18. Campbell, D.J. (1988). Task Complexity: A Review and Analysis. *Academy of Management Review,* V. 13 (1), pp. 40-52.

19. Campbell, S.K., *Applied Business Statistics: Test, Problems and Cases.* Harper & Row Publishers, New York NY, 1987.

20. Carbone, R. and Armstrong, J.S. (1982). Evaluation of extrapolative forecasting methods: Results of a survey of academicians and practitioners. *Journal of Forecasting,* V. 1, pp. 215-217.

21. Carroll, J.B. *Language Comprehension and the Acquisition of Knowledge.* New York: Wiley, 1972.

22. Champion, D.J., *Basic Statistics for Social Research,* Chandler Publishing, Scranton, 1970.

23. Chatfield, C. (1988). Apples, oranges and mean square error. *International Journal of Forecasting*, V. 4, pp. 515-518.

24. Cleveland, W.S. and Terpenning, I.J., (1982), Graphical Methods for Seasonal Adjustment, *Journal of the American Statistical Association* , 77, 377, pp. 52-62.

25. Cohen, J.D., Dunbar, K. And McClelland, J.L. (1990). On the Control of Automatic Processes: A Parallel Distributed Processing Account of the Stroop Effect. *Psychological Review,* V. 97 (3), pp. 332-361.

26. Connolly, T. And Serre. P. (1984). Information search in judgement tasks: the effects of unequal cue validity and cost. *Organizational Behavior and Human Decision Processes*, V. 34, pp. 387-401.

27. Cooper, R.A. and Weekes, A.J., "*Data, Models and Statistical Analysis*"; Philip Allan Publishers, Oxford, UK, 1983.

28. Corragio, L. *Deleterious Effects of Intermittent Interruptions on the Task Performance of Knowledge Workers: A Laboratory Investigation,* unpublished doctoral dissertation, University of Arizona, 1990.

29. Creed, A., Dennis, I. and Newstead, S. (1987). Proof-reading on VDUs. *Behaviour & Information Technology*, V. 6 (1), pp. 3-13.

30. Cushman, W. H. (1986). Reading from Microfiche, a VDT, and the printed page; subjective fatigue and performance. *Human Factors*, 27, pp. 189-200.

31. D'Ambra, J., and Rice, R.E. (2001), Emerging factors in user evaluation of the World Wide Web, *Information & Management,* V. 38, pp. 373-384.

32. Dalrymple, D.J. (1987). Sales Forecasting practices: Results from a United States survey. *International Journal of Forecasting,* V. 3, pp. 379-391.

33. Dalrymple, D.J. and King, B.E. (1981). Selecting Parameters for short-term Forecasting Techniques. *Decision Sciences,* V. 12, pp. 661-669.

34. Davey, A.M. and Flores, B.E. (1993). Identification of Seasonality in Time Series: A Note. *Mathematical Computing Modelling,*V. 18 (6), pp. 73-81.

35. Dickson, G.W., DeSanctis, G., and McBride, D.J., (1986), Understanding the Effectiveness of Computer Graphics for Decision Support: A Cumulative Experimental Approach, *Communications of the ACM,* V. 29 (1), pp. 40-47.

36. Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics,* V. 35 (10), pp. 1297-1326.

37. Downton, A. and Leedham, G. (1993). Human Aspects of Human-Computer Interaction, in A. Downton (ed.) *Engineering the Human-Computer Interface*. London: McGraw-Hill, Ch 2.

38. Downton, A. *Engineering the Human-Computer Interface.* London: McGraw Hill, 1993.

39. Edmundson, R.H. and Terry, J.E., (1986), Graphical Data Presentation for Decision Support Systems, *The Australian Computer Journal*, V.18 (4), pp. 191-195.

40. Ericsson, K.A. and Simon, H.A. (1980). Verbal Reports as Data. *Psychological Review*, V. 87 (3), pp. 215-251.

41. Ferguson, G.A. (1956). On Transfer and Human Ability, *Canadian Journal of Psychology*. V. 10, pp. 121-130.

42. Fildes, R. (1992). The Evaluation of Extrapolative Forecasting Methods, *International Journal of Forecasting*, V. 8, pp. 81–98.

43. Fleishman, E.A. and Quaintance, M.K. (1984) *Taxonomy of Human Performance: The Description of Human Tasks*. Orlando, FL: Academic Press, Inc.

44. Garland, J. (1982), *Ken Garland and Associates: Designers - 20 years work and play*. Cited in, Waller, R. (1978). The Typographic Contribution to Language: Towards a Model of Typographic Genres and their Underlying Structures, PhD Thesis, Dept. of Typography and Graphic Communication, University of Reading.

45. Goodhue, D.L. (1988), IS Attitudes: Toward Theoretical and Definition Clarity, *DataBase,* V. 19 (3/4), pp. 6- 15.

46. Goodhue, D.L. and Thompson, R.L. (1995), Task-Technology Fit and Individual Performance, *MIS Quarterly,* V. 19 (2), pp. 213-236.

47. Goodhue, D.L. (1995), Understanding User Evaluations of Information Systems, *Management Science,* V. 41 (12), pp. 1827-1844.

48. Goodhue, D.L. (1998), Development and Measurement Validity of a Task-Technology Fit Instrument for User Evaluations of Information Systems, *Decision Sciences,* V. 29 (1), pp. 105-138.

49. Goodwin, P., Enhancing Judgemental Sales Forecasting. In G. Wright and P. Goodwin (eds.) *Forecasting with Judgment*, England: John Wiley & Sons, 1998, Ch. 4.

50. Goodwin, P. and Wright, G., (1993), Improving Judgemental Time Series Forecasting: a Review of the Guidance Provided by Research, *International Journal of Forecasting*, V. 9, pp. 147-161.

51. Gould. J.D. and Grischkowsky, N. (1984). Doing the Same Work with Hard Copy and with Cathode-Ray Tube (CRT) Computer Terminals. *Human Factors*, V. 26 (3), pp. 323-337.

52. Gould. J.D., Alfaro. L., Barnes. V., Finn, R., Grischkowsky, N. and Minuto, A. (1987a). Reading Is Slower from CRT Displays than from Paper: Attempts to Isolate a Single-Variable Explanation. *Human Factors*, V. 29 (3), pp. 269-299.

53. Gould. J.D., Alfaro. L., Finn, R., Haupt. B. and Minuto, A. (1987b). Reading from CRT Displays Can Be as Fast as Reading from Paper. *Human Factors*, V. 29 (4), pp. 497- 517.

54. Grandjean, E and Vigliani, E. *Ergonomic Aspects of Visual Display Terminals.* London: Taylor & Francis, 1982.

55. Hackman, J.R. (1969). Toward understanding the role of Tasks in Behavioral research. *Acta Psychologica*, V. 31, pp. 97-128.

56. Hand, D.J., and Taylor, C.C., *Multivariate Analysis of Variance and Repeated Measures: A Practical Approach for Behavioural Scientists*, Chapman and Hall, London, 1987.

57. Handzic, M. (1997). *The Utilization of Contextual Information in a Judgemental Decision Making Task.* Unpublished PhD thesis, University of New South Wales (UNSW).

58. Hansen, W.J., Doring, R., and Whitlock, L.R. (1978). Why an examination was slower on-line than on paper. *International Journal of Man Machine Studies,* V. 10 (5), pp. 507- 519.

59. Hansen, W.J., and Haas, C. (1988). Reading and writing with Computers: A framework for explaining differences in performance. *Communications of the ACM,* V. 31 (9), pp. 1080-1089.

60. Harvey, N. (2001). Improving Judgemental Forecast, in J. S. Armstrong (ed.) *Principles of Forecasting*, Norwell, MA: Kluwer Academic Publishers.

61. Harvey, N., and Bolger, F. (1996). Graphs versus tables: effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, V. 12 (1), pp. 119-137.

62. Harvey, N., Ewart, T. and West, R. (1997). Effects of Data Noise on Statistical Judgement, *Thinking and Reasoning,* V3, pp. 111-132.

63. Helander, M. G., Billingsley, P.A. and Schurick, J.M. (1984), An Evaluation of Human Factors Research on Visual Display Terminals in the Workplace, *Human Factors Review,* Chapter 3, pp. 55 – 129.

64. Huey, E.B. (1908). *The Psychology and Pedagogy of Reading.* New York: Macmillan (Reprinted by MIT Press, 1968).

65. Jacko, J.A. And Salvendy, G. (1996). Hierarchical Menu Design: Breadth, Depth and Task Complexity. *Perceptual and Motor Skills,* V. 82 (3), pp. 1187-1201.

66. Jacko, J.A., Salvendy, G. And Koubek, R.J. (1995). Modeling of Menu Design in Computerized Work. *Interacting with Computers*, V. 7 (3), pp. 304-330.

67. Jacko, J.A. And Ward, K.G. (1996). Toward Establishing a Link Between Psychomotor Task Complexity and Human Information Processing. *Proceedings of the 19th International Conference on Computers and Industrial Engineering,* V. 13 (1), pp. 533 -536.

68. Just, M. A. And Carpenter, P. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review,* V. 87 (4), pp. 329 - 354.

69. Just, M. A. And Carpenter, P. (1992). A Capacity Theory of Comprehension: Individual Differences in Working Memory. *Psychological Review,* V. 99 (1), pp. 122-149.

70. Kagan, D. and Pietron, L.R. (1987). Cognitive Level and Achievement in Computer Literacy. *The Journal of Psychology,* V. 121 (4) pp. 317-327.

71. Kahneman, D. *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall, 1973.

72. Kak, A.U. (1981). Relationships between readability of printed and CRT-displayed text. *Proceedings of the 25th annual meeting of the Human Factors Society.* Baltimore: HFS.

73. Kasper, G.M. and Cerveny, R.P. (1985). Laboratory Study Of User Characteristics And Decision-Making Performance In End-User Computing. *Information & Management,* V. 9 (2) pp. 87-96.

74. Kintsch, W. *The Representation of Meaning in Memory.* Lawrence Erlbaum, Hillsdale, NJ, 1974.

75. Kruk, R.S. And Muter, P. (1984). Reading Continuous text on video screens. *Human Factors,* 26, pp. 339 - 345.

76. Lancaster, F.W., *Towards Paperless Information Systems.* New York: Academic Press, 1978.

77. Lawrence, M.J. (1983). An extrapolation of some practical issues in the use of quantitative forecasting models, *Journal of Forecasting,* V. 2, pp. 169-179.

78. Lawrence, M.J., Edmundson, R.H. and O'Connor, M.J. (1985). An Examination of Accuracy of Judgemental Extrapolation of Time Series, *International Journal of Forecasting,* V. 1, pp. 25-35.

79. Lawrence, M.J. and O'Connor, M.J. (1992). Exploring judgemental forecasting, *International Journal of Forecasting,* V. 8, pp. 15-26.

80. Lawrence, M.J., O'Connor, M.J. and Edmundson, R.H. (2000). A field study of sales forecasting accuracy and processes, *European Journal of Operational Research,* V. 122, pp. 151-160.

81. Locke, E.A., Shaw, K.N., Saari, L.M. and Latham, G.P. (1981). Goal setting and task performance, *Psychological Bulletin,* V. 90, pp.125-152.

82. Lovelace, E.A. and Southall, S.D. (1983). Memory for Words in prose and their locations on the Page, *Memory and Cognition,* V. 11, pp. 429-434.

83. Lucas, H.C., (1981). An Experimental Investigation of the Use of Computer-Based Graphics in Decision Making, *Management Science,* V. 27 (7), pp. 757-768.

84. Lumsden. J. *Elementary Statistical Method.* University of Western Australia Press, Nedlands WA, 1974.

85. MacGregor, D., Lichtenstein, S. and Slovic, P. (1988). Structuring knowledge retrieval: an analysis of decomposed quantitative judgements. *Organizational Behavior and Human Decision Processes*, V. 42, pp. 303 - 323.

86. MacGregor, D. Decomposition for judgemental forecasting and estimation, in J. S. Armstrong (ed.) *Principles of Forecasting*, Norwell, MA: Kluwer Academic Publishers, 2001.

87. Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982). The accuracy of extrapolation (time-series) methods: Results of a forecasting competition. *Journal of Forecasting*, V. 1 (2), pp. 111-153.

88. Makridakis, S. and Wheelwright, S.C. *Forecasting Methods for Management, Fifth Edition.* New York: John Wiley & Sons, 1989.

89. Makridakis, S. and Winkler, R.L. (1983). Averages of Forecasts: Some Empirical Results. *Management Science,* V. 29, pp. 987-995.

90. Mandler, J. and Johnson, N. (1977). Remembrance of things parsed: story structure and recall, *Cognitive Psychology*, V. 9, pp. 111-151.

91. Masson, M. E. J. (1982). Cognitive Processes in Skimming Stories. Journal of Experimental Psychology, V. 8, pp. 400-417.

92. McKnight, C., Dillon, A. and Richardson, J. (1990). A comparison of linear and Hypertext formats in information retrieval, in R. McAleese and C. Green (eds.) Hypertext: Sate of the Art (Intellect, Oxford), pp. 10-19.

93. Mentzer, J.T. and Cox, J.E. (1984). Familiarity, Application and Performance of Sales Forecasting Techniques. *Journal of Forecasting.* V. 3 (1), pp. 27-36.

94. Mentzer, J.T. and Kahn K.B. (1995). Forecasting Technique Familiarity, Satisfaction, Usage, and Application. *Journal of Forecasting.* V. 14 (5), pp. 465-476.

95. Meyer, J., Shinar, D. and Leiser, D. (1997). Multiple Factors that Determine Performance with Tables and Graphs. *Human Factors*, V. 39 (2), pp. 268-286.

96. Mills, C. B. and Weldon L. J. (1987). Reading Text from Computer Screens. *ACM Computing Surveys*. V. 19 (4), pp. 229-358.

97. Muter, P. and Maurutto, P. (1991). Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour & Information Technology*, V. 10 (4), pp. 257-266.

98. Mykytyn, P.P. and Green, G.I. (1992). Effects of computer experience and task complexity on attitudes of managers. *Information & Management*, V. 23 (5), pp. 263-278.

99. Newell, A. and Simon, H.A. *Human Problems Solving*. Englewood, Cliffs, NJ: Prentice-Hall, 1972.

100. Nisbett, R. and Wilson, T. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, V. 84, pp. 231-259.

101. Newsted, P.R. (1985). Paper versus online Presentations of Subjective Questionnaires. *International Journal of Man-Machine Studies*, V. 23, pp. 231-247.

102. Oborne, D.J. and Holton, D. (1988). Reading from screen versus paper: there is no difference. *International Journal of Man-Machine Studies*. V. 28 (1), pp. 1-9.

103. O'Connor, M.J., Remus, W., Griggs, K., (1993), Judgemental Forecasting in Times of Change, *International Journal of Forecasting*, V.9, pp. 163-172.

104. O'Connor, M.J., Remus, W., Griggs, K., (1997), Going up-going Down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting*, V.16, pp. 165-176.

105. O'Connor, M. and Lawrence, M.J., Judgemental Forecasting and the Use of Available Information. In G. Wright and P. Goodwin (eds.) *Forecasting with Judgment*. (1998), England: John Wiley & Sons, Ch. 3.

106. Paquette, L. and Kida, T. (1988). The Effect of Decision Strategy and Task Complexity on Decision Performance. *Organizational Behavior and Human Decision Processes*, V. 41, pp. 128 - 142.

107. Payne, J.W. (1976). Task Complexity and Contingent Processing in Human Decision Making: An Information Search and Protocol Analysis. *Organizational Behavior and Human Decision Processes*, V. 16, pp. 366 - 387.

108. Payne, J. (1982). Contingent decision behaviour. *Psychological Bulletin,* V. 92, pp. 382-402.

109. Picking, R. (1997). Reading music from screen vs paper. *Behaviour & Information Technology*, V. 16 (2), pp. 2-78.

110. Plume, T. 1988. Practical Applications of Optical Disk Image Systems in Document Management, in P. A. Yates-Mercer (ed), *Future Trends in Information Science and Technology* (London: Taylor Graham).

111. Radl, G.W. (1980) Experimental Investigations for Optimal Presentation Mode and Colours of Symbols on the CRT Screen, in Grandjean, E and Vigliani, E. (eds) *Ergonomic Aspects of Visual Display Terminals.* London: Taylor & Francis, pp. 127-137.

112. Raiffa, H. (1968). D*ecision Analysis*, Reading, MA: Addison-Wesley.

113. Remus, W.E., (1984). An Empirical Investigation of the Impact of Graphical and Tabular Data Presentations on Decision Making. *Management Science,* V. 30 (5), pp. 533-542.

114. Remus, W.E., (1986). Graduate Students as surrogates for Managers in Experiments on Business Decision Making. *Journal of Business Research,* V. 14 (1), pp. 19-25.

115. Remus, W.E., (1987). A Study of Graphical and Tabular Displays and their Interaction with Environmental Complexity. *Management Science,* V. 33 (9), pp. 1200-1204

116. Sanders, N.R. (1992). Accuracy of judgemental forecasts: a comparison. *Omega*, V. 120, pp. 353 - 364.

117. Sanders, N.R. and Manrodt, K.B., (1994). Forecasting Practices in US Corporations: Survey Results, *Interfaces,* 24, pp. 92 - 100.

118. Sanders, N.R. and Ritzman, L.P., (1992). The need for Contextual and Technical Knowledge in Judgemental Forecasting, *Journal of Behavioural Decision Making,* V.5, pp. 39-52.

119. Schackel, B. (1980). Dialogues and Language – can Computer Ergonomics help? Ergonomics, V. 23 (9), pp. 857-880.

120. Schütte, M. And Jordan, C. (1996). Dependence between performance, strain and strategies for an activity involving information load. *Ergonomics,* V. 39 (7), pp. 903-923.

121. Schroder, H.M., Driver, M.J. And Streufert, S. (1967). *Human Information Processing.* New York: Holt, Reinhart and Winston.

122. Schumacher, G. and Waller, R. (1985). Testing Design Alternatives: A Comparison of procedures in T. Duffy and R. Waller (eds) *Designing Usable Texts.* Academic Press, Orlando, FL, pp. 377 – 403.

123. Shaw, M.E. (1976). *Group Dynamics: The psychology of small group behaviour.* McGraw-Hill. New York, pp. 309-321.

124. Slovic, P., Kunreuther, H. and White, G.F. (1974). Decision processes, rationality, and judgement of natural hazards, *Natural hazards, local, national, and global. White*, G.F (Editor), London: Oxford University Press.

125. Speier, C., Valacich, J.S. and Vessey, I. (1997). The Effects of Task Interruption and Information Presentation on Individual Decision Making. *Proceedings of the Eighteenth International Conference on Information Systems.* Atlanta, Georgia. pp. 21-35.

126. Stevens, A.L. and Rumelhart, D.E. (1975). Errors in Reading: An Analysis using an Augmented Transition Network Model of Grammar, in Norman. D.A., Rumelhart, D.E. and the LNR Research Group (eds.) *Explorations in Cognition.* San Francisco: Freeman and Co., Ch 6.

127. Switchenko, D.M. (1984). Reading from CRT versus Paper: the CRT-disadvantage re-examined. *Proceedings of the 25th Annual Meeting of the Human Factors Society.* Baltimore: HFS.

128. Tinker, M.A. *Legibility of Print*, Iowa State University Press, Ames. I.A. 1963.

129. Tversky, A. and Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, V. 211, pp. 453 - 458

130. Umanath, N.S. (1994). Effect of IS variables on information acquisition modes: An experimental investigation. *Information & Management,* V. 27, pp. 287-301.

131. Vessey, I. (1991). Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature. *Decision Sciences*, V. 22, pp. 1219-240.

132. Vessey, I. (1994). The effect of information presentation on decision making: A cost-benefit analysis. *Information & Management*, V. 27 (2), pp. 103-119.

133. Waern, Y. (1981). On the Relation between Comprehension and Memory of a Complex Text. *Scandinavian Journal of Psychology*. V. 21, pp. 123-132.

134. Waern, Y. and Rollenhagen, C. (1983). Proof Reading text from visual display units (VDUs). *International Journal of Man-Machine Studies*, V. 18, pp. 441-465.

135. Waller, R. (1986). What electronic books will have to be better than. *Information Design Journal*, V. 5 (1), pp. 72-75.

136. Washburne, J.N. (1927). An Experimental Study of Various Graphic, Tabular and Textural Methods of Presenting Quantitative Material. *Journal of Educational Psychology*, V. 18 (6), pp. 361-376.

137. Webby, R.G. (1994). *Graphical Support for the Integration of Event Information into Time-series Forecasting: An Empirical Investigation.* Unpublished Ph.D. dissertation, University of New South Wales.

138. Webby, R. and O'Connor, M. (1996). Judgemental and statistical time series forecasting: a review of the literature. *International Journal of Forecasting*. V. 12, pp. 91-118.

139. Webby, R., O'Connor, M., and Lawrence, M. (2001). Judgemental Time-Series Forecasting using Domain Knowledge, in J. S. Armstrong (ed.) *Principles of Forecasting*, Norwell, MA: Kluwer Academic Publishers.

140. Wilkinson, R.T. and Robinshaw, H.M. (1987). Proof-reading: VDU and paper text compared for speed, accuracy and fatigue. *Behaviour & Information Technology*, V. 6 (2), pp. 125-133.

141. Wilson, J.H. and Daubek, H.G. (1989). Marketing Managers Evaluating Forecasting Models, *Journal of Business Forecasting*, Spring, pp. 19-22.

142. Wilson, E.V. and Addo, T.B.A. (1994). An Investigation of the Relative Presentation Efficiency of Computer-displayed Graphs. V. 26, pp. 105-115.

143. Wood, R.E. (1986). Task Complexity: Definition of the Construct. *Organisational Behavior and Human Decision Processes,* V. 37, pp. 60-82.

144. Wright, P. and Lickorish, A. (1983). Proof reading texts on screen and paper. *Behaviour & Information Technology*, V. 2 (3), pp. 227-235.

145. Zmud, R.W., Blocher, E. and Moffie, R.P. (1983). The impact of Colour Graphic Report Formats on Decision Performance and Learning, *Proceedings of the Fourth International Conference of Information Systems,* pp. 179-193.

10. Appendicies

Appendix 1

Goodhue and Thompson's (1995) Task-Technology Fit Dimensions*

8 Final TTF Factors	16 Original Task-Technology Fit Dimensions**
1. **Quality**	Currency of the Data Right data is maintained Right level of detail
2. **Locatabality**	Locatability Meaning of data is easy to find out
3. **Authorization**	Authorization for access to Data
4. **Compatibility**	Data Compatibility
5. **Ease of use/Training**	Ease of use Training
6. **Production Timeliness**	Production Timeliness
7. **Systems Reliability**	Systems Reliability
8. **Relationship with Users**	IS understanding of business IS interest and dedication Responsiveness Delivering agreed-upon solutions Technical and business planning assistance

* Reproduced from Table 1. pg 222.

** 16 remaining original TTF dimensions after 5 of the original 21 TTF dimensions were dropped as unsuccessfully measured.

AUTHOR	YEAR	TASK	MEDIA	EXPERIMENT	MAJOR FINDINGS	FUTURE RESEARCH
Kak A.	1981	Reading	Screen v's Paper	Very "crude" early experiment, 4 subjects - sample size too small.	**SPEED COMPREHENSION** **SPEED:** Signif. diff. Reading from screen took longer than reading from paper (hard-copy). **COMPREHENSION:** No signif. diff. (Post-task questions). **FATIGUE**: No signif. diff. (Post-task questions).	On Screen, subjects had to "turn-the-page" once. On Paper, the text was all on one page→ maybe this made a diff. Should try to get screen and paper conditions as close as possible. Varied distance **(D)** between screen and subject and found that as **(D)** increased, screen reading rate decreased.
Muter P., Latremouille S.A., Treuniet W.C. & Beam P.	1982	Reading	TV Screen v's Paper	Two 1 hr sessions – CRT & Paper. Text on CRT displayed -ve polarity (light characters on dark b/ground - white text on blue b/ground). Compared texts with variable and fixed letter widths/heights. Characters in CRT text > than on Paper.	**SPEED COMPREHENSION** **SPEED**: Signif. diff. Reading from TV screens was 28.5% slower than from paper. **COMPREHENSION**: No signif. diff. Post-task questions, scores were about the same. **FATIGUE**: No signif. diff. (Post-task questions).	Believe that reading on CRT is the way of the future. Speed diff. , but no diff. in comprehension. Perhaps further study should review skimming for particular information – perhaps speed diff. would not be so great.
Wright P. & Lickorish A.	1983	Proof-reading text P-R	Screen v's Paper Keyboard	Two groups, 16 people each P-R 4 published texts; half the texts (2) were presented as print on paper, the other half (2) on a screen. Each group had to identify errors found in the texts. (1,206 words; 39 errors)	**SPEED ACCURACY** Speed and accuracy were impaired when text was presented on screen **SPEED**: Signif. diff. between media, "People work more slowly when proof-reading text on a VDU screen than when the text is printed on paper". **ACCURACY**: Diff. not Signif.	Authors advised that "The present study was more concerned with the possible cognitive disruption of the reading process that might arise from manipulating text in an unfamiliar electronic medium" - need to further explore some of the cognitive consequences of reading texts on VDU's.
Switchenko D.M	1984	Reading	Screen v's Paper	Exp. carried out over 2 days. Day 1: Subjects read both articles in each pres. mode- CRT & paper. Day 2: Subjects re-read both articles from alternative pres. modes – paper & CRT.	**SPEED** **SPEED**: No signif. difference **COMPREHENSION:** No signif. diff. "percent correct comprehension was also recorded but not of primary interest".	Switchenko concentrated on the "CRT - Disadvantage" hypothesis in relation to SPEED differences between the media.

AUTHOR	YEAR	TASK	MEDIA	EXPERIMENT	MAJOR FINDINGS	FUTURE RESEARCH
Gould J.D. & Grischkowsky N.	1984	Proof-reading text P-R	Screen v's Paper light-pen Exp 1= Paper, +ve polarity Exp 2 = Screen, -ve polarity	24 clerk-typists; divided into two groups of 12 (**Grp1, Grp2**). Pre-test Questionnaire determined level of experience with screen, all participants had little exp. with screen. EXP: Two days, six 45 min work periods, 12 sets of materials, layout same for both. **Exp 1:** P-R paper and circling errors. Day 1, Grp1, Day 2, Grp2 **Exp 2:** P-R screen and mark errors with light pen. Day 1, Grp2, Day 2 Grp1. (1,000 words; 9 errors)	**SPEED, ACCURACY** **SPEED**: - Signif. difference between screen and paper. Speed on paper 20-30% faster than screen. **ACCURACY**: No Signif. diffn. slightly more errors on screen (33%) than hard copy (30%). **FATIGUE**: About the same for both screen and paper. (But has been criticised because didn't check fatigue *within* a session only at the end). **VISION**: not affected by disp. Mode	Believe that this study shows that the same work done on screen and hard-copy can be as accurate and as fulfilling for the worker. It isn't the screen itself which causes fatigue and stress, maybe it is the fact that the workers feel they have little or no control over their work lives. In reference to the SPEED diff., this needs to be looked at, because the diff. in time between circling with pen (paper) and using the light-pen (screen) isn't enough to account for the diff. between the two.
Kruk R.S. & Muter P.	1984	Reading	Screen v's Book	**Exp 1.** To determine which of two factors contributed to slower reading of text from screen; the amount of info. on the page OR the time taken to fill the screen (9secs). 24 subjects read text in all four conditions. ***SCREEN & BOOK 2 conditions*** S - refresh rate, B - # of words **SCREEN:** 39 chars per line, 20 lines per screen for **BOTH.** **S-1:** Delayed, 9.0sec to fill screen. **S-2**: Instant, 0.5sec to fill screen. **BOOK:** Information on Page: **B-1**: More. 60 Chars per line, 40 lines per page. Approx 400 words **B-2:** Less. 39 Chars per line, 20 lines per page. Approx 130 words **Exp 2&3**: Screen only - distance /contrast	**SPEED COMPREHENSION** **SPEED**: **Exp- 1:** Overall→ Was Signif. different, slower on screen (24%) than from book. **(Exp 2 & 3) NO** effect on speed for either increase in distance, or change in contrast ratio. ***Screen:*** No Signif. diff. between S-1 and S-2. ***Book:*** B-1 read significantly faster than B-2 AND B-2 signif. faster than S-2 (Diff. b/w S-2 and B-1 accounted for only 9.5% diff. in speed) **COMPREHESNION**: No signif. difference between conditions. Post-task questions.	"It appears that some, but not all, of the differences in reading speed between computer screen and book conditions are due to the differences in the density of info. displayed per page". (no. of chars per line and no. of lines per page). It appears delay in filling the screen (up to 9 sec) is not a signif. reason for slower reading on screen. No single cause identified. **EXP 2 and 3 - screen only**. Varying the contrast ratio of screen and the distance between the screen and the subject had no effect on reading speed - contradicts Kak (1981). Single spacing on screen produced even slower (10.9%) reading than double spacing. Future Research - other factors - posture, image polarity, character set, resolution, justification (left or full) and familiarity with the medium.

AUTHOR	YEAR	TASK	MEDIA	EXPERIMENT	MAJOR FINDINGS	FUTURE RESEARCH
Askwall, S	1985	Reading	Screen v's Paper	16 students read 4 texts Screen v's paper. 2 exp. sessions each 2 hrs long. Screen size-24 rows x 40 columns Char size 0.5 x 0.5 cm	**SPEED COMPREHENSION** **SPEED:** No signif. difference b/w media. **COMPREHENSION:** No signif. difference. (Measured in terms of accuracy of correct judgements).	Askwall attributes her finding of "NO SPEED difference" to the fact that her texts were fairly short (22 sentences). Generalises results → reading speed not affected by Media for short periods.
Cushman W.H.	1986	Reading	Screen v's Paper also M/fiche	Exp 1: 16 subjects, five 80 min reading sessions. (1) paper; (2) metal screen, +ve image; (3) metal screen, -ve image; (4) HR screen +ve image; (5) HR screen, -ve image. Exp 2: 60 subjects, 80 min reading session. (1) VDT +ve polarity (dark chars. on light background) (2) VDT –ve polarity (light chars. on dark background).	**SPEED COMPREHENSION FATIGUE** **SPEED**: No signif. diff between screen and paper when both +ve polarity. (Speed slightly slower when for screen with -ve polarity). **COMPREHENSION**: No signif. diff. (Post-task questions) **FATIGUE**: Signif. diff. b/ween screen and paper →when *reading* from +ve polarity VDU (dark chars. on light b/ground - to resemble paper). **PAPER better**. No Signif. diff. b/ween screen and paper → when reading from traditional VDUs (-ve polarity).	Refers to Gould and Grischk (84) and says that except for the magnitude of the difference in proof-reading speed reported by G&G, findings as reported by Cushman are similar, → notes that "the discrepancy could be due to task differences (i.e., reading for content versus proof-reading)". Cushman noted a -ve correlation between reading speed and comprehension; i.e., comprehension tended to be higher for slower readers.
Creed A., Dennis I. & Newstead S	1987	Proof-reading text P-R	Screen v's Paper joystick	30 u/g subjects. 3 texts; VDU – joystick to mark errors Paper – vertical presentation, pen to mark errors. (2,100 words; 18 errors)	**ACCURACY** **ACCURACY**: Signif. diff. between media, accuracy poorer on Screen.	Results from media studies should expand → should be viewed in "the wider perspective of the physical and psychological environment in which VDU's are set".

AUTHOR	YEAR	TASK	MEDIA	EXPERIMENT	MAJOR FINDINGS	FUTURE RESEARCH
Gould J.D. , Alfaro L., Barnes V., Finn, R. Grischkowsky N. & Minuto A.	June 1987a	Proof-Reading text P-R	Screen v's Paper Verbal ident. of errors	Experiments attempted to isolate a single-variable Exp 1 examined the influence of a no. of variables; Classifications; *task* (horizontal v's vertical orientation, visual angle), *display* (contrast, scale, polarity, aspect ratio) and *personal* (e.g. experience or familiarity with the media, age). **Exp 4:** Briefly compared proof-reading and comprehension. Reading for Comp. was faster than proof-reading. (1,000 words; 9 errors)	**SPEED:** Signif diff. between the media, Speed - Paper better **ACCURACY:** NO Signif. diff. Mainly interested in SPEED. No one variable studied in isolation can explain the diff., probably due to a combination of variables, may be reliant on the image quality of the characters. Concluded that however, "... the basic finding is robust - people do read more slowly from CRT displays" (p. 269).	Need to concentrate research on searching for conditions in which people can read as fast from VDU displays as from paper; rather than looking for an explanation of the reading-speed difference. **Exp 4:** Found reading speed for Comprehension was faster than proof-reading under all 3 conditions; paper better.
Wilkinson R.T. & Robinshaw H. M.	1987	Proof-reading text P-R	Screen v's Paper Verbal ident. of errors	Proof-reading experiments to review differences - specifically in relation to speed, accuracy, and fatigue 5,000 words over 5x 1 hr sessions; 9 errors/1,000 words, verbal identification of errors.	**SPEED & ACCURACY & FATIGUE.** SIGNIF. DIFF for all three, Screen signif. poorer for all three. (Findings in sharp contrast to those of Gould & Grischkowsky 1984). Results showed that paper performed better than the screen in terms of all three conditions, increased accuracy and speed and resistance to fatigue.	Need to conduct further experimentation to review possible speed/accuracy trade-off. Need to also conduct future research to determine if present results are simply due to subjects being less "familiar" with the screen environment. Question whether "the difference observed would disappear with practice and familiarity?"
Gould J.D. , Alfaro L., Finn R. , Haupt B. & Minuto A.	Oct 1987b	Proof - Reading text P-R	Screen V's Paper Verbal ident. of errors	Exp. used High-resolution, anti-aliased display on which fonts resembled those on paper. i.e., they created an "enhanced" screen condition. (1,000 words; 8 errors)	**SPEED - NO SIGNIF. DIFF.** **ACCURACY: NO SIGNIF. DIFF** Speed: - No signif. diff b/w screen and paper when quality of screen is "high" and matches paper condition as much as possible.	Experiments have been based upon proof-reading, need to verify whether results would hold for reading for comprehension.

AUTHOR	YEAR	TASK	MEDIA	EXPERIMENT	MAJOR FINDINGS	FUTURE RESEARCH
Oborne D.J. & Holton D.	1988	Reading	Screen v's Paper	Created a "degraded" paper condition that resembled the screen; "... the paper was positioned on a stand in front of the subjects at the same reading distance inclination and conditions as the screen" (p. 7). Tried to control as many variables as possible within the experimental situation. Such as differences in *presentation format* - horizontal for paper, vertical for screen; *reading distances* and *polarity or contrast* - paper generally has dark characters on a white background, screen may not.	**SPEED & COMPREHENSION** **NO SIGNIF. DIFFERENCE b/w media for either.** Results showed that when all variables remain constant, there was **no difference** in reading **speed or comprehension** between the two treatment conditions. In relation to image polarity, their findings supported those of Cushman (1986), whose study reported no significant difference in reading speed when image polarity was the same between the two treatment conditions.	In view of the fact that the presentation of information in electronic form is becoming more pervasive, instead of merely investigating differences in performance between the two media, perhaps what we should be looking at is whether ".. individuals perform better using screen or paper when the material is presented in the most optimum format for that medium"? (p. 8).
Creed A., Dennis I. & Newstead S.	1988	Proof-reading P-R	**VDU ONLY** Difft. VDU display formats and contrast.	** Reviewed difference in speed/accuracy across ONE MEDIA only ➔ VDU only. Used different Display Segmentation and polarity to review accuracy and speed of proof-reading task. **Found SPEED / ACCURACY trade-off.**	Displaying text in different formats affected SPEED and ACCURACY (trade-off). An improvement in accuracy when displaying text one p/graph at a time was accompanied by a reduction in speed. Displaying text by sentences further increased accuracy but again, speed decreased even more.	Findings indicate that a large number of factors affect performance with VDU's, much more work needs to be done so that we can beneficially exploit the flexibility of the VDU by using different display formats for different tasks

AUTHOR	YEAR	TASK	MEDIA	EXPERIMENT	MAJOR FINDINGS	FUTURE RESEARCH
Black A.	1990	Graphic design	Screen v's Paper	**Study 1:** 29 1st year students planned the same text twice, working once on paper and once on screen. They then filled out a questionnaire about their level of satisfaction with each medium. **Study 2:** 18 2nd year students were given a drafting exercise, broken down into individual tasks. They were asked to identify which medium they had predominantly used for each task.	**SATISFACTION WITH MEDIA** Interface constraints can lead to less satisfactory design solutions on screen than on paper. Screen may "inhibit" creative process. DTP software and tools enabled draft layouts to be prepared, which on screen, appeared "finished", perhaps this "finished" appearance may have an adverse effect on the creative process and stops it prematurely.	Focussed on Limitations of Desktop Publishing (DTP) systems as design tools and tries to suggest ways in which screen-based designing could be improved both by user education and by developing software to meet user's needs. Black acknowledges the contribution of technological developments in this area, but calls for greater levels of interaction between user and machine so as to create an environment more congenial to visible planning.
Muter P. Maurutto P.	1991	Reading & Skimmi ng	Screen V's Book	H0: there is no difference in speed or comprehension between screen and normal book for Reading or Skimming. **EXP 1:** 24 Subjects each read 6 short stories; 3 (book), 3 (screen). Comprehension test given after each story, 10 short answer Q's. , randomly presented for 10sec each Media (screen / book) = within subject variable. Reading & Skimming = between subject variable. **EXP 2:** 18 Subjects read 6 stories, shorter than EXP 1, 3 conditions, ***CRT-A***= typical PC screen format single line spacing - 1980's style. ***CRT-B***= Improved screen, as for EXP1, double spacing etc. →90's ***BOOK***= as for EXP1.	**SPEED & COMPREHENSION** **EXP 1 & 2: READING:** No Signif. diff in Speed or Comprehension. **EXP 1: SKIMMING:** Signif. diff. in Speed, **Book / Paper Faster** Comprehension - Signif. diff. **Screen Better** **EXP 1 & 2: READING:** No signif. diff. in SPEED. Comprehension was better on screen, but not signif. **EXP 1: SKIMMING:** SPEED signif. diff. between book and screen. 41% slower on screen **BUT** Comprehension was better on screen, maybe there was a speed-accuracy trade-off.	The increase in reading speed in comparison to earlier studies may be attributable to the quality of the screen and the clarity of characters. Comprehension was higher in the screen condition, but did not reach significance "The paperless office may be imminent after all". "Reading from computer screens that are readily available in 1991 can be equivalent in Speed and Comprehension to reading from a book".

AUTHOR	YEAR	TASK	MEDIA	EXPERIMENT	MAJOR FINDINGS	FUTURE RESEARCH
Dillon A.	1992	Review of Litera-ture. Mainly Proof-reading	Screen v's Paper	A critical review of the empirical literature. "Although reading from screens may be slower and occasionally less accurate than reading from paper, no ONE variable is likely to be responsible".	Observed differences between the media classified as either outcome (what the reader gets from the text) or process (how the reader uses a text) measures. **Outcome measures** - Speed, Accuracy, Fatigue, Comprehension, Preference. **Process measures** - Eye movements, Manipulation, Navigation. "Lack of scientific rigour has reduced the value of many of the studies".	"Single variable explanations are insufficient to capture the range of issues involved in reading from screens". Need to look at more cognitively demanding tasks other than proof-reading; future research needs to be grounded in theory. "As with many variables, the task being performed is likely to be a deciding factor".
Picking R.	1997	Proof - Reading Music P-R	Screen v's Paper	H0: There is no difference b/w presentation styles - media. 19 Subjects P-R 3 pieces of music from screen v's paper. 3 intentional errors (pitch errors) incl. in each piece of music. 2 Screen conditions. S-1 = static presentation (Frozen) S-2 = animated score tracking. (Tracker, Stepper, Jumper - P. 73) Considered reader preference, author acknowledges possible problems with validity. Subjective preferences recorded post-task.	**ACCURACY, Style Preference** **ACCURACY:** No Signif. diff. b/w presentation styles (media). BUT, S-1, *Frozen* performed worst of all. Paper proof-reading score was 10% better than *Frozen* (but worse than all three S-2 conditions) **Style Preference:** Using subjective data, 5 level Likert scale, signif. diff b/w presentation styles, P, S-1 and S-2. (Again S-1, *Frozen,* performed worst of all)	"The results do suggest that the computer display itself may cause readers some problems in the ergonomic sense, but that visual aids may cancel out that problem" (p. 77). "This study has shown that there is a strong indication that animation can offer very powerful reading assistance, especially to less able readers of music" (p. 77). Study provided evidence to justify the extra effort of incorporating animated support tools in this domain - reading of music from screens.

PROOF-READING TEXT

AUTHORS	YEAR	TASK	MAJOR FINDINGS
Wright and Lickorish	1983	proof-reading	SPEED: Signif. difference in Speed, paper better. ACCURACY: Difference in Accuracy, but not signif.
Gould and Grischkowsky	1984	proof-reading	**SPEED:** Signif. difference, Speed was 20-30% faster on paper. **ACCURACY:** No signif. difference b/w media. **FATIGUE:** No signif. difference b/w media. **VISION:** No effect, Display mode did not affect vision.
Creed, Dennis and Newstead	1987	proof-reading	**SPEED:** No signif. difference in relation to Speed, but there was a trend towards slower performance on screen. **ACCURACY:** Signif. difference, Accuracy was significantly worse on screen.
Gould, Alfaro, Barnes, Finn, Grischkowsky and Minuto	1987a	proof-reading	**SPEED:** Signif diff. paper better.
Wilkinson and Robinshaw	1987	proof-reading	**SPEED:** Signif. difference, paper better. **ACCURACY:** Signif. difference, paper better. **FATIGUE:** Signif. difference, paper better. Paper performed better than screen on three counts: speed of reading, detection of proof-reading errors and resistance to fatigue.
Gould, Alfaro, Finn, Haupt and Minuto	1987b	proof-reading	**SPEED:** No Signif. difference between media. **ACCURACY:** No Signif. difference between media.
Creed, Dennis and Newstead	1988	proof-reading	**SPEED / ACCURACY** trade-off *WITHIN* the Media. Media = Screen only. Used difft. Display Segmentation and polarity to review accuracy and speed of proof-reading task. Displaying less text on screen increased Accuracy but decreased Speed.
Picking	1997	proof-reading music	**ACCURACY:** No Signif. difference between presentation styles: Paper, S-1 (Frozen), and S-2 (Tracker, Stepper, Jumper). BUT, S-1(Frozen) performed worst of all. Paper proof-reading score was 10% better than Frozen (but worse than all three S-2 conditions).

READING FOR COMPREHENSION

AUTHORS	YEAR	TASK	MAJOR FINDINGS
Kak	1981	Reading	**SPEED:** Signif. difference, paper better. Reading from screen took longer than hard copy. **COMPREHENSION:** No signif. difference. **FATIGUE:** No signif. difference.
Muter, Latremouille, Treuniet and Beam	1982	Reading	**SPEED:** Signif. difference, paper better. Reading from screen was 28.5% slower than reading from a book. **COMPREHENSION:** No signif. difference. **FATIGUE:** No signif. difference.
Switchenko	1984	Reading	**SPEED:** No signif. difference b/w media. **COMPREHENSION:** No signif. difference b/w media.
Kruk and Muter	1984	Reading	**SPEED:** Signif. difference, reading from screen significantly slower than reading from a book. **COMPREHENSION:** No signif. difference.
Askwall	1985	Reading	**SPEED:** No signif. difference b/w media. **COMPREHENSION:** No signif. difference. (Measured in terms of accuracy of correct judgements).
Cushman	1986	Reading	**SPEED:** No signif. difference b/w media when both +ve Polarity (dark chars. on light b/ground). Speed slightly slower for Screen with -ve Polarity. **COMPREHENSION:** No signif. difference. **FATIGUE:** +ve Polarity VDU: Signif. difference b/w media. - ve Polarity VDU: No signif. difference b/w media.
Oborne and Holton	1988	Reading	**SPEED:** No signif. difference b/w media when both +ve Polarity (dark chars. on light b/ground). **COMPREHENSION:** No signif. difference.
Muter and Maurutto	1991	Reading & Skimming	**READING:** **SPEED:** No signif. difference b/w media. **COMPREHENSION:** No signif. difference. **SKIMMING:** **SPEED:** Signif. difference b/w media. Book / Paper Faster **COMPREHENSION:** Signif. difference. Screen Better

Appendix 3

Media Studies - Range of Tasks used to Measure Comprehension

The following list, categorised by Carroll (1972), represents a range of tasks which have been used to measure comprehension (Waern and Rollenhagen p.454) in various Media Studies as presented in Table 2.6.6.2.1.

A. Subjective reports assessing comprehension or comprehensibility.

B. Reports about truth or identity with other information as, for instance, verification against pictures, own prior knowledge, or logical truth.

C. Non-verbal reactions to orders.

D. Completing missing parts of a message, as in a common close-procedure, or sentence completion.

E. Answers to questions based upon the message.

F. Recognition of the message or part of it in a later presentation.

G. Reproduction of the message as a whole or in parts as, for instance, in rote, cued or free recall or in the translations to some other language or symbolism.

Appendix 4

Media Studies – Environmental Attributes*

1. **Scale:**

This refers to the dimensions of the material/s presented to task-doers. Including the dimensions of the actual "page" of material/s presented as well as the size of the characters and graphical displays.

For the purpose of this Thesis, the scale of materials presented to task-doers in the paper-based treatment condition was identical to that presented to task-doers in the screen-based treatment condition.

2. **Orientation:**

This refers to the orientation of the material/s presented to task-doers.

For the purpose of this Thesis, the natural orientation for the paper-based medium was been deemed to be horizontal, with the task-doer generally looking down onto the medium. Whilst the general orientation for the screen-based medium was been deemed to be an "upright position" with the task-doer generally facing the medium.

3. **Aspect Ratio:**

This generally refers to the relationship of width to height of the presented material/s. Typically, paper sizes are higher than they are wider (portrait layout), whilst the opposite can be said for typical screen-based displays (landscape layout).

For the purpose of this Thesis, all materials were presented in landscape layout under both treatment conditions (paper-based and screen-based).

4. **Image Polarity:**

This term refers to the relationship between the colour/s of the characters and the background of the medium. A display in which dark characters appear on a light background (e.g. black on white) is referred to as positive image polarity. Conversely, a display in which light characters appear on a dark background (e.g. white on black) is referred to as negative image polarity.

For the purpose of this Thesis, all materials were presented using positive image polarity under both treatment conditions (paper-based and screen-based).

5. **Display Characteristics:**

This refers to the issues related to fonts such as character size, line spacing and character spacing.

For the purpose of this Thesis, materials presented in the paper-based treatment condition were identical to those presented to task-doers in the screen-based treatment condition. The materials were basically a "screen dump" or "screen capture" with all display characteristics being identical.

6. **Contrast:**

This term also refers to the relationship between the colour/s of the characters and the background of the medium.

For the purpose of this Thesis, all materials – characters and graphics – were presented on a white background under both treatment conditions.

7. **Posture of the Reader/Task-doer:**

This refers to the general posture and orientation of the task-doer including the distance from the material/s.

For the purpose of this Thesis, task-doers under both the paper-based and screen-based treatment conditions were seated in a typical classroom-type situation. Task-doers were seated on a chair and the material/s were presented to them were locate don a table in front of them.

In general, posture of the task-doer could be said to be equivalent under each of the treatment conditions. However, as noted by Gould *et al.*, 1987a, "...people do sit farther away from a CRT screen than from paper..." p. 281. Given the brevity of the experimental task to be undertaken in each of the four studies, this issue was not deemed to be significant.

* Adapted from Dillon (1992) pp. 1309 – 1315

Appendix 5.1

General Instructions – Tool 1 – Screen- based

School of Information Systems, Technology and Management
Judgemental Forecasting Experiment

1. You have been chosen to participate in a software supported Judgemental Forecasting Research Project.

   This Research Project is being conducted as part of my Thesis.

   The results are purely for the basis of experimentation. This exercise is not related to any assessment procedure, there are no right or wrong answers.

2. You will be presented with 4 Time Series graphs.

3. You are required to plot the values for the next 12 periods on the graph presented.

   Where you Plot the points is where you think the actual value for that period within the Time Series will be. In other words, it is merely your prediction of what the graph will do over the next 12 periods.

   This is purely a judgemental exercise so each one of you will have a different opinion about where each point will or should be.

4. I have a demonstration System to show you.

   Instructor Notes Only:

   Demonstrate the PC system on the laptop. Start the system by typing in:

   C: \ JFS \ DEMO

   Use the slider, mouse and keyboard and buttons to select and accept values for each of the 12 points - go back and edit some values to show the editing feature/s. When finished, press the **PREDICTIONS COMPLETED** button .

5. You may change your period values if you wish by re-selecting and accepting each of the period values you wish to edit.

6. Once you are satisfied with each selection, please accept the values by pressing the **PREDICTIONS COMPLETED** button and moving on to the next graph.

Appendix 5.1.1

Handout – Tool 1 – Screen- based

Handout provided to all Subjects in the screen-supported treatment condition:

Tool 1.

- *Media* – screen based;
- *Technology* – slider, mouse and keyboard;
- *Task complexity* – unstructured, more complex.

School of Information Systems, Technology and Management
Judgemental Forecasting Experiment

You will be able to practice using the Tool with your own demonstration system. To activate this demonstration system please run the following:

C: \ JFS \ DEMO

Once you're happy with the demonstration system, click on the Predictions Completed button, this will return you to windows.

You may now activate the Judgemental Forecasting System. Run the following:

C: \ JFS \ JFS

Thank You for your time:

Appendix 5.2

General Instructions – Tool 3 – Screen- based

School of Information Systems, Technology and Management
Judgemental Forecasting Experiment

1. You have been chosen to participate in a software supported Judgemental Forecasting Research Project.

   This Research Project is being conducted as part of my Thesis.

   The results are purely for the basis of experimentation. This exercise is not related to any assessment procedure, there are no right or wrong answers.

2. You will be presented with 4 Time Series graphs. These graphs may show additional information such as a trend line or seasonality.

3. You are required to plot the values for the next 12 periods on the graph presented.

   Where you Plot the points is where you think the actual value for that period within the Time Series will be. In other words, it is merely your prediction of what the graph will do over the next 12 periods.

   This is purely a judgemental exercise so each one of you will have a different opinion about where each point will or should be.

4. I have a demonstration System to show you.

   Instructor Notes Only:
   Demonstrate the system, start the system by typing

   DM website INFS5992
   Notices → JFS DEMO

   Use the mouse and buttons to select and accept values for each of the 12 points - go back and edit some values to show the editing feature/s.

5. You may change your period values if you wish by re-selecting and accepting each of the period values you wish to edit.

6. Once you are satisfied with each selection, please accept the values by pressing the **PREDICTIONS COMPLETED** button and moving on to the next graph.

Appendix 5.2.1

Handout – Tool 3 – Screen- based

Handout provided to all Subjects in the screen-supported treatment condition;

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (1) unstructured, more complex
  – (2) structured, less complex.

School of Information Systems, Technology and Management Judgemental Forecasting Experiment

1. You will be able to practice using the tool with your own demonstration system. To activate this demonstration system, please use the following instructions:

   Access the Data Management Website INFS5992;

   Go to **NOTICES**:

   Click on the JFS Demo ICON

   Once you're happy with the demonstration system, click on the Predictions Completed button, this will return you to the website.

2. You may now activate the Judgemental Forecasting System. Please use the following instructions:

   Access the Data Management Website INFS5992;

   Go to **LECTURES**:

   Click on the JFS Experiment ICON

   Once you have completed the exercise please put up your hand, I will come around to collect the data. Please do not logout until the data has been collected.

Thank You very much for your time:

Appendix 5.3

General Instructions – Tool 2 – Paper-based

School of Information Systems, Technology and Management
Judgemental Forecasting Experiment

1. You have been chosen to participate in a paper-based Judgemental Forecasting Research Project.

   This Research Project is being conducted as part of my Thesis.

   The results are purely for the basis of experimentation. This exercise is not related to any assessment procedure, there are no right or wrong answers.

2. You will be presented with 4 Time Series graphs.

3. You are required to plot the values for the next 12 Periods for each graph presented.

   Where you Plot the points is where you think the actual value for that period within the Time Series will be. In other words, it is merely your prediction of what the graph will do over the next 12 periods.

   This is purely a judgemental exercise so each one of you will have a different opinion about where each point will or should be.

4. I have a demonstration System to show you.

   Instructor Notes Only:

   Put up overhead and plot the next 12 point with a marker.

   DO NOT JOIN THE POINTS AND PLOT THE GRAPH.

5. You may change your selected period values if you wish, erase previous values and plot new ones.

6. Once you are satisfied with each selection, please move on to the next graph and **DO NOT** make any further changes to this graph.

Appendix 6A.1

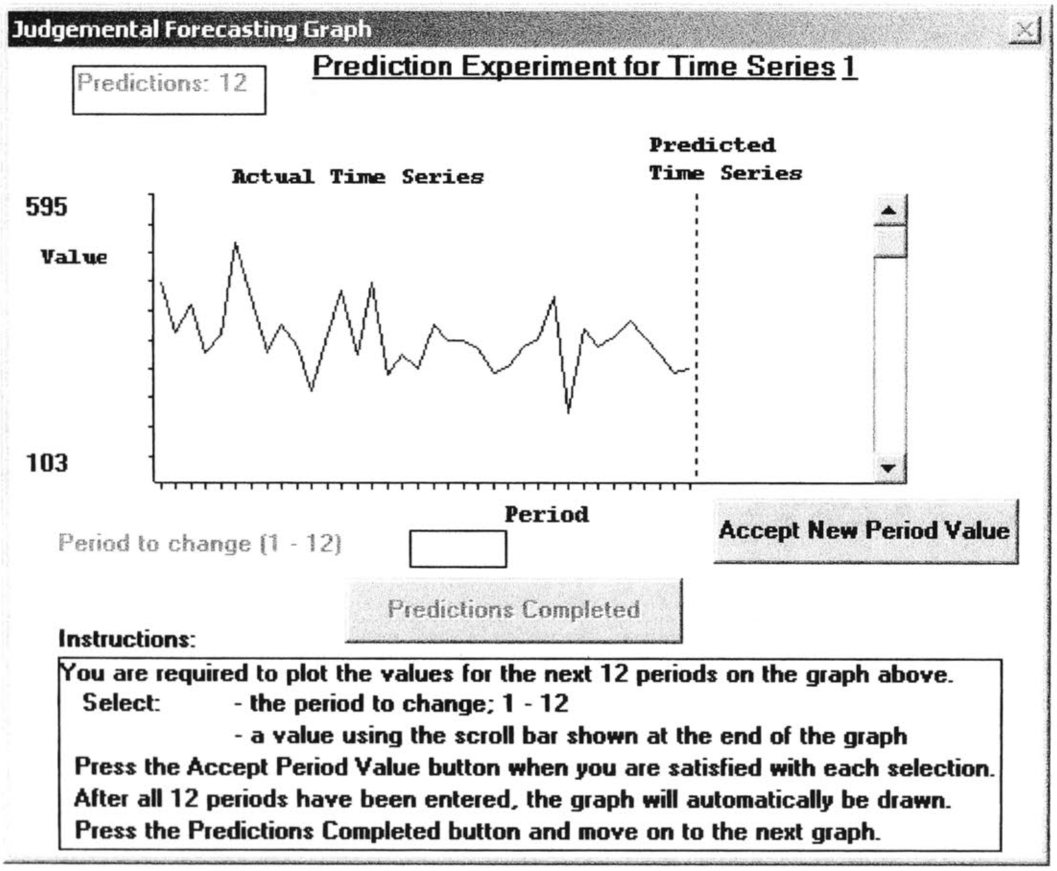
Tool 1 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 1 screen-based treatment condition.

Tool 1:

- *Media* – screen based;
- *Technology* – slider, mouse and keyboard;
- *Task complexity* – unstructured, more complex.

Time Series B17



Appendix 6A.2

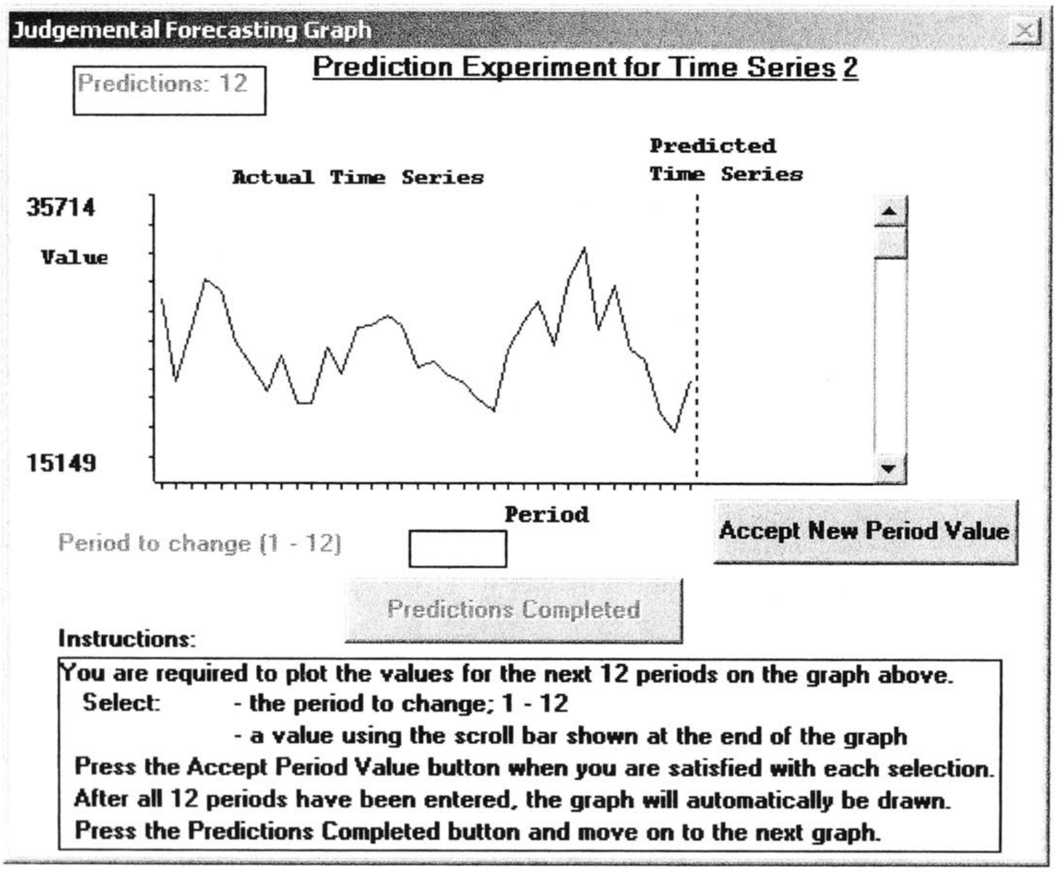
Tool 1 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 1 screen-based treatment condition.

Tool 1:

- *Media* – screen based;
- *Technology* – slider, mouse and keyboard;
- *Task complexity* – unstructured, more complex.

Time Series B26



Appendix 6A.3

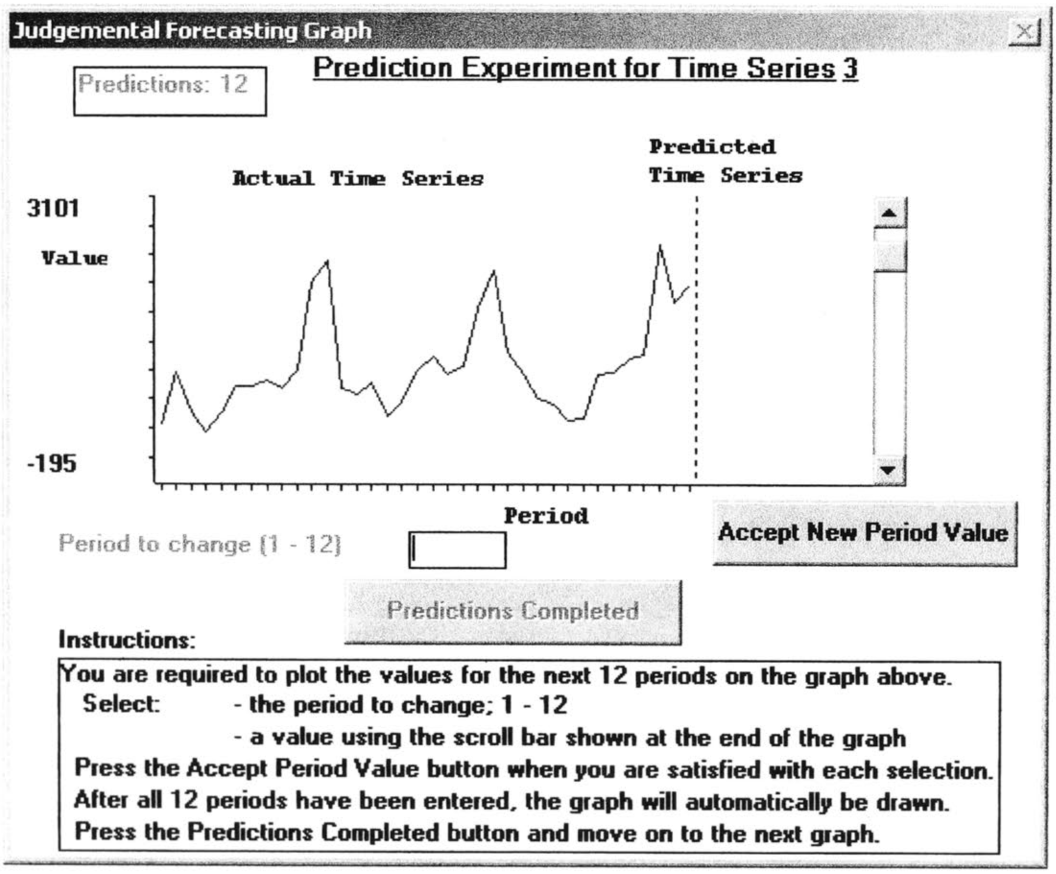
Tool 1 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 1 screen-based treatment condition.

Tool 1:

- *Media* – screen based;
- *Technology* – slider, mouse and keyboard;
- *Task complexity* – unstructured, more complex.

Time Series B8



Appendix 6A.4

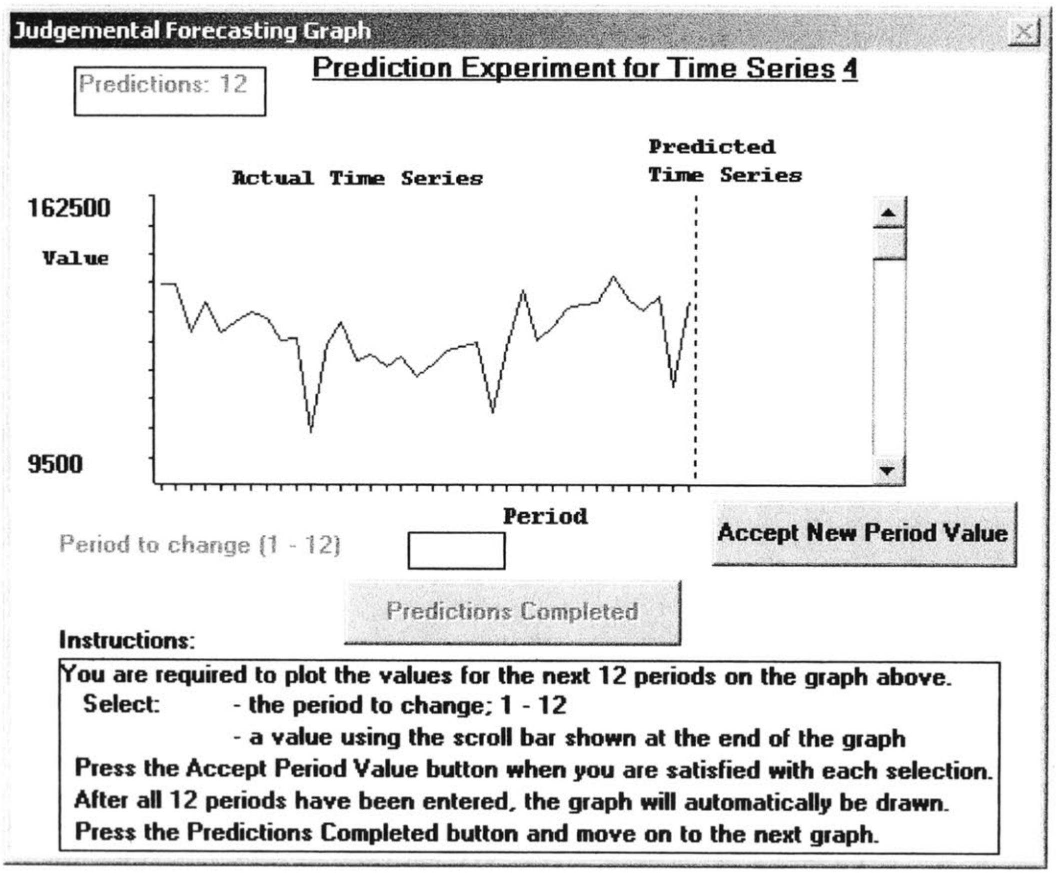
Tool 1 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 1 screen-based treatment condition.

Tool 1:

- *Media* – screen based;
- *Technology* – slider, mouse and keyboard;
- *Task complexity* – unstructured, more complex.

Time Series M6



Appendix 6A.5

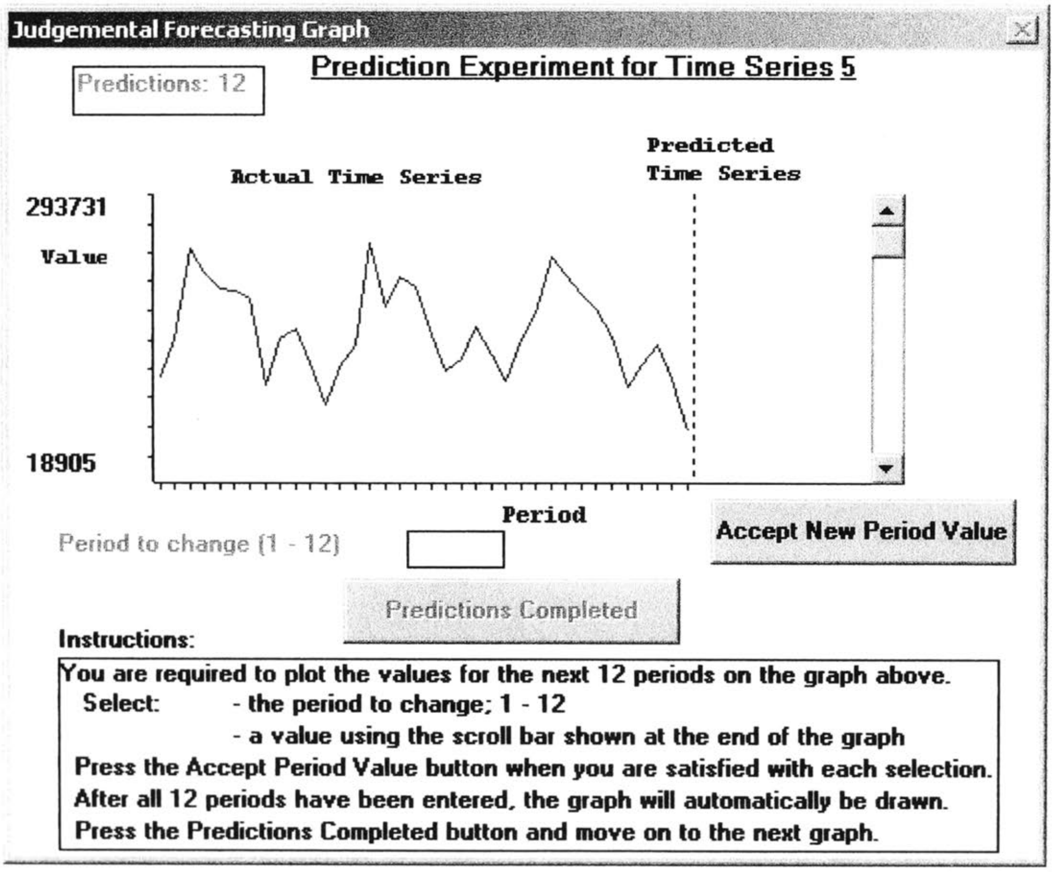
Tool 1 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 1 screen-based treatment condition.

Tool 1:

- *Media* – screen based;
- *Technology* – slider, mouse and keyboard;
- *Task complexity* – unstructured, more complex.

Time Series M7



Appendix 6B.1

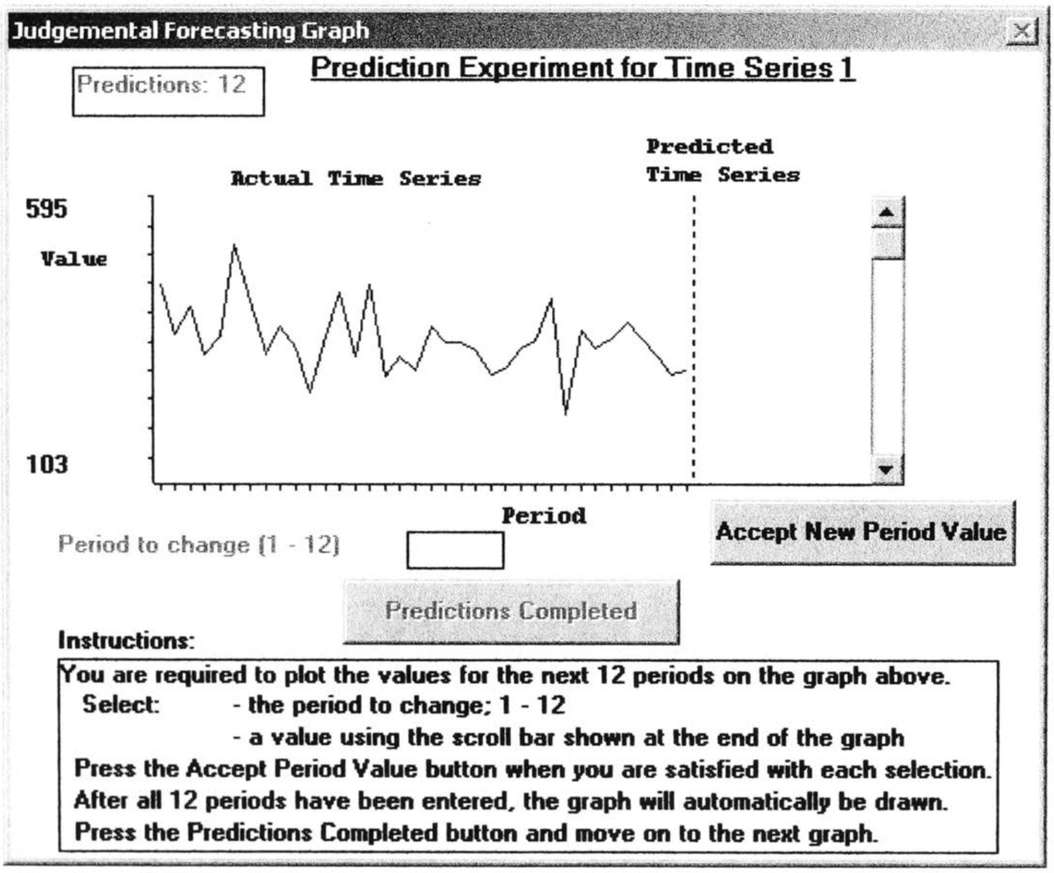
Tool 2 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 2 paper-based treatment condition.

Tool 2:

- *Media* – paper based;
- *Technology* – pencil and eraser;
- *Task complexity* – unstructured, more complex.

Time Series B17



Appendix 6B.2

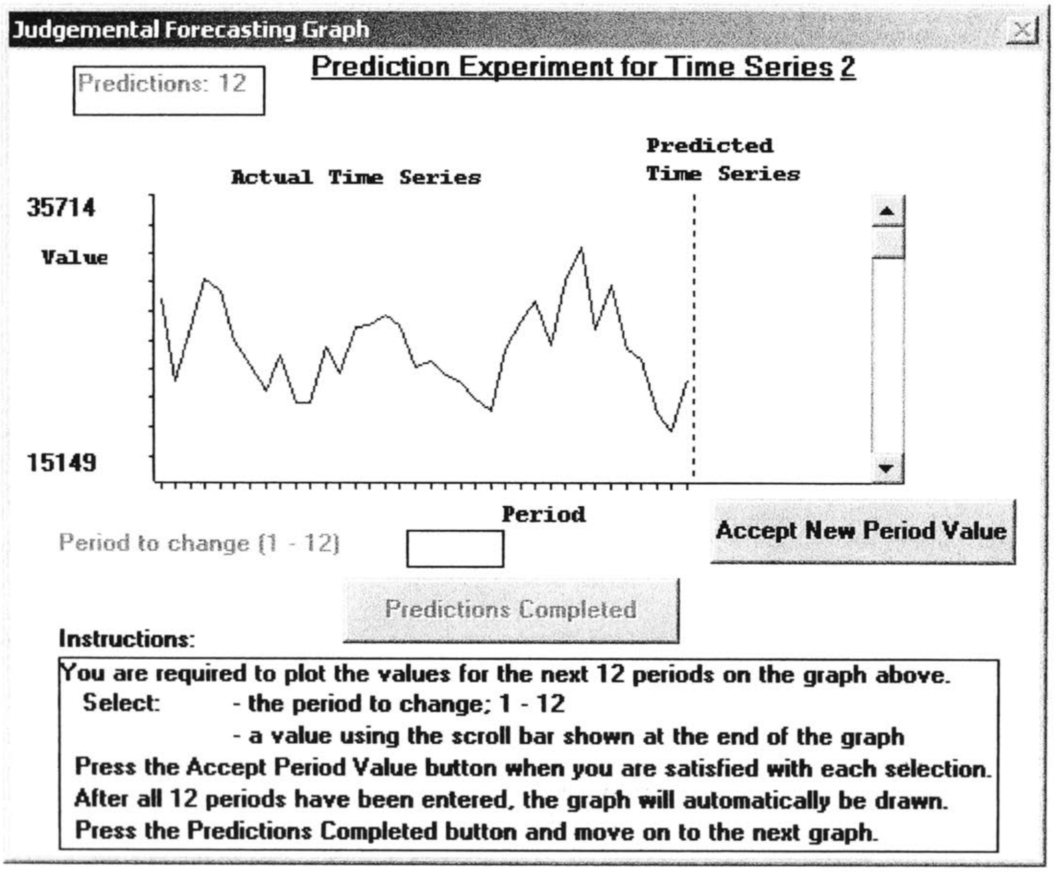
Tool 2 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 2 paper-based treatment condition.

Tool 2:

- *Media* – paper based;
- *Technology* – pencil and eraser;
- *Task complexity* – unstructured, more complex.

Time Series B26



Appendix 6B.3

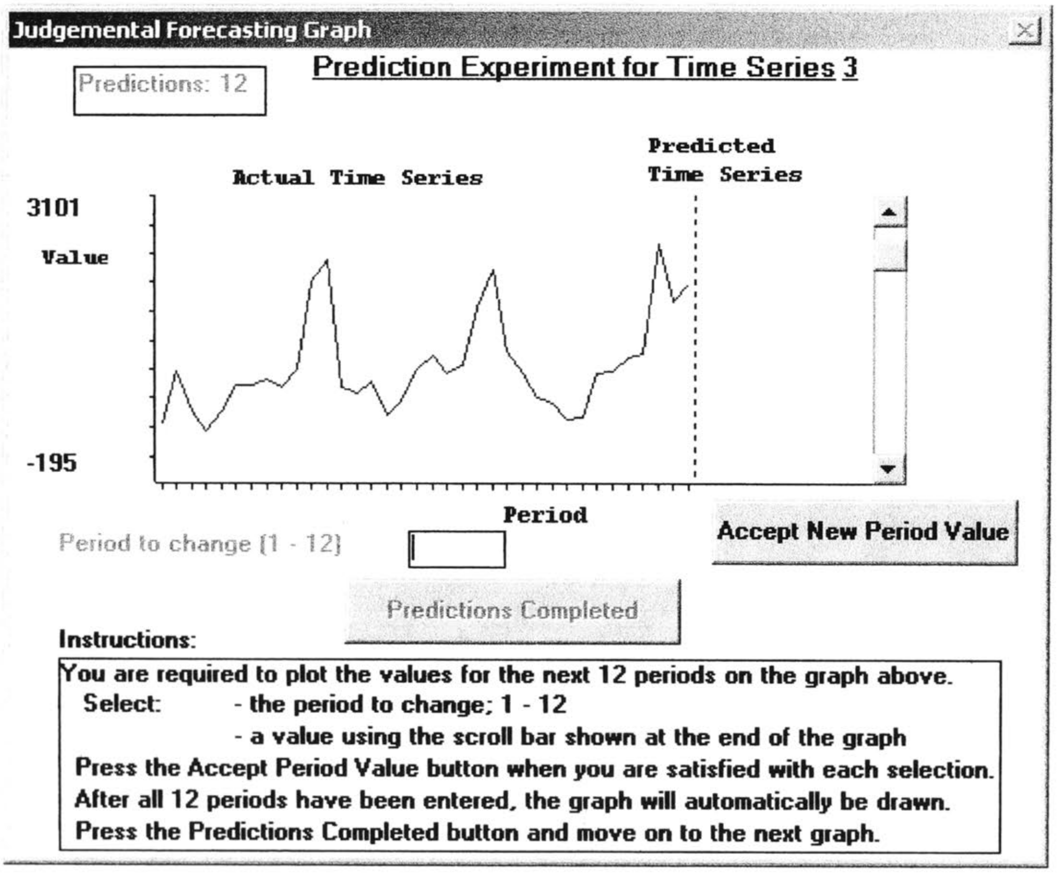
Tool 2 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 2 paper-based treatment condition.

Tool 2:

- *Media* – paper based;
- *Technology* – pencil and eraser;
- *Task complexity* – unstructured, more complex.

Time Series B8



Appendix 6B.4

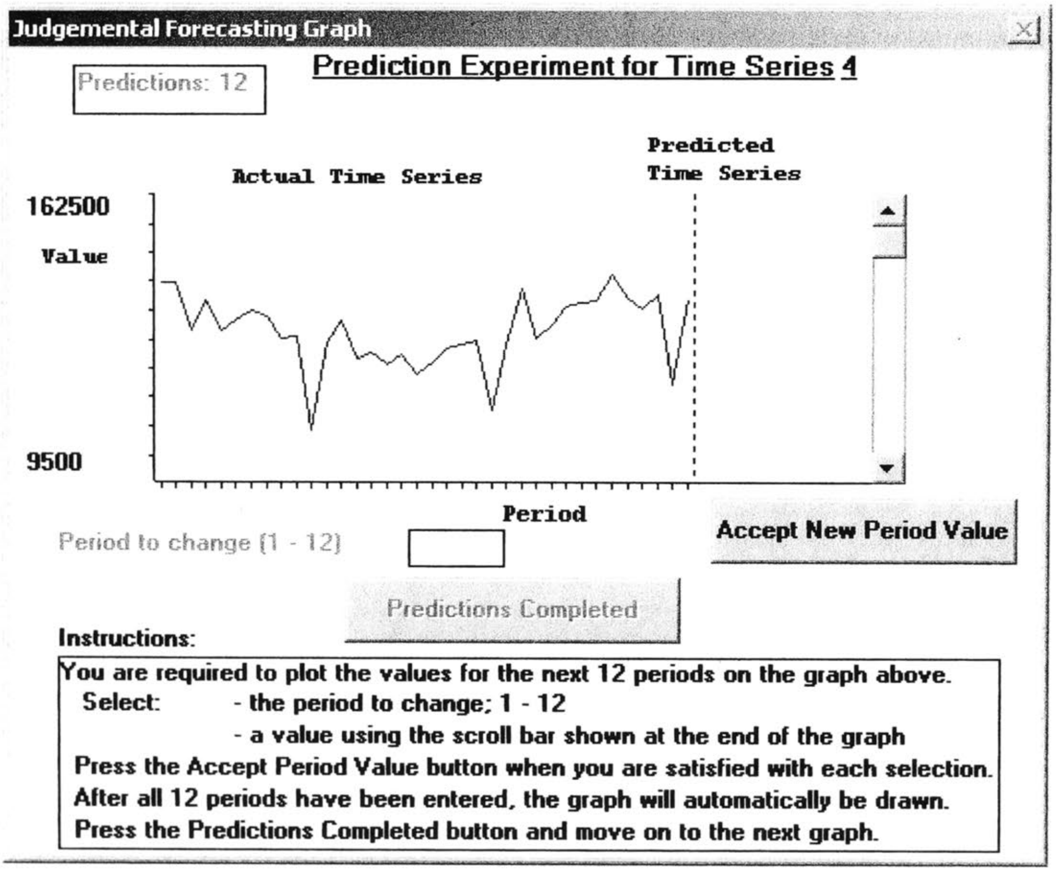
Tool 2 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 2 paper-based treatment condition.

Tool 2:

- *Media* – paper based;
- *Technology* – pencil and eraser;
- *Task complexity* – unstructured, more complex.

Time Series M6



Appendix 6B.5

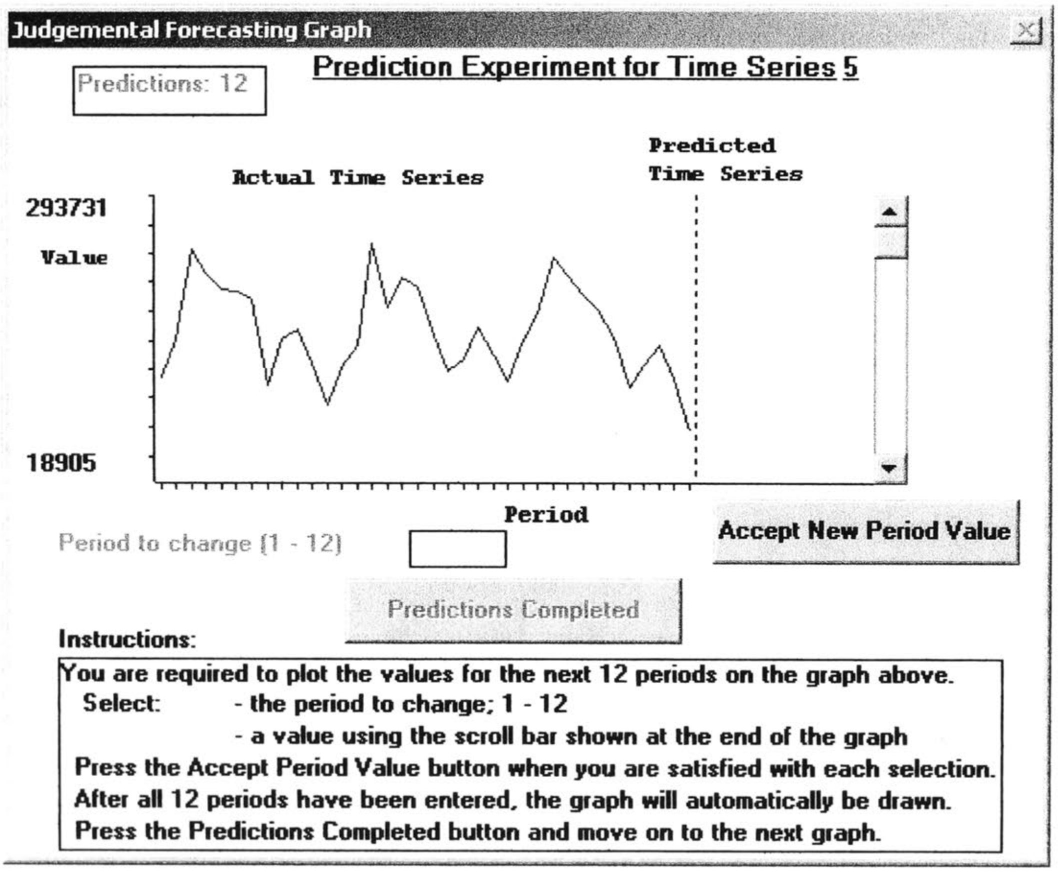
Tool 2 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 2 paper-based treatment condition.

Tool 2:

- *Media* – paper based;
- *Technology* – pencil and eraser;
- *Task complexity* – unstructured, more complex..

Time Series M7



Appendix 6C.1

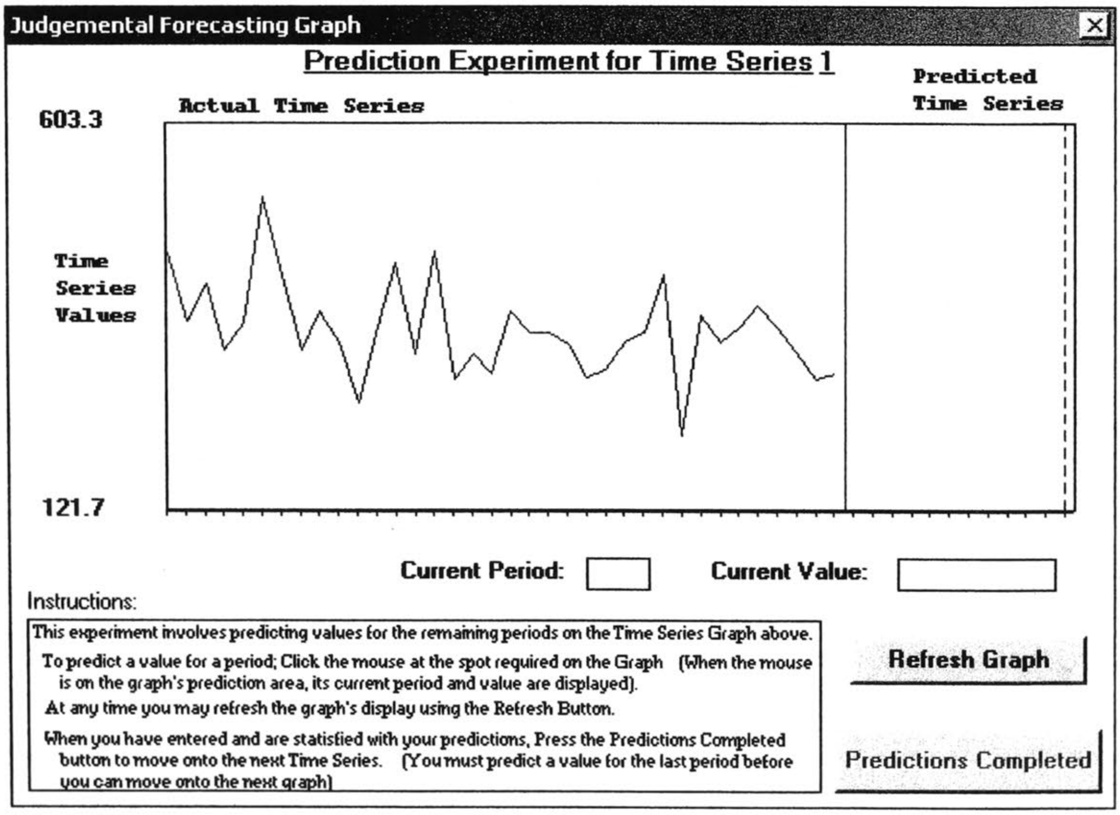
Tool 3 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 3 screen-based treatment condition.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (1) unstructured, more complex

Time Series B17



Appendix 6C.2

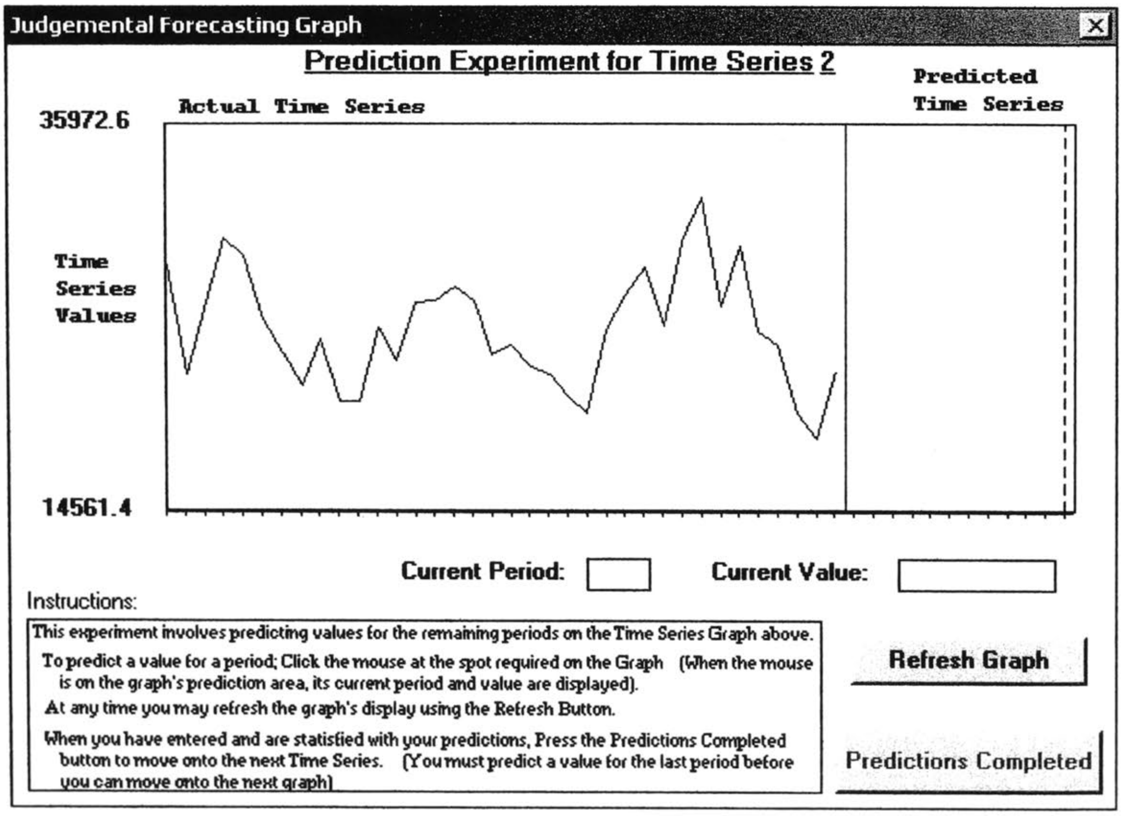
Tool 3 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 3 screen-based treatment condition.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (1) unstructured, more complex

Time Series B26



Appendix 6C.3

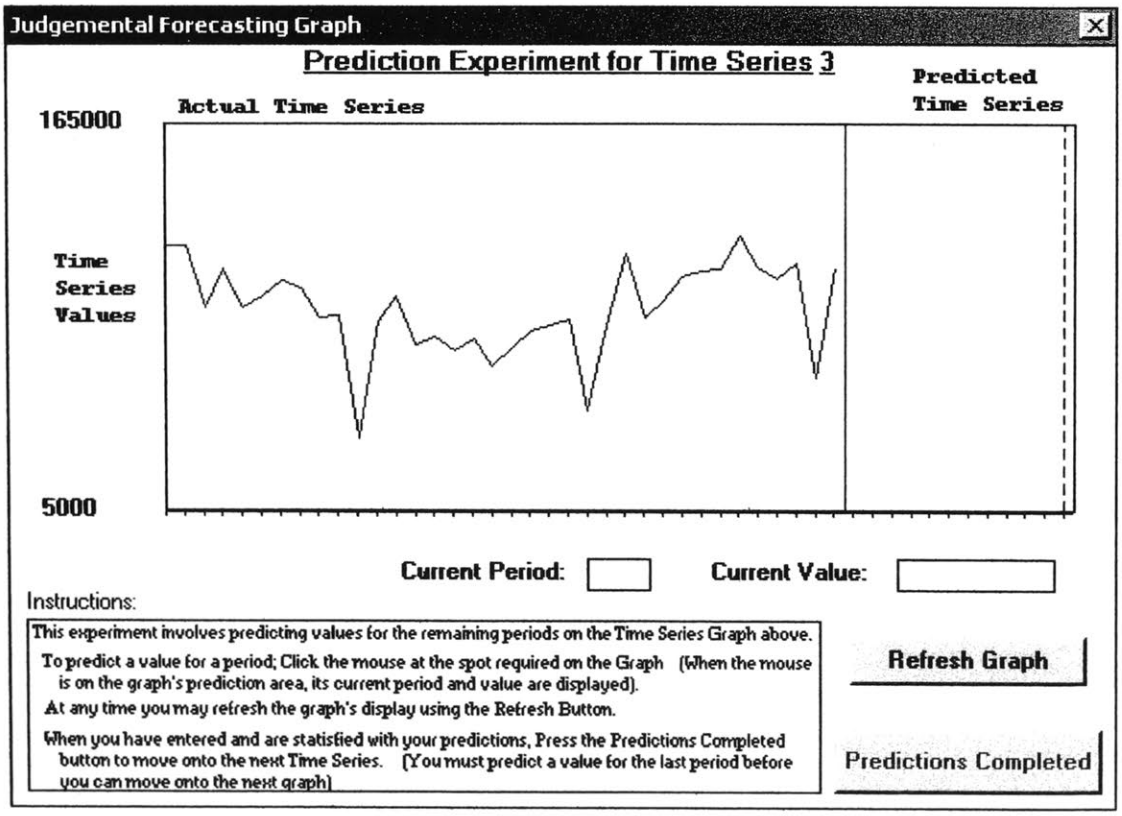
Tool 3 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 3 screen-based treatment condition.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (1) unstructured, more complex

Time Series M6



Appendix 6C.4

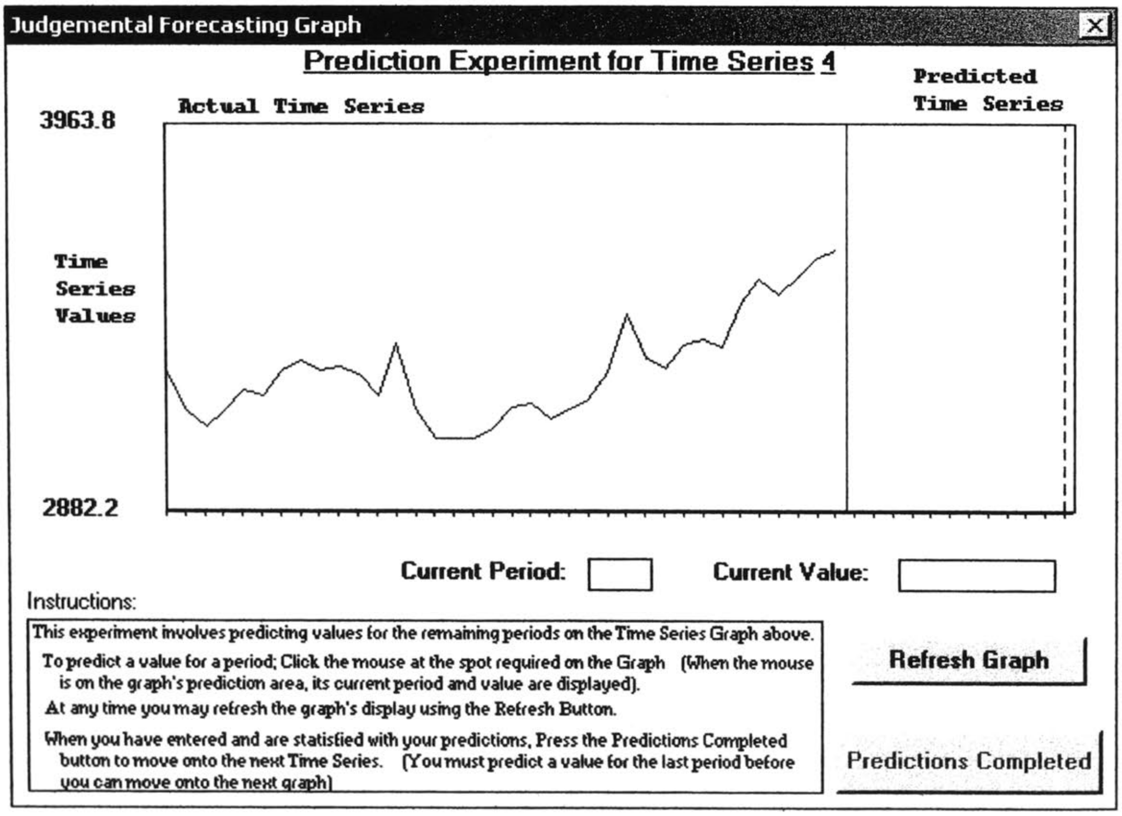
Tool 3 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 3 screen-based treatment condition.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (1) unstructured, more complex

Time Series F3



Appendix 6D.1

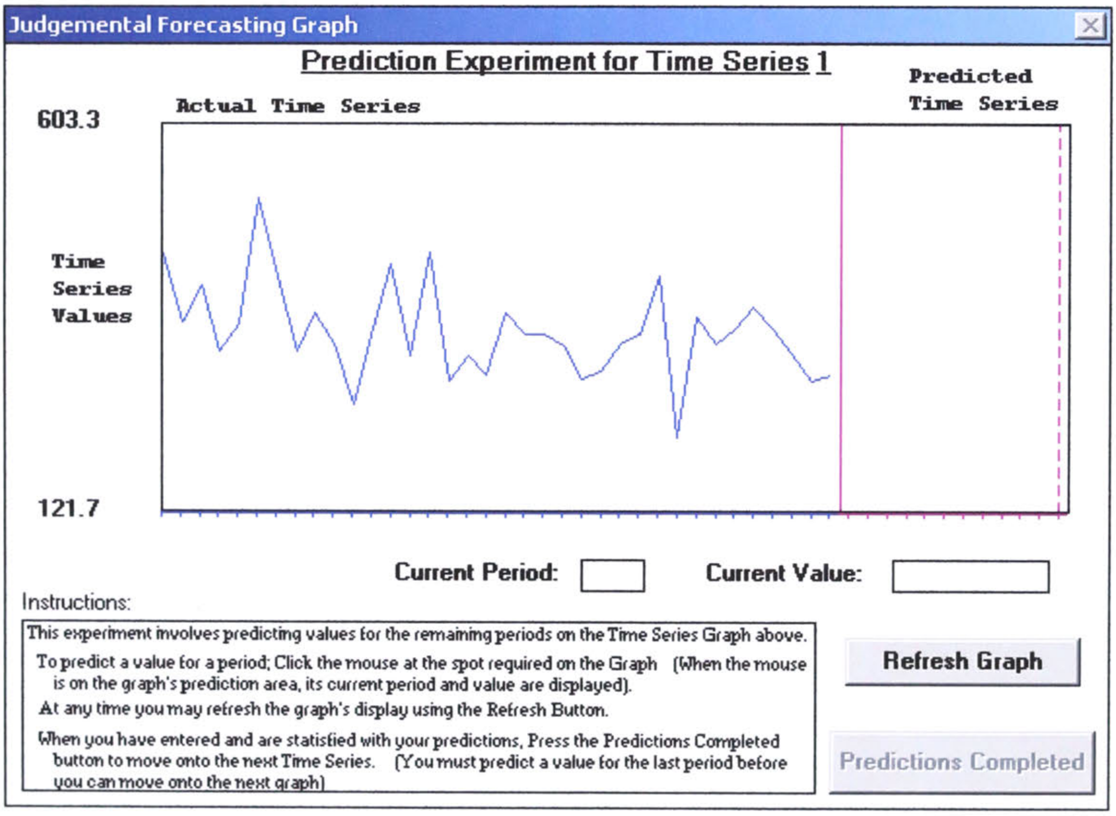
Tool 3 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 3 screen-based treatment condition.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (2) structured, less complex

Time Series B17 (no trend and no seasonality)



Appendix 6D.2

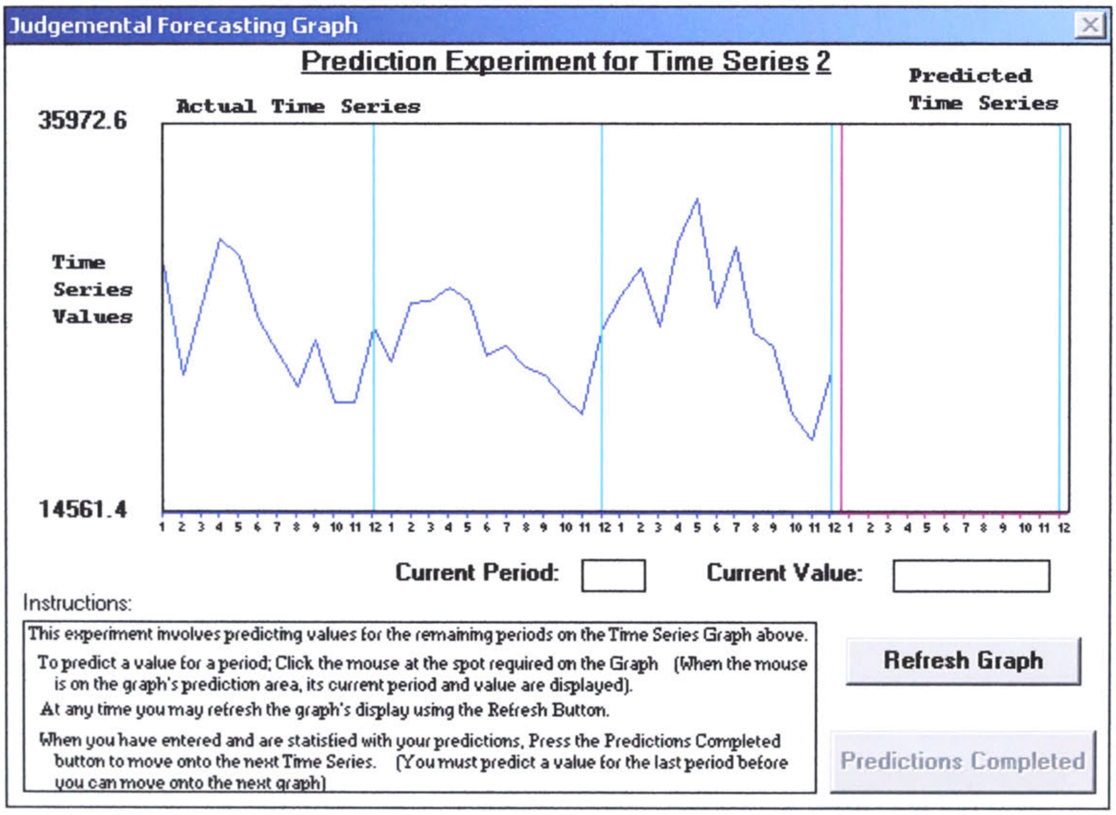
Tool 3 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 3 screen-based treatment condition.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (2) structured, less complex

Time Series B26 (no trend but seasonality)



Appendix 6D.3

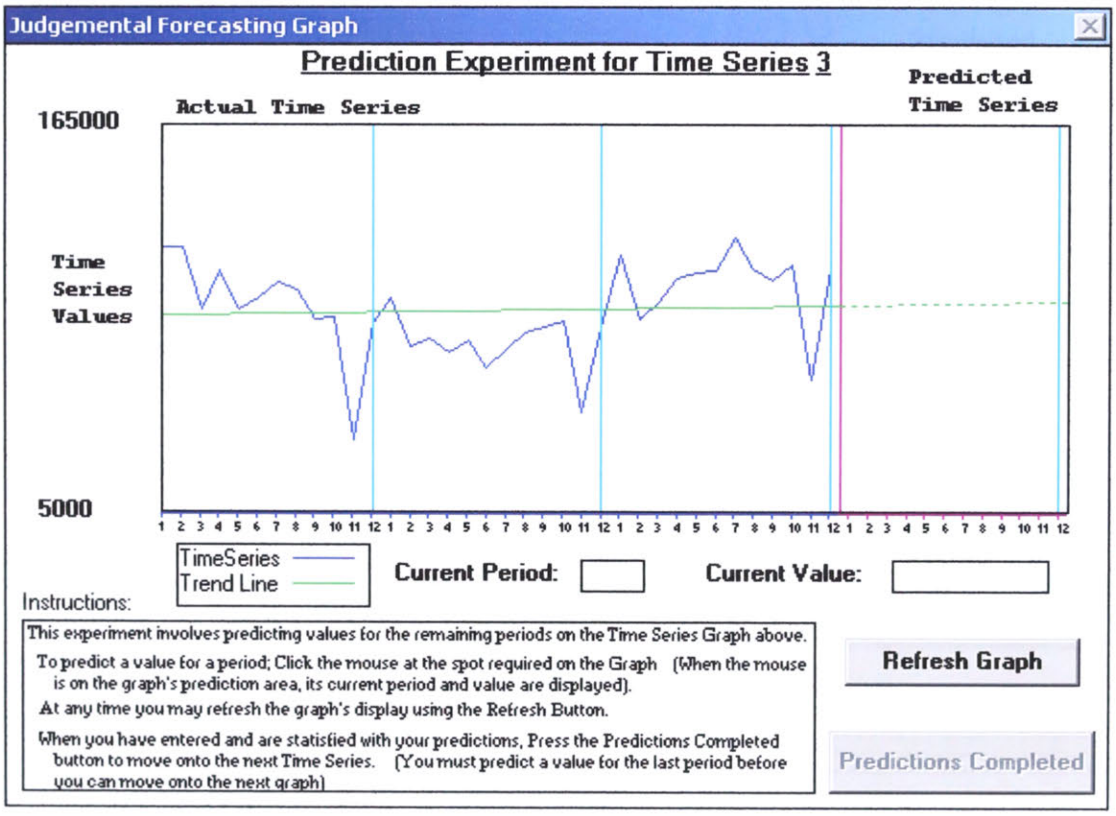
Tool 3 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 3 screen-based treatment condition.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (2) structured, less complex

Time Series M6 (trend and seasonality)



Appendix 6D.4

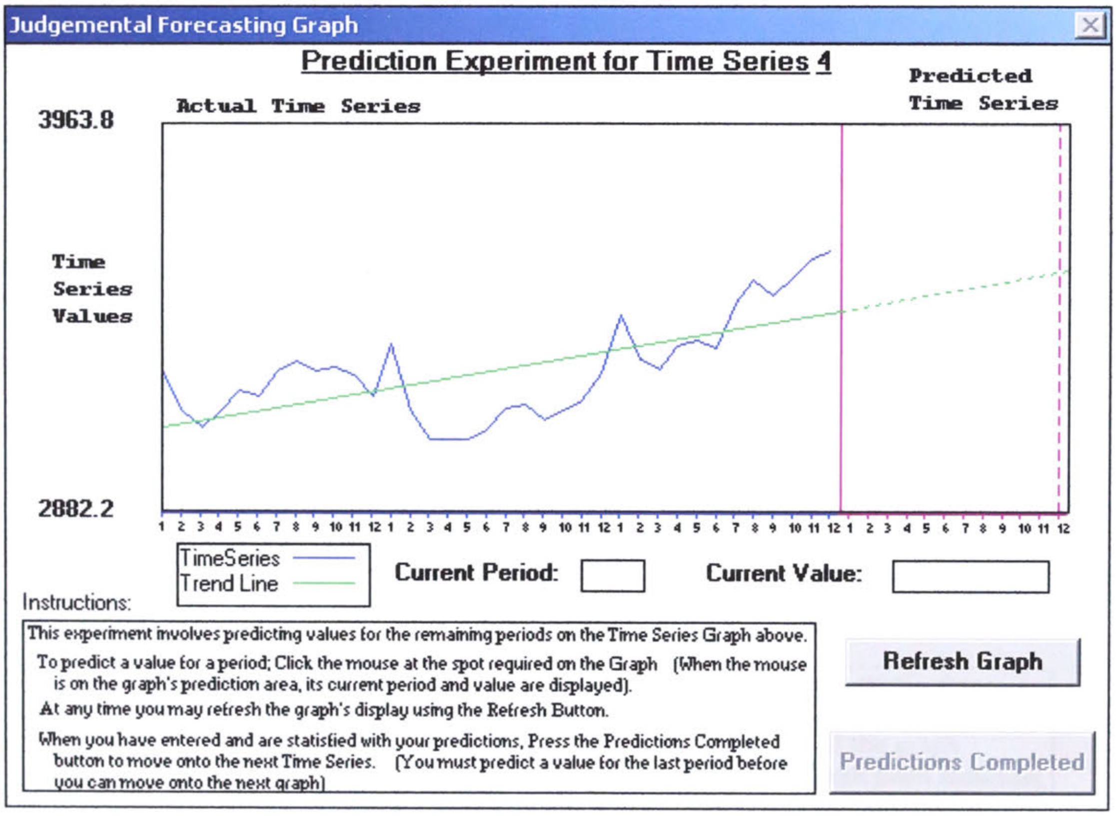
Tool 3 – Graphical Presentation of Time Series Images

Images as presented to subjects in the Tool 3 screen-based treatment condition.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (2) structured, less complex

Time Series F3 (trend but no seasonality)



Appendix 6E.1

Demonstration System - Graphical Presentation of Time Series Images

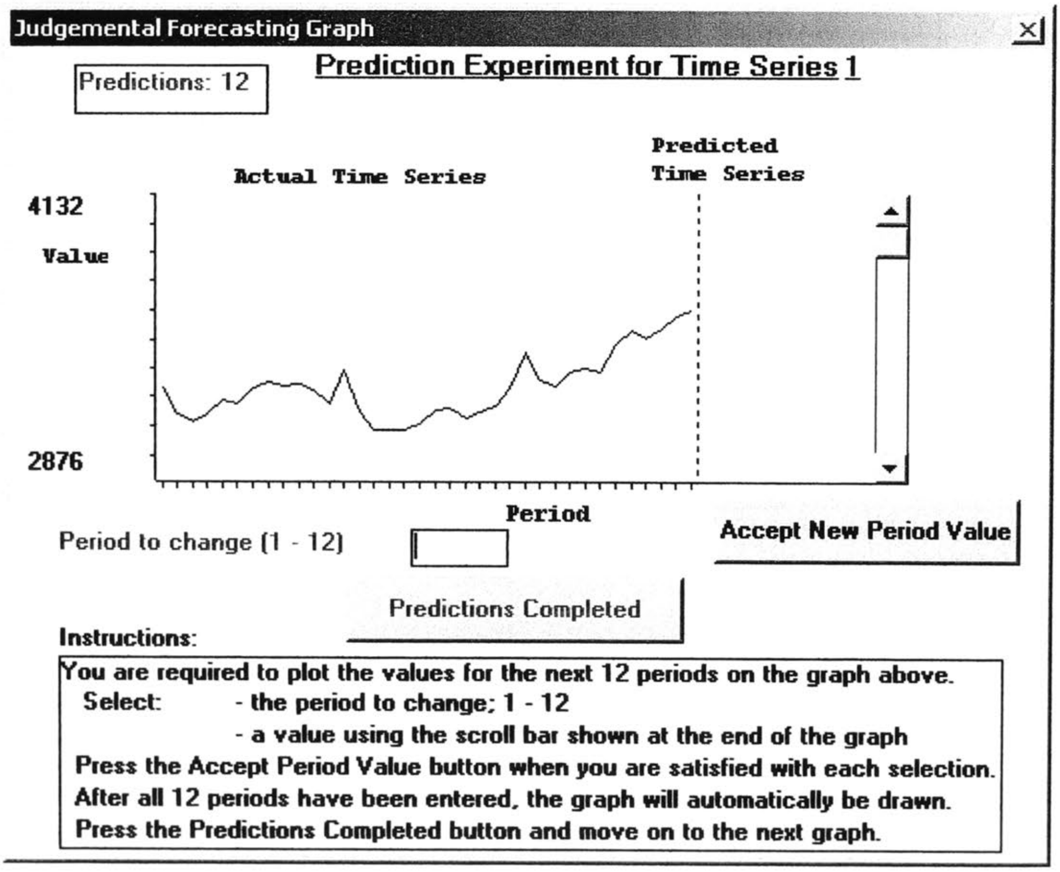
Demonstration System images as presented to subjects in Experiment 1 – screen-based (Tool 1) and paper-based (Tool 2).

Tool 1:

- *Media* – screen based;
- *Technology* – slider, mouse and keyboard;
- *Task complexity* – unstructured, more complex.

Tool 2:

- *Media* – paper based;
- *Technology* – pencil and eraser;
- *Task complexity* – unstructured, more complex.



Appendix 6E.2

Demonstration System - Graphical Presentation of Time Series Images

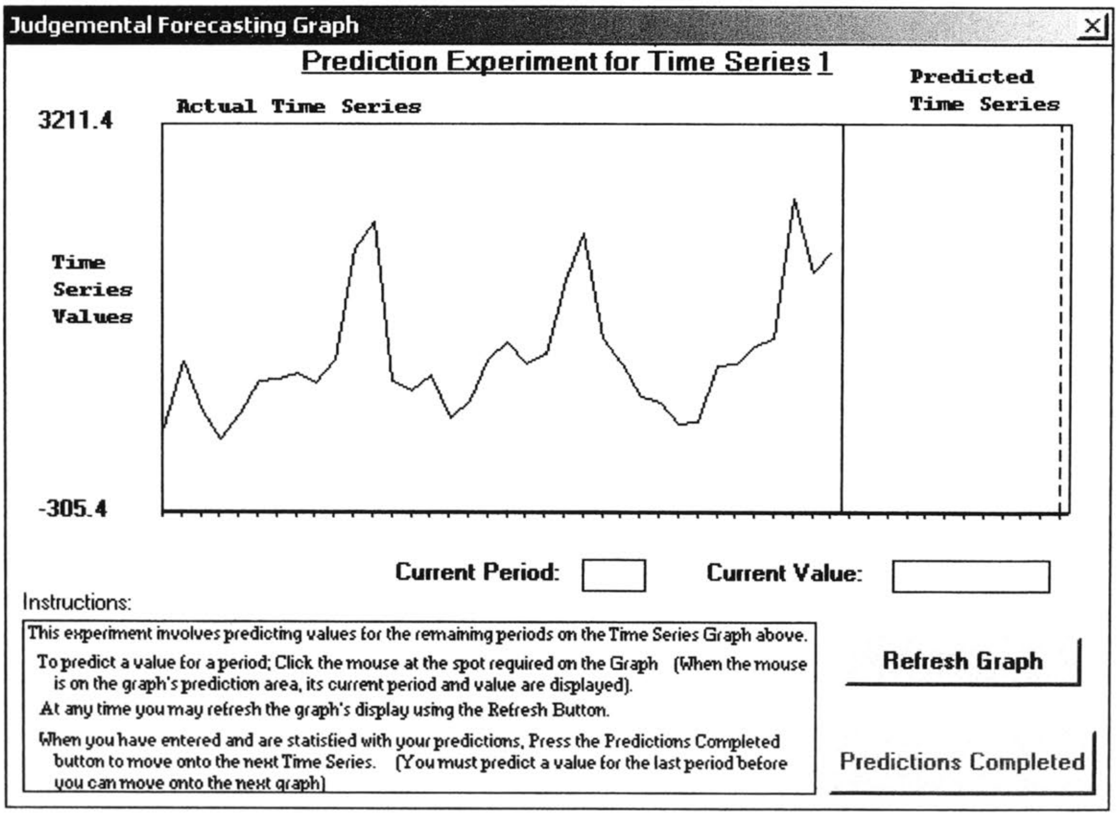
Demonstration System images as presented to subjects in Experiment 2 – screen-based (Tool 1) and paper-based (Tool 2).

Tool 1:

- *Media* – screen based;
- *Technology* – slider, mouse and keyboard;
- *Task complexity* – unstructured, more complex.

Tool 2:

- *Media* – paper based;
- *Technology* – pencil and eraser;
- *Task complexity* – unstructured, more complex.



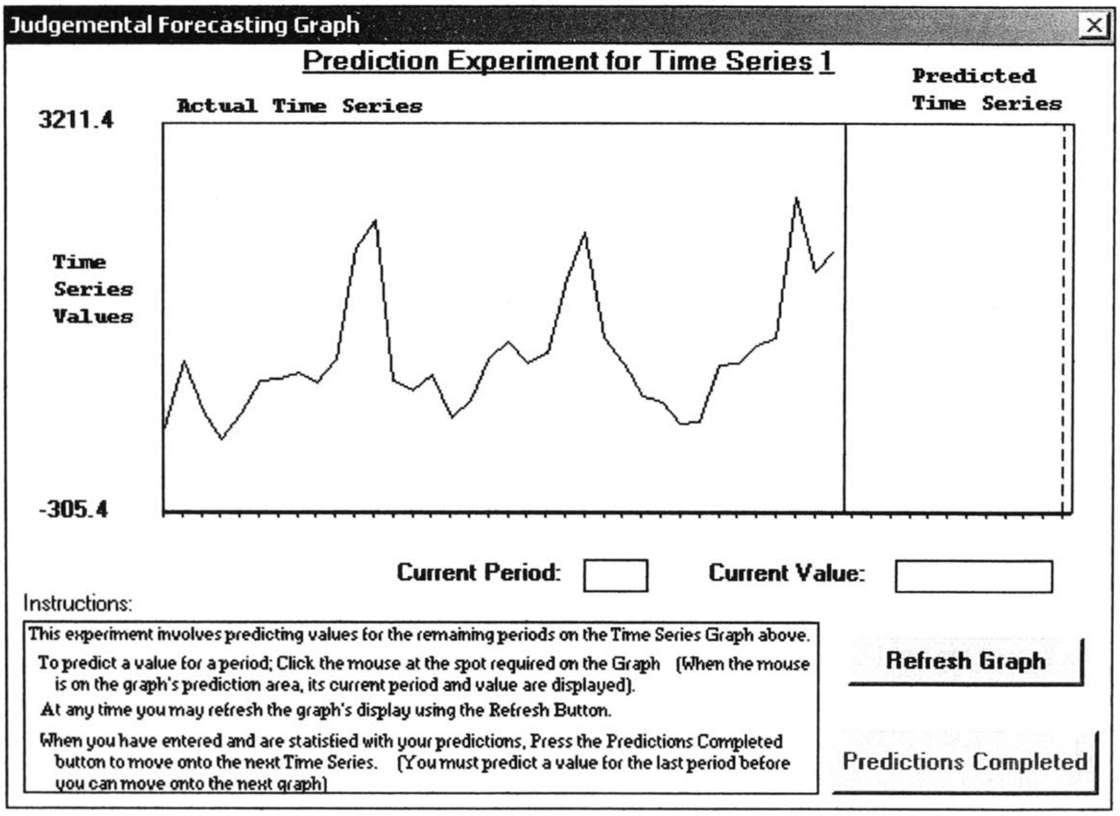
Appendix 6E.3

Demonstration System - Graphical Presentation of Time Series Images

Demonstration System images as presented to subjects in Experiment 3 – screen-based treatment condition using Tool 3.

Tool 3:

- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (1) unstructured, more complex



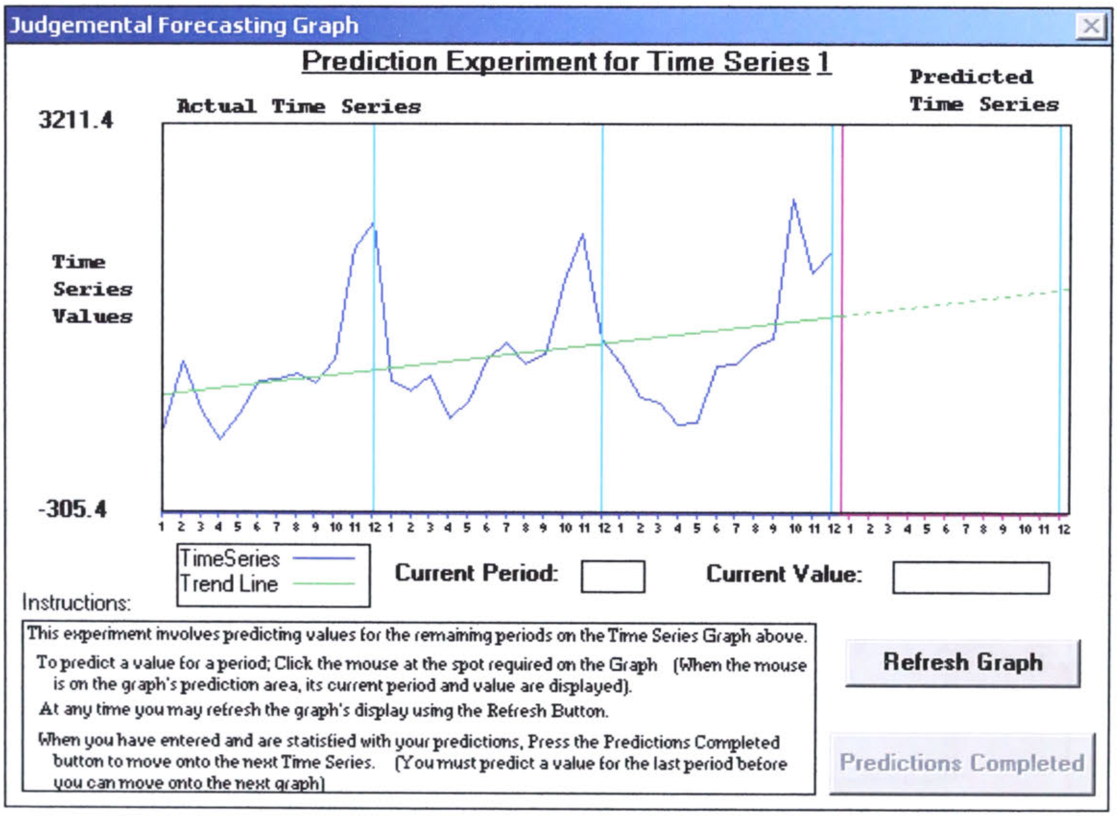
Appendix 6E.4

Demonstration System - Graphical Presentation of Time Series Images

Demonstration System images as presented to subjects in Experiment 4 – screen-based treatment condition using Tool 3.

Tool 3:

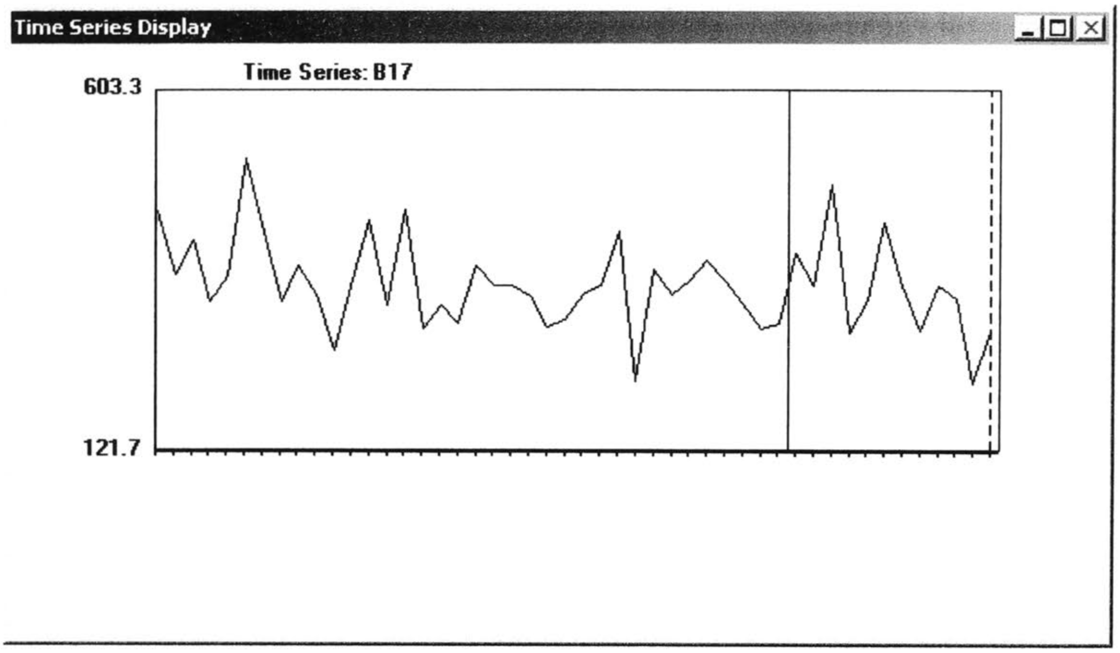
- *Media* – screen-based;
- *Technology* – mouse;
- *Task Complexity* – (2) structured, less complex



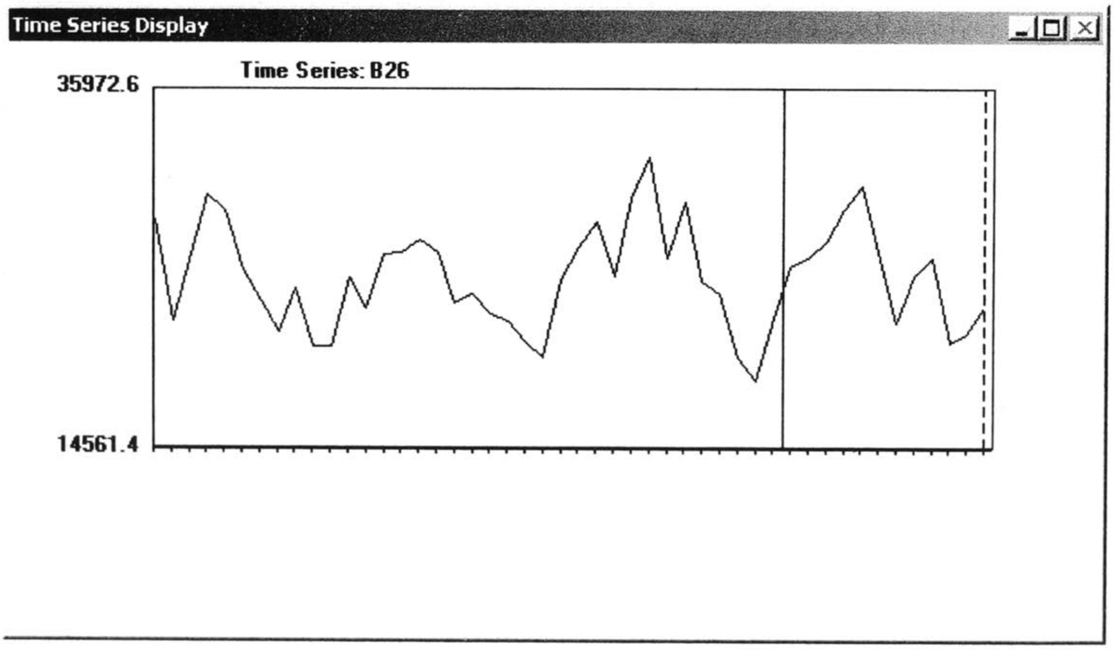
Appendix 6F.1

Time Series Images with Actual Values

Time Series B17



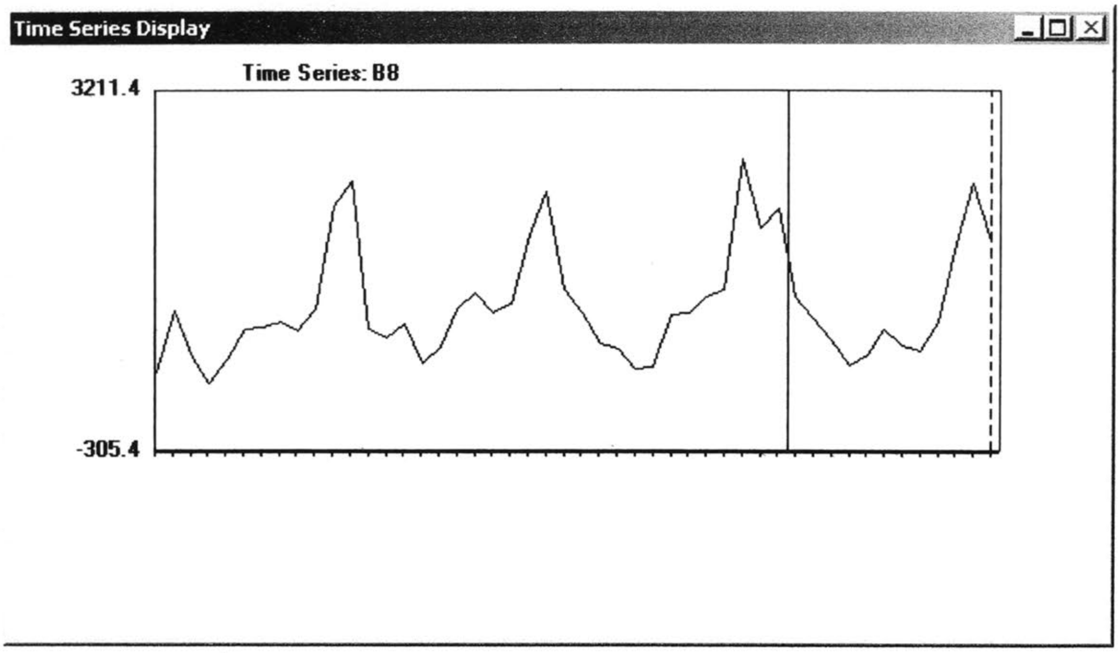
Time Series B26



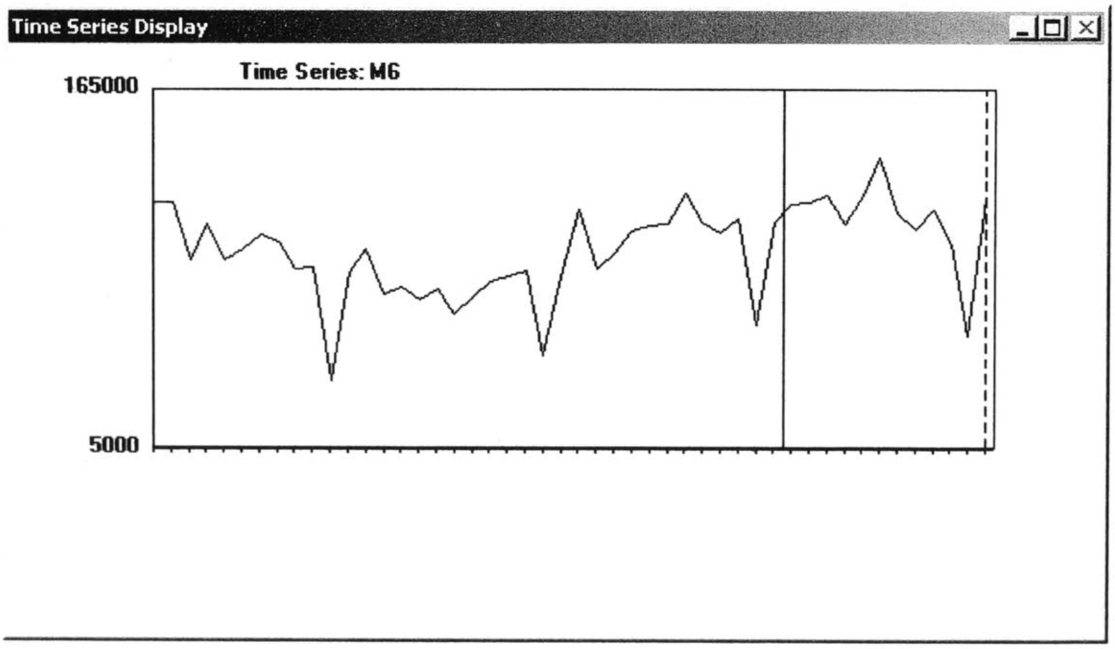
Appendix 6F.2

Time Series Images with Actual Values

Time Series B8



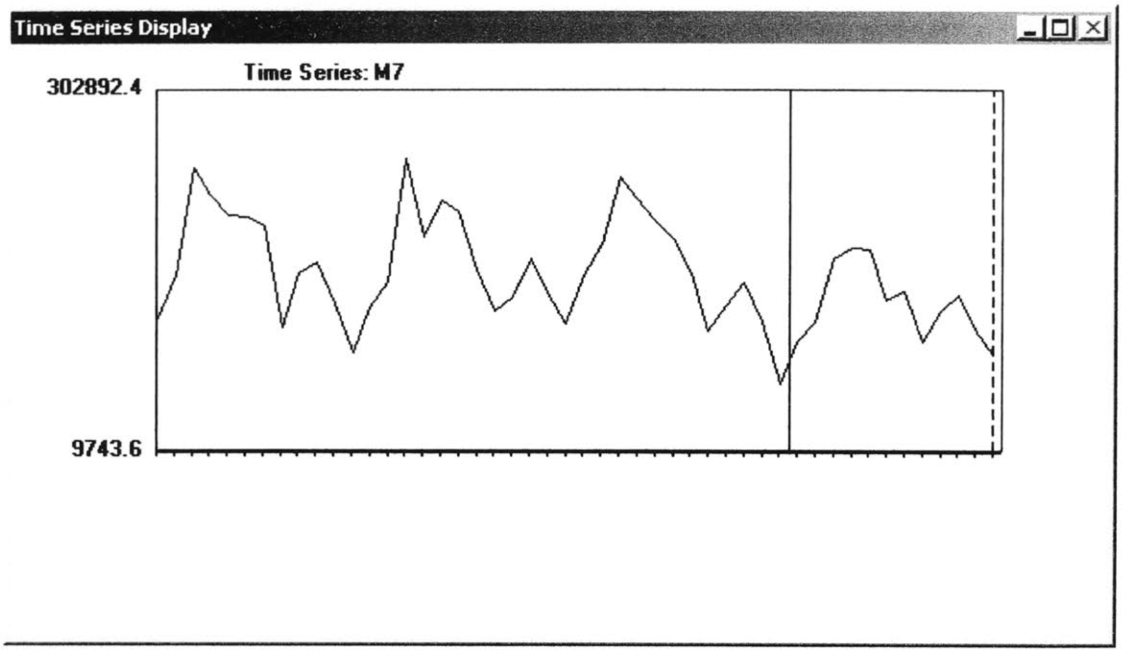
Time Series M6



Appendix 6F.3

Time Series Images with Actual Values

Time Series M7



10.7 Appendix 7

Screen Based Tools/ Source Code

10.8 Appendix 8

Experimental Data