The Relationship between Complex Problem Solving and Intelligence:
An Analysis of Three Computer Simulated Scenarios

Katherine Jane Ryan

School of Psychology
The University of Sydney
New South Wales
AUSTRALIA

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

2006
STATEMENT OF ORIGINALITY

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

Katherine Jane Ryan

Signed: __________________________    Date: __________

29 March, 2006
DEDICATION

I would like to thank my original supervisor, Lazar Stankov, for teaching me everything I know about ‘intelligence’ theory and for having faith in my abilities. I am both proud and humbled to be a (small) part of the Spearman, Cattell, Horn, Stankov, and Fogarty lineage of supervisors. Special thanks must go to Alan Craddock and Damian Birney who took over my supervision and provided guidance and generous support. Sincere thanks also to John Crawford for his role as a mentor throughout. Thank you to Robert Wood, Dietrich Wagener, and Heinz-Martin Süß for giving me access to their computer simulations and for teaching me the intricacies of the programs. I would especially like to thank Robert Wood for guiding my interest in ‘organisational behaviour’.

Thank you to Sabina Kleitman who showed a Melbournian the ropes at Sydney University and made summer statistics camps in Canberra actually enjoyable! I also owe a huge thank you to Andrew Cartwright for imbuing me with his computer expertise and making the computerised testing for this thesis possible. Thank you to my maternal Austrian side ‘the Renz a.k.a. Renc family’, particularly my aunt Rita Miles, for assistance with the translation of much of the material for this thesis from German to English. I would also like to thank my aunt Carol and uncle/Professor Chris Harman and uncle Peter Ryan for inspiring me to follow intellectual pursuits.

Most importantly, I would like to thank my parents Irene (Renc) and Terry Keating, and Doug Brimble, to whom I dedicate this thesis. Thank you for providing me with a loving and supportive environment that allowed me to fulfil my dreams. In loving memory of my father, Stephen Charles Ryan [15 March1950 – 06 January1995].
# TABLE OF CONTENTS

1  The Measurement of Problem Solving  
   1.1  *Simple Problem Solving*  
   1.1.1  Gap Definition  
   1.1.2  Simple Problem Solving and Complex Problem Solving: Task Comparison  
   1.2  *Complex Problem Solving: The Task*  
   1.2.1  Complexity  
   1.2.2  Connectivity  
   1.2.3  Dynamic Environment  
   1.2.4  Intransparency  
   1.2.5  Polytely  
   1.2.6  Complex Problem Solving Tasks: A Definition  
   1.2.7  Complex Problem Solving and Traditional Intelligence Tests: Task Comparison  
   1.3  *Complex Problem Solving: The Problem Solver*  
   1.4  *Complex Problem Solving: The Environment*  
   1.5  *Complex Problem Solving in the Thesis*  
2  Theories of Complex Problem Solving and Intelligence  
   2.1  *Methodological Approaches to Complex Problem Solving: North America*  
   2.1.1  Computer Games and Complex Problem Solving Tasks: A Comparison  
   2.1.2  Computerisation of Traditional Intelligence Tests  
   2.1.3  Ecological Validity of Computer Simulations  
   2.1.4  Summary: North American Research  
   2.2  *Methodological Approaches to Complex Problem Solving: Europe*  
   2.2.1  Dörner School vs. Broadbent School  
   2.2.2  Complex Problem Solving and Intelligence  
   2.2.3  Explicit and Implicit Complex Problem Solving  
   2.2.4  Theoretical Approach of the Thesis  
   2.3  *Naturalistic Decision Making*  
   2.3.1  Field Research vs. Laboratory Research  
   2.3.2  Naturalistic Decision Making vs. Complex Problem Solving  
   2.3.3  Summary: Naturalistic Decision Making Research  
3  Hierarchical Theories of Intelligence  
   3.1  *The Theory of Fluid and Crystallised Intelligence: A Review*
### 3.2 Complex Computer Simulated Scenarios: Some Examples

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary: Complex Computer Simulated Scenarios</td>
<td>62</td>
</tr>
</tbody>
</table>

### 3.3 Overview of the Thesis

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 Specific Components of Intelligence as Predictors of Complex Problem Solving: Study 1</td>
<td>69</td>
</tr>
</tbody>
</table>

#### 4.1 Introduction

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different Demands Hypothesis</td>
<td>69</td>
</tr>
<tr>
<td>Low Reliability Hypothesis</td>
<td>72</td>
</tr>
<tr>
<td>Brunswik Symmetry</td>
<td>74</td>
</tr>
<tr>
<td>Berlin Intelligence Structure Model</td>
<td>76</td>
</tr>
<tr>
<td>Specific Components of intelligence as Predictors of Complex Problem Solving</td>
<td>78</td>
</tr>
<tr>
<td>Gender and Group Differences in Complex Problem Solving</td>
<td>81</td>
</tr>
<tr>
<td>Strategy in Complex Problem Solving</td>
<td>84</td>
</tr>
<tr>
<td>Naturalistic Decision Making in Stocks and Flows Tasks</td>
<td>91</td>
</tr>
<tr>
<td>Overview of Predictions: Study 1</td>
<td>94</td>
</tr>
</tbody>
</table>

#### 4.2 Method

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>98</td>
</tr>
<tr>
<td>Materials</td>
<td>99</td>
</tr>
<tr>
<td>Cognitive Abilities Tests</td>
<td>99</td>
</tr>
<tr>
<td>Complex Problem Solving Tasks</td>
<td>107</td>
</tr>
<tr>
<td>Dynamic Forecasting Questionnaire</td>
<td>107</td>
</tr>
<tr>
<td>Furniture Factory</td>
<td>110</td>
</tr>
<tr>
<td>Decision Rules for Furniture Factory Simulation</td>
<td>116</td>
</tr>
<tr>
<td>Furniture Factory Measures: Performance</td>
<td>117</td>
</tr>
<tr>
<td>Furniture Factory Measures: Strategy</td>
<td>118</td>
</tr>
</tbody>
</table>

#### 4.3 Procedure

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Analyses</td>
<td>120</td>
</tr>
</tbody>
</table>

#### 4.4 Results and Discussion

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Statistics for Cognitive Abilities Variables</td>
<td>121</td>
</tr>
<tr>
<td>Descriptive Statistics for Complex Problem Solving Variables</td>
<td>123</td>
</tr>
<tr>
<td>Correlations between Cognitive Abilities and Furniture Factory Trials</td>
<td>124</td>
</tr>
<tr>
<td>Correlations between Cognitive Abilities and Complex Problem Solving Tasks</td>
<td>128</td>
</tr>
<tr>
<td>Regression of Dynamic Forecasting Questionnaire Performance on Cognitive Abilities Factors</td>
<td>132</td>
</tr>
<tr>
<td>Gender Differences</td>
<td>133</td>
</tr>
<tr>
<td>Group Differences</td>
<td>134</td>
</tr>
<tr>
<td>Exploration Strategy</td>
<td>135</td>
</tr>
</tbody>
</table>

#### 4.5 Conclusion: Study 1

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>137</td>
</tr>
</tbody>
</table>
5 Introduction: Study 2 145

5.1 Scoring Issues in Complex Problem Solving Research: Brunswik Symmetry 146

5.1.1 Goal Achievement Scores 148
5.1.2 Factor Analysis of Complex Problem Solving Performance 150
5.1.3 Relationships of Complex Problem Solving Performance across Three Computer Simulations 151
5.1.4 Loopholes 151

5.2 Relationship of Complex Problem Solving and Intelligence 152

5.2.1 Relationship between Intelligence and Complex Problem Solving Performance on Three Computer Simulations 152
5.2.2 Predictors of Complex Problem Solving Performance 154

5.3 Relationship between Complex Problem Solving and Non-Cognitive Variables 155

5.3.1 Personality 155
5.3.2 Interests 155

5.4 Summary of Key Predictions 156

6 Method: Study 2 159

6.1 Participants 159

6.2 Materials 160

6.2.1 Cognitive Abilities Tests: Computerised 160
6.2.2 Cognitive Abilities Tests: Paper and Pencil 162
6.2.3 Personality 167
6.2.4 Interests 167
6.2.5 Complex Problem Solving Tasks 168

6.2.5.1 Tailorshop: The Task 169
6.2.5.2 Tailorshop: The Measures 174
6.2.5.3 Forestry System (FSYS 2.0) 176
6.2.5.4 Forestry System: Translation 176
6.2.5.5 Forestry System: The Task 179
6.2.5.6 Forestry System: The Measures 184

6.3 Development of a Consistent Complex Problem Solving Score: Goal Achievement 186

6.3.1 Goal Achievement Score: Furniture Factory 187
6.3.2 Goal Achievement Score: Forestry System 188

6.4 Procedure 189
6.4.1 Statistical Analyses 191

7 Results and Discussion: Study 2 192

7.1 Data Screening 192
<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Appendix A: Computer simulated scenarios employed in individual differences research</td>
<td>270</td>
</tr>
<tr>
<td>B</td>
<td>Appendix B: Post-test questionnaire measuring rule knowledge of Furniture Factory simulation</td>
<td>284</td>
</tr>
<tr>
<td>C</td>
<td>Appendix C: Participant handout for the Furniture Factory task</td>
<td>286</td>
</tr>
<tr>
<td>D</td>
<td>Appendix D: Tailorshop Simulation - Participant Reference Sheet</td>
<td>287</td>
</tr>
<tr>
<td>E</td>
<td>Appendix E: Tailorshop Simulation – Directions for The Second Practice Month</td>
<td>289</td>
</tr>
<tr>
<td>F</td>
<td>Appendix F: Test manual for Forest microworld task FSYS 2.0</td>
<td>292</td>
</tr>
<tr>
<td>G</td>
<td>Appendix G: Forest Simulation - Participant Reference Sheet</td>
<td>315</td>
</tr>
<tr>
<td>H</td>
<td>Appendix H: Confirmatory factor analysis of cognitive abilities test battery</td>
<td>321</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 4.1 Summary Statistics for Cognitive Abilities Tests 122
Table 4.2 Summary Statistics for Complex Problem Solving Tasks 123
Table 4.3 Correlation Matrix of Cognitive Abilities and Furniture Factory Trials 1-18 125
Table 4.4 Correlation Matrix of Cognitive Abilities and Complex Problem Solving Tasks (the Furniture Factory and the Dynamic Forecasting Questionnaire) 128
Table 4.5 Stepwise Regression of Dynamic Forecasting Questionnaire Performance on Cognitive Abilities Factors 132
Table 4.6 Correlation Matrix of Furniture Factory Performance and Furniture Factory Strategy 135
Table 7.1 Descriptive Statistics for Cognitive Abilities Tests 194
Table 7.2 Correlation Matrix of Cognitive Abilities Tests 195
Table 7.3 Factor Loadings, Communalities (h\(^2\)), Percents of Variance for Principal Component Analysis and Oblique Rotation on Cognitive Abilities Tests 199
Table 7.4 Descriptive Statistics for Complex Problem Solving Variables 203
Table 7.5 Correlation Matrix of Complex Problem Solving Variables 205
Table 7.6 Correlation Matrix of Cognitive Abilities and Furniture Factory Performance 208
Table 7.7 Correlation Matrix of Intelligence Factors and Furniture Factory Performance 209
Table 7.8 Correlation Matrix of Cognitive Abilities and Tailorshop Performance 209
Table 7.9 Correlation Matrix of Intelligence Factors and Tailorshop Performance 210
Table 7.10 Correlation Matrix of Cognitive Abilities and Forestry System Performance 211
Table 7.11 Correlation Matrix of Intelligence Factors and Forestry System Performance 212
Table 7.12  Stepwise Regression of Furniture Factory Performance on Intelligence Factors  214
Table 7.13  Stepwise Regression of Tailorshop Performance on Intelligence Factors  215
Table 7.14  Stepwise Regression of Forestry System Performance on Intelligence Factors  216
Table 7.15  Correlation Matrix of Personality and Complex Problem Solving Performance on the Furniture Factory  218
Table 7.16  Correlation Matrix of Personality and Complex Problem Solving Performance on the Tailorshop  219
Table 7.17  Correlation Matrix of Personality and Complex Problem Solving Performance on the Forestry System  220
Table 7.18  Correlation Matrix of Interests and Furniture Factory Performance  221
Table 7.19  Correlation Matrix of Interests and Tailorshop Performance  223
Table 7.20  Correlation Matrix of Interests and Forestry System Performance  225
Table 7.21  Correlation Matrix of Biodata with Intelligence and Complex Problem Solving  227
Table 7.22  Component Matrix for Furniture Factory performance  230
Table 7.23  Correlation Matrix of Cognitive Abilities Factors and the Furniture Factory Component Score  231
Table 7.24  Component Matrix for Tailorshop performance  232
Table 7.25  Correlation Matrix of Cognitive Abilities Factors and the Tailorshop Component Score  233
Table 7.26  Factor Loadings, Communalities (h²), Percents of Variance for Principal Component Analysis and Oblique Rotation on Forestry System Performance  235
Table 7.27  Correlation Matrix of Forestry System Factor Scores and Cognitive Abilities  238
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 4.1</td>
<td>Theoretical model of Furniture Factory complex problem solving competencies and specific components of intelligence.</td>
<td>87</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>DFQ Task. Graphical integration task with population growth cover story. DFQ questions were based on the dynamic system shown.</td>
<td>108</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Part of the explanation of the dynamic system and practice question 2 from the DFQ.</td>
<td>109</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Screen image from Week 1 of the Furniture Factory computer Simulation.</td>
<td>111</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Example item from the Paper Folding task</td>
<td>162</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>Example item from the Raven’s Progressive Matrices task.</td>
<td>164</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>Screen image from the Tailorshop computer simulation.</td>
<td>171</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>Screen image from Month 1 of the Forestry system (FSYS 2.0).</td>
<td>180</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>Scree plot indicating the extraction of three factors.</td>
<td>198</td>
</tr>
<tr>
<td>Figure 7.2</td>
<td>Scree plot indicating the extraction of four factors for Forestry System performance.</td>
<td>234</td>
</tr>
</tbody>
</table>
ABSTRACT

The gap between field research and laboratory research has always been a problem in psychology. With the introduction of computers into the laboratory, computer simulated tasks allowed the observation of complex problem solving performance in the laboratory with a higher degree of ecological validity than ever before. The main aim of this thesis was to explore the relationship between complex problem solving ability and intelligence by presenting the results of two studies, using over 400 adults. Complex problem solving ability was assessed by performance on three computer simulations: Furniture Factory, Tailorshop, and Forestry System. The theory of fluid and crystallised intelligence guided the selection of cognitive abilities tests. Relationships between broad cognitive abilities including Fluid reasoning (Gf), Acculturation knowledge (Gc), Visual processing (Gv), Quantitative knowledge (Gq), and Processing speed (Gs) with computer simulation performance were explored.

Previous research exploring the relationship between complex problem solving and intelligence has led to inconsistent and often contradictory findings. Scoring problems in previous research were addressed and for all three computer simulations, relationships between intelligence and complex problem solving were found. Overall, Gf and Gc explained 20% of the variance in complex problem solving. Correlations between intelligence and complex problem solving increased when specific cognitive abilities tests and aggregated computer simulation scores were employed, rather than the employment of general or factor scores of intelligence and final computer simulation scores. A new aggregated scoring technique (goal achievement) that allowed consistent scoring across different computer simulations was developed. The strongest relationship between intelligence and complex problem solving was observed between
goal achievement scores and specific tests of cognitive abilities such as esoteric analogies and critical reasoning. There were significant correlations between goal achievement on the Furniture Factory and both esoteric analogies and critical reasoning ($r = .37, p < .05, r = .41, p < .05$) respectively. Correlations between goal achievement on the Tailorshop and both esoteric analogies and critical reasoning were significant ($r = .25, p < .05, r = .29, p < .05$) respectively. Correlations between goal achievement on the Forestry System and both esoteric analogies and critical reasoning were also significant ($r = .38, p < .05, r = .30, p < .05$) respectively. In addition, performance scores on all three computer simulations were correlated with one another. These findings support the application of the Brunswik lens model to complex problem solving research. Negative correlations, albeit rather modest, were observed between neuroticism and complex problem solving performance on the Furniture Factory ($r = -.17, p < .05$) and the Tailorshop ($r = -.21, p < .05$), indicating that emotion may also mediate complex problem solving performance. Results of this thesis may bring individual differences research in this area a step closer to obtaining stable results from which generalisations about complex problem solving tasks can be made.
ACKNOWLEDGEMENTS

Study 1 of this thesis was conceived of by Prof Robert Wood, Prof Lazar Stankov, and Dr Paul Atkins. Each of these advisors developed a test that was employed in the study. Prof Wood is the author of the Furniture Factory computer simulation and offered advice on computer programming, instructions for participants, and the scoring protocol. Prof Stankov developed Stankov’s Tests of Cognitive Abilities (STOCA) intelligence test battery. As the main supervisor at the beginning of the candidature, Prof Stankov imbued the candidate with advice regarding the theory of fluid and crystallised intelligence (Gf-Gc theory), and psychometric testing in general. As a reflection of the extent of knowledge that was passed on, the candidate lectured in this area during the final stages of candidature. Paul Atkins developed the Dynamic Forecasting Questionnaire (DFQ). Paul Atkins provided the candidate with methodologies and scoring protocols for the DFQ.

Study 1 was undertaken under the guidance of Prof Stankov, from which process, the candidate gained considerable insight into the workings of advanced multivariate statistics, specifically factor analytic techniques. The assistance of Prof Stankov throughout the candidature is gratefully acknowledged. This study represents the first time that the relationship between complex problem solving tasks and 2nd order factors of cognitive abilities based on Gf-Gc theory has been explored. Previous research has employed a general factor of intelligence or subcomponents of intelligence from the Berlin Intelligence Structure (BIS) test.
The assistance and instruction provided by Dr Andrew Cartwright regarding computer programming, licensing, and technical difficulties for all experiments is gratefully acknowledged. Dr Cartwright imbued the candidate with computer knowledge in the early stages of the thesis, so that subsequent studies were carried out without technical assistance.

The main experiment, Study 2 of this thesis, was conceived of solely by the candidate. The forestry computer simulation, FSYS 2.0, was translated from German to English specifically for this thesis. Study 2 represents the first time this computer simulation has been administered in English. Grateful acknowledgement is given to Dr Dietrich Wagener, the test developer, for his contribution to the translation process, computer programming advice, and assistance with the scoring protocol. The assistance from several other translators, including Rita Miles and Heidi Krause is also acknowledged. The candidate (of Austrian extraction) was solely responsible for the translation process that was conducted concurrently with correspondence with the test developer.

The Tailorshop simulation, employed in Study 2, was provided by Dr Heinz-Martin Süß, who has conducted extensive research exploring the relationship between this task and the Berlin Intelligence Structure (BIS) test. This thesis represents the first time that the relationship between performance on the Tailorshop task and 2nd order cognitive abilities factors from Gf-Gc theory has been explored. The advice provided by Dr Süß regarding computer programming, administration, and the scoring protocol is gratefully acknowledged.
The candidate was solely responsible for the conception of the main experiment, Study 2, of this thesis and the development of the paper and pencil intelligence test battery. For all chapters, test battery preparation, test administration, analysis of data, writing up, and discussion of the outcomes was undertaken by the candidate alone.

Prior to commencement of all experiments, Human Research Ethics Committee approval was granted, Ref No. 1710 (00/04/37). The chief investigator for Study 1 was Prof Lazar Stankov. The chief investigator for Study 2 was Dr Alan Craddock, authorised personnel also included Dr Damian Birney. The participants who gave up their time to be a part of this important research are also gratefully acknowledged. Special thanks must go to those who participated solely due to their interest in the subject matter, rather than as a course requirement. A study of this scale (over 2000 subject hours) would simply not have been possible without the generosity of these volunteer participants who arose from different departments within the university and from the general public. Finally, Dr Alan Craddock and Dr Damian Birney are gratefully acknowledged for reviewing and providing insightful comments on drafts of this thesis.
CHAPTER 1

THE MEASUREMENT OF PROBLEM SOLVING

1 An Introduction to Problem Solving

Throughout human history, the ability to solve problems has been the most important human faculty. In our ancient past, complex problem solving allowed us to achieve great feats, such as the construction of the Egyptian pyramids. Currently, complex problem solving is facilitating our understanding of complex human biology with developments in the Human Genome Project. In our future, the largest and most complex scientific project in human history will achieve its objective to move humans from our home planet, with advancements in the International Space Station Project. In contrast, as Turner and Pidgeon (1997) suggest in their volume on man-made disasters, errors in complex problem solving have led to many of the catastrophic events that chequer human history, for example, the Chernobyl nuclear disaster and the space shuttle Challenger tragedy. Regardless of how researchers choose to define it, problem solving has shaped human history and will continue to shape our future (Wenke, Frensch, & Funke, 2005) both in terms of extraordinary achievements and in terms of avoidable disasters in which great sums of money, or more tragically human lives, stand to be unnecessarily lost.

In their work on why decisions fail, Kahneman and Tversky (1979) integrated insights from psychological research into economic theory. They suggested that the weight humans place on their decisions does not correspond to objective probabilities in reality. As a reflection of the importance of this work, psychologist Daniel Kahneman was awarded the 2002 Nobel Prize for economics. In a similar vein, in his volume ‘The
Logic of Failure’ Dietrich Dörner utilises complex computerised scenarios to investigate why there are errors in human judgement and decision-making despite all of our intelligence, experience, and information (Dörner, 1996).

Computerised scenarios not only offer new ways to assess problem solving in the laboratory, they can be developed as contextualised assessment tools to be used for selection and training in educational, organizational, and military applications. Over the last 10 years, there has been a trend towards assessment of problem solving abilities in more practical settings. These include measures of constructs such as emotional intelligence (Mayer & Salovey, 1997; Mayer, Salovey, & Caruso, 2000), practical intelligence (Sternberg, 1997), situational judgement (Motowidlo, Dunnette, & Carter, 1990), and critical thinking (Derry, Levin, & Schauble, 1995). These contextualised assessment tools possess greater face validity than traditional intelligence tests (Kyllonen & Soonmook, 2005) and have been widely embraced by business, educational institutions, and the military.

In Sternberg and Berg’s (1986) comparison of definitions of intelligence since the 1920s to the modern day, the ability to solve problems has featured in nearly every attempt to describe human intelligence. Intelligence and problem solving are theoretically inextricably linked. However, the direction of causation (i.e., whether intelligence predicts problem solving competencies, or whether the dependence is the other way around) is unknown due to a lack of evidence (Wenke & Frensch, 2003). The main goal for the present research was to investigate the extent to which individual differences in problem solving ability and individual differences in specific intelligence components
are related. The focus is on complex problem solving (CPS) rather than simple problem solving (SPS). These two types of problem solving will be described next.

1.1 Simple Problem Solving

The results of problem solving are not always of the magnitude described in the examples earlier of problem solving gone awry as in the Chernobyl disaster or beneficial problem solving such as the human genome project. Not every task that humans encounter requires complex problem solving (CPS). In fact, many of our day-to-day decisions require simple problem solving (SPS). Simple problem solving tasks were developed in the 1930s by the early experimental work of the Gestaltists in Germany (Wenke, Frensch, & Funke, 2005).

Simple problem solving tasks require creative ingenuity and restructuring of information. A classic example is Maier’s (1930) nine-dot problem. The participant is instructed to connect nine dots arranged in a square without lifting the pen from the paper and without retracing any lines. This is where the phrase “think outside the square” originated. Other classic examples of simple problem solving tasks include Duncker’s (1945) candle task and Duncker’s radiation problem. In Duncker’s candle task, participants are given candle, a box of nails, and matches. They are instructed to fix the candle to the wall so that no wax will drip on to the floor. These tasks have also been called insight tasks and have been said to measure ‘functional fixedness’. Very few participants were able to provide the correct response of nailing the empty nail box to the wall to use it as a candleholder. This is a good example of the difference between difficulty and complexity, Duncker’s candle task was difficult because of the low
success rate of participants to find a solution. However, the task is ‘simple’ in terms of complexity because there are a small number of variables that must be taken into consideration.

Participants also had great difficulty with Duncker’s radiation problem (1945, cited in Mackintosh, 1998):

A patient has an inoperable stomach tumour. The tumour can be destroyed by radiation, but although weak radiation will not damage normal flesh, it will also not destroy the tumour. Radiation strong enough to destroy the tumour will also destroy normal flesh. How can the surgeon use radiation to treat the tumour? (p. 308)

These intuitively simple, though not psychometrically simple, laboratory tasks continued to be used for problem solving research until the early 1970s. The disk problem (Ewert & Lambert, 1932), later known as the famous Tower of Hanoi (e.g. Simon, 1975), has been extensively used in neuropsychological research (Shallice, 1988). The problem requires the participants to move disks across different pegs to achieve a set order in a minimum amount of moves. The problem has a working memory load because for optimal performance participants must keep track of the consequences of their actions before committing themselves to a set of moves, rather like chess. The prefrontal cortex is engaged while participants attempt to solve the problem as evidenced by PET scans (Morris, Ahmed, Syed, & Toone, 1993). Neurocognitive evidence from patients with damage (e.g., stroke, surgery, blows to the head, brain tumours) to the prefrontal cortex is associated with a range of deficits and
given the umbrella term ‘frontal syndrome’. Patients with frontal syndrome are severely impaired when it comes to problem solving tasks such as the Tower of Hanoi. Baddeley and Wilson (1988) characterise this impairment functionally as dysexecutive syndrome. Neuropsychological evidence supports the existence of complex problem solving as a real phenomenon rather than a theory.

Frontal patients are severely damaged in their attempts to solve problems such as the Tower of Hanoi or the Tower of London (3 peg version). This deficit extends to other problem solving tasks where patients are unable to initiate solutions to novel problems. They ignore relevant feedback and do not adapt their responses to changing circumstances but rather persevere with their original solution strategies. Thus, neurocognitive evidence suggests that the prefrontal cortex is activated during problem solving. However, since one third of the human cortex is prefrontal cortex, it cannot be claimed to serve a problem solving function exclusively (Mackintosh, 1998).

Problem solving tasks such as those just described were relatively simple and novel to participants (Mayer, 1992). These tasks could be administered and solved by the participant within a relatively short time frame. Experimenters could relatively easily trace the participants’ steps towards a clearly specified optimal solution with straightforward scoring protocols. In addition to convenience of administration, it was thought that generalisations could be made from the results of these simple tasks to more complex real world problems. Simple tasks such as the Tower of Hanoi were believed to present the same task demands to participants as real world problems. That is, researchers (e.g., Newell & Simon, 1972) assumed that the cognitive abilities required to solve laboratory tasks of low complexity were the same as those required to
solve highly complex real world problems. A “gap” definition of problem solving (Frensch & Funke, 1995) to explain why this was not the case will be provided next.

1.1.1 Gap Definition of Problem Solving

When an attempt is made to solve a problem, there is a gap between the information presented and the solution to the problem. The extent of knowledge and possibly intelligence participants must possess to solve the problem, the clarity of the problem state, and the goal state are all variables that will vary in magnitude depending on the person and the problem they are presented with. In other words, there are two gaps: a) the difference between the current state and the goal state, and b) the difference between the ‘ability’ of the person and the ‘ability-level’ required by the problem. The important component of the gap definition is the distance between the task and the solver. There is a barrier between the problem state of the task and the goal state the participant is trying to achieve. Simple problem solving and complex problem solving are qualitatively different because simple problems typically have only one ‘barrier’ whereas complex problems have multiple concurrent ‘barriers’ (Frensch & Funke, 1995).

In complex problem solving (CPS) the participants must plan and prioritise their actions in order to reach the goal state. In simple problem solving (SPS) a single action is all that is required to overcome single barrier, so the cognitive processes required to reach the solution can be equated with the requirements of the task in a one to one ratio. However, in CPS there is a complex interaction between the requirements of the task
and the actions of the problem solver who will use multiple ‘mental steps’ to get closer to the solution. In SPS the participant is explicitly told the goal of the problem. In contrast, there is no one single correct solution strategy for CPS, rather problem solving competency is measured as a function of how close the participant gets to the goal which may or may not be clearly specified (Wenke & Frensch, 2003). This makes scoring of CPS problems rather more complicated than scoring of SPS problems and is an issue that is investigated in the studies in following chapters. In addition to posing scoring problems for experimenters, the absence of a single optimal solution strategy in CPS poses problems for the use of these tasks as recruitment tools (Funke, 1995).

1.1.2 Simple Problem Solving and Complex Problem Solving: Task Comparison

During the 1970s, it was conceded that empirical findings and theoretical concepts that had been developed based on the use of simple problem solving tasks could not easily be generalised to more complex, real world problems (Wenke, Frensch, & Funke, 2005). To compound the matter further, it was suggested that the cognitive processes underlying complex problem solving were different across different domains and across levels of expertise (Sternberg, 1995). Fortuitously, these realisations took place around the same time as the inception of the Personal Computer (PC) as experimental equipment in psychology laboratories. Researchers in the area of problem solving could use PCs to design complex problem solving tasks that would closer approximate real world problems.
In the studies presented in the following chapters, there is an emphasis on problems that consist of relatively large “gaps” between the information presented and the solution to the problem. The computer-simulated scenarios adopted in the present research more closely resemble real world problems than simple problem solving tasks such as the Tower of Hanoi. There is already extensive research exploring the relationship between intelligence and simple problem solving (e.g. Sternberg, 1982), the correlations are typically quite modest, around .30 (Wenke & Frensch, 2003). However, the relationship between complex problem solving, as measured by computer simulated tasks, and intelligence has received relatively minor discussion. Therefore, this relationship is the focus of the present research.

1.2 Complex Problem Solving: The Task

Early stages in the history of complex problem solving research will be described in the next chapter after defining what is meant by complex problem solving in terms of the task, person, and environment. During this initial research period, the following five features were suggested as intrinsic to complex problem solving tasks (Dörner, 1980; Dörner & Kreuzig, 1983): 1) Complexity (number of variables); 2) Connectivity (causal connections between variables); 3) Dynamics (changes carried out automatically by the system and not under participants’ control); 4) Intransparency (opaqueness, whether the underlying rules of the system are provided or must be inferred); and 5) Polytely (the pursuit of multiple goals). This list remained largely unchallenged and was adopted by the majority of complex problem solving researchers. However, some researchers critiqued this taxonomy of complex problem solving features (Funke, 1990,
2001; Hussy, 1984; Strohschneider, 1991). These criticisms will be reviewed and discussed next. Overall, the list remains predominantly the same and is accepted in some combination or another even by those who reviewed and critiqued it. Each of these features will now be described in turn and their adequacy in explaining complex problem solving tasks will be assessed. Finally, a definition of complex problem solving, as it will be used in this thesis will be provided.

1.2.1 Complexity

According to (Brehmer & Dörner, 1993), computer simulated scenarios used in complex problem solving research demonstrate complexity because the participant is required to take multiple aspects of the situation into account simultaneously. At the very basic level, complexity is defined as the number of variables within a computer simulated scenario. For example, in the computer simulated scenario the Furniture Factory (Wood & Bailey, 1985), the participant plays the role of the manager of a company. As the manager, the participants’ goal is to maximise the efficiency of their five employees by making four types of decisions: 1) Assign the right job to the right employee; 2) Set the employee a production target; 3) Advise the employee of his or her performance and 4) Allocate rewards to the employees. Thus, this version of the Furniture Factory is said to have a moderate degree of complexity because there are 20 decisions per trial, that is four decisions for each of the five employees.

Other computer simulated scenarios may have more or fewer variables depending on the approach or research school of the researcher. Obviously, researchers with different theoretical goals and different methodologies will choose more or less complex
computerised scenarios. An overview of the most commonly cited computer simulations and their relative applications will be provided in Section 3.2. The number of variables usually ranges from 3 to 20 (Funke, 2001) and can range up to 2000 variables, for example the complex computerised scenario, Lohhausen project (Dörner, Kreuzig, Reither, & Stäudel, 1983) in which the participant plays the role of the mayor of a small town. The different approaches to complex problem solving research and the different computer simulations they employ will be discussed in detail in Chapter 2, in a review of the history of complex problem solving research.

The simplistic definition of complexity as equated with the number of variables in the system has been criticised (Funke, 1984, 2001; Kotkamp, 1999; Strauß, 1993; Wallach, 1998). The above definition of complexity does not take into account the dependent relationships among two or more variables. This relationship among the variables is termed ‘connectivity’. Some researchers (Brehmer & Dörner, 1993; Rigas, Carling, & Brehmer, 2002) include ‘connectivity’ in their definition of ‘complexity’ rather than listing ‘connectivity’ as a completely separate and distinct feature of complex problem solving tasks. In the definition provided by Brehmer and colleagues, the number of variables and the number and structure of the causal connections between them are how the complexity of the system is determined.

Funke (2001) suggests that researchers should concentrate less on the component of ‘complexity’ as defined by the number of variables alone. Researchers should focus on both the number of variables and, more importantly, the causal connections between them. Funke also suggests that complexity and connectivity cannot be viewed as separate components as they cannot be differentiated. Overall, use of the term
complexity usually implies both the number of variables in the system and the connectivity between the variables (Brehmer & Dörner, 1993; Rigas, Carling, & Brehmer, 2002). However, it argued in this thesis that complexity and connectivity must be included as separate components in any definition of a complex problem solving task in order to distinguish these computerised complex systems from manipulations of complexity in traditional cognitive abilities tasks such as the Swaps task.

The Swaps task was developed by Stankov and associates (Stankov, 1999) to manipulate task complexity. Participants are instructed to mentally ‘swap’ the positions of 3 stimuli (e.g., pictures or letters). For example, in the level 1 ‘swaps’ condition, participants are presented with three pictures in the following order: Bottle, Eggs, Cards. They are then asked what the order of the pictures would be if you ‘swap’ 1 and 3. Answer: Cards, Eggs, Bottle. Complexity was manipulated by altering the number of times that the pictures were to be rearranged and included level 1, level 2, level 3, and level 4 swaps. For example, a level 4 swaps condition would ask the participants what the new order of the pictures will be if you Swap 1 and 3, then swap 1 and 2, then swap 1 and 3, then swap 2 and 3. Participants must remember the order of previous ‘swaps’ and then carry out further manipulations on additional ‘swaps’. Thus, the level of complexity can be increased by choosing conditions with a greater number of swaps. Increasing complexity was found to improve the correlation of these types of traditional cognitive abilities tasks with fluid reasoning, making them more psychometrically sound intelligence measures (Ackerman, 1988; Crawford, 1991; Fogarty & Stankov, 1988; Stankov & Crawford, 1993). The Swaps task employs varying levels of steps which manipulate task complexity, however there are no causal connections between
the steps, with each new step being completely independent of each previous step. Thus, traditional cognitive abilities tasks (e.g., Swaps; Stankov, 1999) can feature ‘complexity’ without ‘connectivity’, which will be described next.

1.2.2 Connectivity

Connectivity refers to the causal connections between two or more task variables and is an important feature of complex systems (Casti, 1979), such as computer simulated scenarios. The connections between variables are instantiated in the underlying mathematical algorithms of the computer simulations. The participant cannot control interconnected variables separately, as they are dependent on one another (Süß, Oberauer, & Wittmann). Thus, participants are forced to make tradeoffs (Brehmer & Dörner, 1993). For example, in the complex computer simulated scenario the Forestry task (FSYS 2.0; Wagener, 2001; Wagener & Conrad, 1996) the participant plays the role of the manager of a forestry business that grows trees and then cuts them down to sell the timber for profit. Tradeoffs must be made when deciding whether to continue to grow the trees, which will take more time (con) and generate greater profit (pro), or whether to cut the trees down early for immediate financial gain (pro) that would generate less profit (con).

Participants must also consider the side effects of possible actions (Brehmer & Dörner, 1993). In the Forestry task (FSYS 2.0; Wagener, 2001; Wagener & Conrad, 1996) just described, participants may choose to use a pest control poison that will result in two side effects. One possible side effect is positive in that the pest population will be reduced which in turn will minimise the damage the pests are inflicting on the trees.
Another possible side effect is negative in that the pest control poison may also be toxic to the trees and some proportion of the forest may die. For every action taken by the participant, there is a dependent side effect that may or may not be initially known to the participant.

If participants can learn the connections between the variables in the system, they can construct a causal model of how the system operates and adjust their actions accordingly. The causal connections between the variables may be explicitly provided to the participants at the outset. In this case, the system is said to be ‘transparent’. Alternatively, if the causal connections between the variables are hidden, the system is ‘intransparent or opaque’. Different computer simulations vary in their level of transparency and this is considered another separate feature of complex problem solving tasks. Intransparency will be discussed later in this chapter.

The variables in a complex problem solving task can be interconnected in a variety of ways. Nonlinear causal relationships are the most common, although linear relationships also exist. In computer simulations of economic or ecological systems, the relationships between variables will consist of logistic or exponential growth functions (Brehmer & Dörner, 1993), as these are the type of functions that tend to describe such non-linear relations in the real world. Another important feature of interconnected variables is the structure of their relations, presented to the participant as feedback. Both negative and positive feedback relations occur as a consequence of the participant’s actions and also as a result of the changes made automatically by the system. For example, in the Furniture Factory task (Wood & Bailey, 1985), the aim for participant-managers is to motivate their employees to work as efficiently as possible.
After the participant-manager has made decisions (e.g., allocation of employees to jobs, production targets, etc.) the system implements the underlying algorithm, taking into consideration the current state and the input of the participant-manager, and displays the actual number of hours the simulated employees took to complete their assigned jobs. Hence, the feedback provided to the participant occurs both as a consequence of the participant’s actions (under participant control) and the automatic calculations of the system (not under participant control). There is a causal relationship between the participant’s actions and the actions arising from automated system changes.

1.2.3 Dynamic Environment

The dynamic development of the situation is another distinguishing feature of complex problem solving tasks (Brehmer & Dörner, 1993; Frensch & Funke, 1995). The dynamics of a complex computer simulated scenario are related to the level of automatic (computer generated) system change (Rigas, Carling, & Brehmer, 2002). According to Funke (2001), connectivity and dynamics are the only two features that differentiate complex problem solving tasks from paper and pencil tests. Both connectivity and dynamics rely on computer presentation of the complex system. Connectivity refers to the structural relationships within the system, whereas dynamics refers to the processes within the system.

The system is dynamically changing over time and the participants must make decisions when the system requires them to be made in the current trial, they cannot take actions when they choose and at their own pace (Brehmer, 1990) as they could for a paper and
pencil test. The participant must also try to anticipate what will happen in future trials of the computer simulation (Wenke, Frensch, & Funke, 2005). They must try to anticipate the short and long term effects of their actions and the way in which the system develops over time (Funke, 2001).

During the performance of a computer simulation, the current state of the system depends on the history of interaction between the participants’ actions and the automatic (computer generated) response of the system (Brehmer & Dörner, 1993). For example, in the forestry task FSYS 2.0 (Wagener, 2001; Wagener & Conrad, 1996), the participant-manager may have been planting a new crop of trees every few simulated months on the same plot of land while neglecting to fertilize the soil. In the subsequent simulated months of the game, the soil will be depleted of all minerals and the growth of timber will be stunted. Thus, the state of the system in the current trial (i.e., the current month) is dependent on previous trials of the simulation. In this sense, the games could be described as having a memory of previous actions and their related effects (Brehmer & Dörner, 1993).

1.2.4 Intransparency

The intransparent or opaque character of complex problem solving tasks refers to the level of information that is provided to the participant regarding the causal connections between the variables. For example, in a previous study by Putz-Osterloh (1981) the Tailorshop computer simulation (originally developed by Dietrich Dörner and first used in a study by Putz-Osterloh) was presented in either a transparent or an intransparent
condition. The Tailorshop task requires participants to maximise the company’s profit by modifying the production capacity of a small shirt production and sales company. In the transparent condition of the task, the participants are provided with a model depicting all the causal relationships among the variables in the task. In the intransparent condition, the participants are unaware of the underlying rules of the task and must test their own hypotheses and learn from the effects of their actions (Brehmer & Dörner, 1993). Thus, in CPS tasks, only part of the information is available to the participants and the rest must be inferred as they progress through the system trials (Rigas, Carling, & Brehmer, 2002). The level at which the system is set on the intransparency to transparency continuum is a feature of complex problem solving tasks that can be modified by the experimenter. Funke (2001) argues that the amount of information provided by the experimenter to the participants could be also be varied in paper and pencil tasks and thus intransparency or opaqueness is not unique to complex computer simulated scenarios. Although it is not a unique feature of these tasks, it is certainly considered an important component nonetheless (Wenke, Frensch, & Funke, 2005).

1.2.5 Polytely

Polytely refers to the pursuit of multiple goals within a complex problem solving task (Frensch & Funke, 1995; Funke, 2001). Brehmer & Dörner (1993) do not explicitly use the term polytely in their definition of a complex problem solving task. However, their definition does refer to polytely when they describe the goals that the participant is instructed to follow as sometimes being contradictory, so that a trade off is required. In
addition to intransparency, polytely is another feature of complex computer simulated scenarios that is modifiable by the experimenter. Polytely is not inherent in dynamic systems but could also be applied to static experiments (Funke).

Polytely has not been included in more recent definitions of complex problem solving tasks (Wenke, Frensch, & Funke, 2005). The psychometric quality of the data collected in complex problem solving research can be lowered by polytelic or contradictory goals (Süß, Oberauer, & Wittmann, unpublished). Reliability of performance scores on complex computer simulated scenarios is increased when an optimum solution can be calculated, the testing conditions are standardised, an explicit goal is provided, and the participants are aware of the performance criteria (Süß, Oberauer, & Wittmann). Polytelic goals are likely to be more realistic, but they come at the cost of psychometric clarity.

Novelty is included in the definition of complex problem solving proposed by many researchers, as reviewed by (Frensch & Funke, 1995).

1.2.6 Complex Problem Solving Tasks: A Definition

In summary, the five most common distinguishing features of complex problem solving tasks are complexity, connectivity, dynamics, intransparency (opaqueness), and polytely. Connectivity and dynamics are features that distinguish computer administered complex tasks from paper and pencil tests (Funke, 2001). Connectivity cannot exist without complexity (Funke). However, complexity can exist without
connectivity, as described earlier in the “Swaps task” example. Therefore, in contrast to Funke’s suggestion to combine the two, both complexity and connectivity must be included as separate components when defining a complex problem solving task. With the current design of computer simulations, polytelic is not a desirable feature as it reduces the psychometric quality of the data (Süß, 2001). While connectivity has been included as a defining property of complex problem solving tasks in this thesis, the feature of polytelic will be excluded from the definition. In addition, performance scores derived from polytelic systems are inconsistent and subjective because they are based on the experimenter’s consideration of the importance of one goal over another, which may or may not coincide with what the participant infers to be the most important goal. Polytelic systems with multiple and contradictory goals that lead to vague scoring protocols are avoided in the studies presented in this thesis. Instead, well-defined systems with single goals will be employed to measure complex problem solving performance. The computer simulations employed in the present thesis do not have polytelic goals because there is contention over how to score complex problems with multiple goals. It is acknowledged that polytelic is often a property of real world problem solving. However, resolution of this problem and development of new design features and scoring protocols to capture polytelic in complex computer simulated scenarios is beyond the scope of this thesis. The complex computerised scenarios used in the current thesis have a single clearly defined goal, either to maximise company capital, in the cases of the Tailorshop simulation (Putz-Osterloh, 1981) and the Forestry simulation FSYS 2.0 (Wagener, 2001; Wagener & Conrad, 1996), or to maximise company efficiency in the case of the Furniture Factory simulation (Wood & Bailey, 1985). These clearly defined goals are needed in order to maintain experimental control over complex problem solving performance, so that its relationship with intelligence
may be investigated, albeit at the cost of using somewhat artificial simulations of the environment.

Based on the above review of the literature, complex problem solving tasks in this thesis will consist of four features: complexity, connectivity, dynamics and intransparency.

1.2.7 Comparison of Tasks: Complex Problem Solving and Traditional Intelligence Tests: Task Comparison

Although a 4-feature taxonomy is employed in the definition of complex problem solving tasks in this thesis, the 5-component taxonomy is useful for comparing and contrasting traditional intelligence tests such as the WAIS-III (Wechsler, 1997) and the WJ-R (Woodcock, McGrew, & Mather, 2001) with complex computer simulated scenarios. In contrast to a complex problem solving task, a traditional intelligence test includes relatively simple questions, where there are one or two independent variables of interest (generally low complexity and no connectivity). The state of the current item does not depend on the results of previous actions; for example, each question in the test is a complete unit (static system). All the information needed to solve the problem is presented to the participant at the outset and sometimes, in a multiple-choice set up, even the answers are provided to the participant (transparent). There is one specific goal, that is, to choose the correct answer for that particular item which may or may not be presented as a set of alternatives depending on whether the question is open ended or multiple choice format (not polytelic). Note that these five qualities all refer to the nature of the task itself, (i.e., the complex computer simulated scenario). This
taxonomy describes the task properties alone. However, in addition to the task, the act of complex problem solving also involves the person (the problem solver) and the environment (Frensch & Funke, 1995).

1.3 Complex Problem Solving: The Problem Solver

The “gap” definition of problem solving presented in Section 1.1.1 describes a gap or barrier between the information presented in the task (the current state) and the solution to the problem (the goal state) that the problem solver is trying to reach (Frensch & Funke, 1995). So far, this chapter has mainly focussed on task properties. However, the distance between the task and the person (problem solver) is the most important aspect of the definition. Thus, a complex problem can be defined by task features (barriers) alone as described in the 5-feature taxonomy (complexity, connectivity, dynamics, intransparency, polytely) presented above and in terms of the interaction between the task requirements and the solver. Moderate correlations have been found between the general intelligence of the problem solver and some aspects of complex problem solving (Beckmann & Guthke, 1995) although the causal direction is unclear (Wenke, Frensch, & Funke, 2005). Longitudinal studies may help establish whether individual differences in intelligence cause individual differences in complex problem solving or whether the direction of the relationship is the other way around. So far, there has not been any longitudinal research investigating the relationship between intelligence and complex problem solving, and that type of research is beyond the scope of this thesis. The transition from the current state to the goal state depends on several
aspects of the problem solver including memory contents, information processing, and non-cognitive variables (Frensch & Funke).

‘Memory contents’ refers to the static aspects of the problem solver such as implicit and explicit knowledge of the underlying rules of the complex problem solving task. Explicit knowledge is able to be articulated by the participant and can be assessed by rule knowledge, while implicit knowledge is unable to be articulated by the participant and can be assessed by rule application (Kröner, Plass, & Leutner, 2005). Complex problem solving performance can occur whilst the participant is unable to display explicit rule knowledge (Berry & Broadbent, 1995). Within “memory contents”, both domain general and domain specific knowledge are suggested to affect complex problem solving performance (Frensch & Funke, 1995).

An important distinction must be made between knowledge of the complex system and background knowledge regarding the semantic content of computer simulated scenarios. Explicit and implicit system knowledge assessed by rule knowledge and rule application is an important feature of complex problem solving. A very different type of knowledge is background knowledge regarding the “cover story” of a computer simulation. “Cover story” specific background knowledge should be highly relevant for assessment if designing complex computer simulated scenarios for training or recruitment purposes (Brehmer & Dörner, 1993). However, it can cause problems in studies of complex problem solving performance, where the cover story should be of little intrinsic interest. Previous research has found that cover stories (semantics of the labelling of input and output variables) can influence overall complex problem solving performance (Hesse, 1982; Putz-Osterloh, 1993; Stanley, Mathews, Buss, & Kotler-
Cope, 1989). Participants had higher overall performance scores, when controlling the behaviour of a fictional person in the PERSONAL INTERACTION computer simulation, than when controlling the size of a workforce to reach and maintain a given production level in the SUGAR FACTORY. Both PERSONAL INTERACTION and SUGAR FACTORY were based on exactly the same underlying mathematical algorithms, only the cover stories were semantically different. These 2 simulations are described in more detail in Appendix A. Thus, while knowledge of rules underlying the complex system is an important aspect of the problem solver, previous knowledge regarding the “cover story” of the computer simulation is of subsidiary interest. It is important to distinguish between these two types of knowledge in complex problem solving research. Effects of cover stories on complex problem solving performance will be investigating in Study 1, Chapter 4, of this thesis.

Krems (1995) suggested that experience in the domain affects the exploration and strategies that participants use to discover the underlying rules of the game. Individual differences exist in the sort of hypothesis testing participants engage in when they try different actions and learn from the resulting system feedback (Burns & Vollmeyer, 2002). Strategies, cognitive style, and the processes of task monitoring and progress evaluation influence complex problem solving performance (Frensch & Funke, 1995). Although not a main aim of this thesis, information processing aspects of the person such as exploration strategy (Goodman & Wood, 2004; Robert Wood & Bandura, 1989) will be included in the data analyses.

The term “non-cognitive variables” refers to the problem solvers’ self-efficacy, self-confidence, perseverance, enjoyment, and motivation (Frensch & Funke, 1995).
Complex problem solving performance is suggested to be improved by these non-cognitive aspects of the problem solver (Washburn, 2003; Wood & Bandura, 1989; Wood, Bandura, & Bailey, 1990; Wood, Atkins, & Tabenero, 2000; Wood, Kakebeeke, Debowsk & Frese, 2000). Recently, researchers have suggested that the emotions of the problem solver impact on problem solving strategy while no differences were found in overall performance. Participants with negative emotions more thoroughly explored the task and accessed the information text of the complex Forestry system FSYS 2.0 (Wagener, 2001) more often than their positive affect counterparts (Spering, Wagener, & Funke, 2005). Relationships have also been found between complex problem solving performance and personality and social factors (Dörner & Wearing, 1995). However, the results are inconsistent with other findings in which insignificant to negligible relationships were found (Wagener). The highest correlations were only around .15 between the Forestry task and the personality factors of Extraversion and Neuroticism. Due to a lack of consistency in previous findings, the relationship between complex problem solving performance and noncognitive variables such as personality and interests will be further explored in this thesis.

1.4 Complex Problem Solving: The Environment

Complex problem solving performance involves an interaction of the task, the person, and the environment. There is now consensus amongst researchers in the area that the person (problem solver) and the environment are separate and distinct entities (Frensch & Funke, 1995; Kersting, 1999; Strauß, 1993). Modifications to the environment in complex problem solving research include varying levels of feedback, expectations,
cooperation, peer pressure, disturbances, and individual versus team performance (Frensch & Funke). These types of experiments are of particular interest to both organisational behaviour and cognitive psychologists who are concerned with how manipulations in the environment affect complex problem solving performance.

For example, it has been suggested that different types of feedback impact not only on overall problem solving performance but also the types of exploration strategies that are adopted when participants use hypothesis testing to try to understand the underlying rules of a complex system. Specific feedback usually leads to improved CPS performance (Burns & Vollmeyer, 2002). However, too much feedback can actually lower the amount of rule knowledge that participants gain and consequently hinder subsequent performance. In specific feedback conditions, participants tend to rely on the automatic feedback provided by the system rather than engage in active exploration. Thus, they are discouraged from learning the underlying rules of the complex system for themselves. From experimental manipulation of feedback specificity, the complex problem solving environment was changed and it was concluded that specific feedback could lower exploration and information processing in CPS tasks (Goodman, Wood, & Hendrickx, 2004).

In another example, a team environment was created to investigate group complex problem solving performance. C³Fire (Granlund, 2002) is a complex computer simulated scenario useful for exploring team activities. C³Fire is currently employed in preliminary research for the development of a new command environment for the Swedish armed forces called the ROLF (joint mobile command and control concept) 2010 project. The end result of this project is to create a high-technical user interface
that several commanders will view at the same time while working as a team in the
decision making process (Johansson, Granlund, & Waern, 2005). In preliminary
research based on C³Fire, participants take on the roles of reconnaissance persons, fire
fighting unit chiefs, and various other team members who must cooperate to extinguish
a forest fire. The “cover story” of C³Fire is a forest fire fighting scenario and is
immaterial to the results from which generalisations regarding team performance
environments can be extracted and applied to the ROLF 2010 project (Johansson, et al.),
for military purposes as opposed to fire fighting.

Changes in the environment have an important and significant impact on both the task
and the person (problem solver). The environment can affect persons by changing their
exploration and information processing activities affecting their learning and
consequently their accessible knowledge of the complex system. The environment can
affect the task by restricting the types of tools that can be employed. Overall, complex
problem solving research seeks to understand the interaction between the task, person,
and environment (Frensch & Funke, 1995). However, the affect of the environment on
complex problem solving performance is beyond the scope of this thesis.

1.5 Complex Problem Solving in the Thesis

In summary, the definition of complex problem solving provided has influenced the
choice of tasks and methods used in this thesis and consequently the conclusions drawn
from the results. Three computer simulated scenarios were employed in the present
research: Tailorshop simulation (Putz-Osterloh, 1981); Forestry simulation FSYS 2.0
(Wagener, 2001; Wagener & Conrad, 1996); and Furniture Factory simulation (Wood & Bailey, 1985). All the complex problem solving tasks chosen for the following studies are in accordance with the definition of a complex problem solving task provided in this chapter comprising complexity, connectivity, dynamics, and intransparency. These qualities are representative of the cognitive processing demanded by real world complex problem solving (Dörner, 1989). It is acknowledged that the interplay of the task, person, and environment are important for complex problem solving research. However, task and person aspects will be the focus of the following chapters, which will explore the relationships between complex problem solving performance and specific components of intelligence.
CHAPTER 2
THEORIES OF COMPLEX PROBLEM SOLVING AND INTELLIGENCE

2 Introduction

During the 1930s, simple problem solving tasks were developed in the experimental work of the Gestaltists in Germany, for example, the nine-dot problem (Maier, 1930); the candle task and radiation problem (Duncker, 1945), and the disk problem (Ewert & Lambert, 1932), later known as the Tower of Hanoi (Simon, 1975). Task demands were believed to be the same for both these simple experimental problems and real world problems. There was a widespread assumption that the cognitive abilities required to solve these simple tasks were the same as those required to solve more complex real world problems (Wenke, Frensch, & Funke, 2005). It was not until the 1970s that researchers realised that the theories they had developed based on simple problem solving tasks, could not be generalised to more complex, real world problems. Their previous assumptions were even further challenged by findings that the cognitive processes involved in complex problem solving were different across various domains and levels of expertise (Sternberg, 1995). North American and European researchers responded quite differently to these findings and their subsequent diverse experimental approaches will be described next.
2.1 Methodological Approaches to Complex Problem Solving: North America

In North America in the 1970s, the very same researchers who had attempted to generalise findings from simple problem solving performance to real world problem solving (Simon, 1975) began to investigate the development of complex problem solving in semantically rich domains (Anzai & Simon, 1979; Bhaskar & Simon, 1977). The discovery that the processes underlying problem solving performance were different for separate domains (Sternberg, 1995) lead researchers away from attempting to develop a comprehensive theory of complex problem solving (Sternberg & Frensch, 1991). Their research attention shifted to problem solving in specific, natural knowledge domains (e.g., chess, engineering, physics) and the development of expertise in these areas (Anderson, Boyle, & Reiser, 1985; Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981).

2.1.1 Computer Games and Complex Problem Solving Tasks: A Comparison

Around the same time as research focus shifted to problem solving in different natural knowledge domains, Arnold (1976) published the first psychological article in which computerised tasks were mentioned. This study investigated the effect of feedback, regarding performance, on motivation. Subsequently, the number of psychological papers in which computerised tasks were mentioned continued to climb steadily each year. Psychological research employing computerised tasks consisted of several categories including training and instruction, effect on behaviour, computer use, human factors and programming, experimental, physiological effects, aggression, and
assessment. In 1994, the number of computerised task publications surged to around 70, more than half of these studies were experimental manipulations programmed as game-like tasks, and 8% focussed on assessment (Washburn, 2003).

The popularity of computer games continues to rise at a rapid rate. For example, TOCA touring car (Microsoft, 1986-2004), a driving simulation, has made more profit than all three movies in the Star Wars trilogy combined. Another popular simulation is SIM CITY (Electronic Arts Inc., 2002-2003), where there is no clear goal and the players are free to decide whether they want to build a large city, a beautiful city, or a safe city (Johnson, 2005). Superficially, it is difficult to see how computer games, like TOCA or SIM CITY, differ from experimental tasks such as the air traffic control simulation (Ackerman & Cianciolo, 2000; Kanfer & Ackerman, 1989) or the Lohhausen simulation of a small town mayor (Dörner, Kreuzig, Reither, & Stäudel, 1983).

However, it is important to make a distinction between computer games and experimental manipulations programmed as game-like tasks. Firstly, psychologists refer to their computer games as game-like tasks or complex computer simulations. Secondly, the “players” in computer games are referred to as “participants” in psychological experiments. Thirdly, even though cognitive processes underlie performance on commercial games such as SIM CITY, these types of games do not allow experimental control over the variables and cannot provide the type of data that would be useful for psychological analyses. Therefore, psychologists must program their specific experimental manipulations as computerised tasks at great cost both financially and in the amount of time invested. The design and development of complex computerised tasks for psychological research is certainly a much greater investment.
than the construction of a paper and pencil test or even a simple computerised task (e.g., the Stroop colour word test). Consequently, much of the psychological research into computer task performance has historically been driven and funded by the military and this is still the case to the present day (Washburn, 2003).

During World War II, the US Army employed video games as psychological tests to assess reaction time and coordination. The video games were shown to have high predictive validity for pilot training success. However, the apparatus was too cumbersome, non-portable, and high maintenance (Jones, Dunlap, & Bilodeau, 1986). One of the main aims of this thesis is to explore the relationship between performance on complex computerised scenarios and intelligence. Interestingly, the relationship was explored during the 1980s, where the relationship between performance on computer game tasks and intelligence was investigated. Rabbitt, Banerji, and Szymanski (1989) found a correlation ($r=0.69$) between performance on the computer game SPACE FORTRESS and the AH4 test of general intelligence (Heim, 1968). The strength of this correlation is comparable to that obtained between different tests of intelligence. However, the correlation was obtained only after a substantial amount of practice with the Space Fortress task. Similarly, Jones et al (1986) found moderate positive manifold between arcade games (e.g., Atari’s Indy 500 and Street Racer) and a range of cognitive abilities, with the exception of verbal skills. These results are in contrast to current findings of zero correlation between computer simulation performance and global intelligence (see review by Beckmann & Guthke, 1995). Note, however, that the latter research utilises complex computerised scenarios, whereas the former research employed computer games. Complex computerised scenarios were employed in this
thesis, and performance on them will be compared to performance on specific components of intelligence, rather than global IQ.

Despite their extensive development costs, even in the new millennium, computer simulated tasks offer the advantage of having a common platform (Washburn, 2003). Researchers employing the same computer simulated tasks can compare results across different laboratories and across a wide variety of psychological processes such as attention, learning, or memory. Psychologists from many different countries and from different theoretical backgrounds collaborated in the “Learning Strategies Program” (Donchin, 1995). This group of psychologists (Arthur, Strong, Williamson, Jordan, & Regian, 1995; Shebilske, Goettl, Corrington, & Day, 1999), were able to compare their results regarding training and skill acquisition by employing a common computerised task, Space Fortress (Mané & Donchin, 1989). Similarly, Case (1995) and Gonzalez and Cathcart (1995) employed a single common computerised task to explore operant behaviour and strategy formation respectively. Again, comparison among psychological processes was possible because the researchers were using a common platform. Some researchers, for example, Washburn and Rumbaugh (1992) have even utilised the common platform provided by computerised tasks to compare human performance with that of non-human primates (apes and monkeys). It was found that monkeys could learn on such tasks and that they found these tasks enjoyable. They would choose to play computer games over and above playing with other toys or socialising with other monkeys (Washburn, 2003). Thus, performance on computerised tasks could be compared across different research interests, different communities and cultures, and across species. Simulations could be made more enjoyable by
incorporating game like characteristics such as score keeping, graphics, dynamics, interactivity and competition.

2.1.2 **Computerisation of Traditional Intelligence Tests**

The finding that computerised tasks were perceived as more enjoyable than paper and pencil tasks and could improve motivation, prompted a relatively new line of research, aimed at developing computerised versions of traditional intelligence tests. For example, improvements in learning were found when participants were required to learn the names of fictitious airplanes presented in a dynamic rather than a static format (Romski & Sevcik, 1996). This task is analogous to the “Memory for Names”, long-term memory and retrieval (Glr) subtest of the WJ-R (Woodcock, 1990). Thus, computerised tasks lead to increased retention of the information presented as well as improved motivation and increased performance (Washburn, 2003).

Similar conversions of traditional paper and pencil tests into computerised tasks have also occurred for the assessment of processing speed (Gs). For example, Washburn (2003) administered a continuous performance task where a letter was flashed on screen every 2 seconds for 12 minutes. On 20% of trials the letter H randomly appeared, on all other trials (80%) the letter X appeared. Participants were instructed to use their mouse to click on the letter H as quickly as they could whenever it appeared on screen. An analogous task was also administered where a Star Wars cover story was employed. The H letter was now an H-shaped enemy fighter aircraft and the X shape was now a friendly fighter aircraft. Again participants were instructed to click on the H-shaped
fighter aircraft whenever it appeared, to destroy the enemy. Points were earned by hitting the H’s and points were deducted for false alarms (i.e., hitting the X’s). It was found that responses were more accurate in the task condition than in the game condition (i.e., there were fewer false alarms). However, the response rate was about 12% faster in the game condition than in the continuous performance task version. Results are consistent with the speed-accuracy trade off that would be expected in competitive environments (Washburn & Hopkins, 1989).

Following the American tradition, McPherson and Burns (2005) are currently developing game-like versions of specific cognitive abilities tests, based on the Cattell-Horn-Carroll (CHC) theory of intelligence, which will be described in the next section. Computer game features include sound, competition, motivation, enjoyment, win/lose ending, joystick and trigger control, and a score that is displayed on screen (Mané & Donchin, 1989). In contrast, these qualities are rarely featured in computer simulations designed for the assessment of complex problem solving. In order to create a new measure of processing speed (Gs), McPherson and Burns used a similar computerised game to that developed by Washburn (2003). The Space Code Test employed the cover story of an enemy war ship that must be destroyed by clicking the mouse button. Performance on the newly developed game-like task of processing speed was strongly correlated, around .65, with performance on traditional cognitive abilities tests, e.g., the Digit Symbol subtest of the WAIS-III (Wechsler, 1997) and the Cross Out subtest of the WJ-R (Woodcock & Johnson, 1990). Note, in the area of assessment that employs computerised tasks, it is important to distinguish complex computer simulations designed to assess complex problem solving, for example, the Forestry Task, FSYS 2.0 (Wagener, 2001), from computerised versions of relatively simple traditional cognitive
abilities tests, for example, Space Code Test (McPherson & Burns, 2005) based on CHC theory. Whether they are assessing complex problem solving competency or relatively simple traditional cognitive abilities, all computerised tasks appear to have improved face validity and a better fit with increasingly computerised organisational and educational settings (Kyllonen & Soonmook, 2005).

2.1.3 **Ecological Validity of Computer Simulations**

Ecological validity is defined as the degree to which findings can be generalised to the real world. Field based research allows psychologists to examine complex problem solving in the real world. However, there is a trade off between experimental control and ecological validity in laboratory based research.

In addition to increasing motivation and performance, computer game-like tasks can also increase the ecological validity of psychological research. Results from studies using simulations of real world complex problem solving tasks have diverse real world applications. Driving simulators are employed to increase road safety (Gugerty & Tirre, 2000). Aircraft flight simulators are utilized in studies of air traffic control (ATC, Kanfer & Ackerman, 1989). Thus, it is now assumed that the psychological processes underlying performance on complex tasks, such as computer simulations in the laboratory, can be generalised to real world contexts (Washburn, 2003). Psychologists also conduct research into complex problem solving in the field, in an area called naturalistic decision making (NDM; Klein, Calderwood, & Zsambok, 1993), which will be described in Section 2.3 of this chapter. However, computer simulations offer a way
to simplify the complexity of real world situations so that they can be analysed in a controlled laboratory environment.

2.1.4 Summary: North American Research

In summary, North American researchers in the 1970s shifted their focus from problem solving in simple tasks (Simon, 1975) to complex problem solving in semantically rich natural knowledge domains (Anzai & Simon, 1979; Bhaskar & Simon, 1977). While this research continued through the 1980s, other researchers (Jones, Dunlap, & Bilodeau, 1986; Rabbitt, Banerji, & Szymanski, 1989) were extending military investigations into the relationship between video games and intelligence, with positive results. Psychologists from different countries and subdisciplines were able to collaborate their research findings, because computerised tasks offer a common platform, for example, Space Fortress (Mané & Donchin, 1989). In the area of assessment utilising computerised tasks, several lines of research have emerged. A small number of studies have been conducted in an attempt to develop computerised versions of relatively simple traditional cognitive abilities tests (McPherson & Burns, 2005; Washburn, 2003). However, the aim of this thesis, and the majority of research regarding computerised assessment, employs complex computerised scenarios to explore complex problem solving competencies, for example, air traffic control (ATC, Ackerman & Cianciolo, 2000; Kanfer & Ackerman, 1989). Much of complex problem solving research in North America has been summarized by Sternberg and Frensch (1991). In Europe, complex problem solving research has predominantly focussed on
the development and research of complex computer simulated scenarios. The European approach will be described next.

2.2 Methodological Approaches to Complex Problem Solving: Europe

Until the 1970s, problem solving research in both North America and Europe was historically conducted with relatively simple laboratory tasks. It was realised that the processes underlying performance on these simple tasks could not be generalised to the types of abilities required for solving real life complex problems (Sternberg, 1995). Thus, the emphasis changed from simple problem solving to complex problem solving. Herbert Simon led the response in North America. Subsequent research abandoned the pursuit of a global theory of problem solving and focussed on problem solving in specific natural knowledge domains (Anzai & Simon, 1979; Bhaskar & Simon, 1977).

In contrast to the approach adopted in North America, the response in Europe was rather different and focussed on complex computerised laboratory tasks that simulated real world problems (Funke, 1991). Until this time, development of complex computerised systems for experimental purposes was simply not possible because the technology was unavailable. Broadbent (1977) in Great Britain and Dörner (1975) in Germany initiated the response in Europe, to change the focus of their laboratory research from performance on simple problem solving tasks to performance on complex computer simulated scenarios. The European approach described in this section includes work from Germany (both the former West-Germany and the former East-Germany), Great Britain, Sweden, and Switzerland.
2.2.1 **Dörner School vs. Broadbent School**

Reviews regarding the history of the European approach to complex problem solving research (Wenke, Frensch, and Funke, 2005) frequently cite Buchner (1995) who makes a distinction between the approaches employed by Broadbent (1977) and Dörner (1975). The distinction is partly based on the number of variables underlying the computerised scenarios that both research schools utilised. The Lohhausen system, designed by Dietrich Dörner, simulates a small town, where the participant plays the role of the mayor and is able to manipulate variables such as taxes, working conditions, leisure time activities, etc. The Lohhausen system comprises more than 2000 variables and was beyond total comprehension of participants even after 8 two-hour experimental sessions (Buchner, 1995). Highly complex computer simulations such as Lohhausen are termed ‘Microworlds’ (Dörner, 1993). In contrast, Broadbent suggested that researchers should employ computerised scenarios that were mathematically well-defined, and less complex with fewer variables “to allow an analysis of psychological processes” (p. 192). Buchner classified the different approaches taken by Dörner and Broadbent as two separate research schools, based on their different methodologies. The “Dörner school” characterised by highly complex systems comprising thousands of variables versus the “Broadbent school” characterised by mathematically well-defined systems comprising relatively few variables.

Buchner’s (1995, p. 33) historical account of European complex problem solving research concluded with the following exceptions to the contrasting features of the two research schools:
Of course, the distinction between the two lines of research is not quite as clear-cut as has been portrayed here. For instance, researchers employing naturalistic scenarios do in fact manipulate some features of their tasks – features that do not require any knowledge of the formal system properties such as the semantic context of the system (Hesse, 1982) or the degree to which the system variables’ interrelations are made transparent to subjects (Putz-Osterloh & Luer, 1981).

It is important to clarify that Buchner’s distinction between the two European research schools is not solely based on the respective complex problem solving tasks each employed. For example, the Tailorshop task was originally developed by Dietrich Dörner and first used in a study by Putz-Osterloh (1981). In the Tailorshop task, participants play the role of a shirt company manager and can manipulate variables such as purchasing raw materials, opening sales outlets, and hiring staff. Although the Dörner school also developed the Tailorshop computer simulation, it is rather different from Dörner’s Lohhausen microworld. The Lohhausen system comprises over 2000 variables, which necessitate an exploratory approach. The Tailorshop has only 24 variables which allows an experimental approach. Of the 24 variables, 12 of them are exogenous variables and can be manipulated directly by the participant manager. 12 of them are endogenous variables that are computed by the simulation after the participant manager has decided on interventions. Thus, the Tailorshop task appears more similar to the features of complex tasks described by the “Broadbent” school and is amenable to the experimental approach.
The Tailorshop task is a mathematically well defined system comprising relatively few variables. This is evidenced by the fact that Putz-Osterloh (1981) employed both a transparent and a non transparent version of the Tailorshop. In the transparent condition, participants were provided with a model depicting the underlying causal relations between task variables. In contrast, participants in the intransparent condition were required to infer the underlying mathematical model of the task by using hypothesis testing and learning from the effects of their actions.

In addition to the Tailorshop task, which was designed for individual differences research purposes, the Furniture Factory task is also employed in this thesis. The Furniture Factory was designed for cognitive research and the participant plays the role of the manager of a furniture company, who must motivate employees to work as efficiently as possible. The Furniture Factory consists of 20 exogenous variables that can be manipulated by the participant manager, such as assigning employees to jobs, and 5 endogenous variables that are calculated by the system regarding the actual time simulated employees took to complete their assigned jobs. All together the Furniture Factory task consists of 25 variables. Thus, the Furniture Factory, which was designed in the tradition of the “Broadbent School” of research, has an almost identical number of variables, level of complexity, and simplicity of the underlying mathematical model, as the Tailorshop, which was designed following the “Dörner School” of research.

The above quote from Buchner (1995), cautioned that the distinction between the “Dörner School” and the “Broadbent School” was not clear-cut. The present author agrees, and this is where the distinction between the two schools, based simply on task complexity, must end. Dividing complex problem solving tasks into two groups, based
on the number of variables employed in the task, and the level of simplicity of the underlying mathematical model, depicting causal relationships between the variables, is of little value. It has already been mentioned in Chapter 1 of this thesis that although real life complex problems may have multiple goals (polytelic), this is not a desirable feature of complex problem solving tasks. For example, the Lohhausen simulation (Dörner, Kreuzig, Reither, & Staudel, 1983) has multiple goals and is typically scored by the experimenters choosing the variables of interest to them (Buchner, 1995). However, in later tasks such as the Tailorshop, participants are given a specific performance goal, that is to increase the total assets of the company. Thus, computer simulations with only one goal (not polytelic) provide results from which a dependent measure for participant’s complex problem solving performance can be derived (Buchner, 1995). Consequently, the computer simulations used in this thesis each have a specific performance goal which can be used to assess participants’ performance depending on how closely they reach the goal state. Overall, the Broadbent school and the Dörner school could employ the same tasks (e.g., the Furniture Factory). The difference between the two schools is not based on the tools they use but on their different theoretical approaches. The Broadbent school takes a cognitive approach, for example, the level of feedback provided in the Furniture Factory task could be manipulated to investigate the effect of feedback specificity on complex task performance. In contrast, the Dörner school takes an individual differences approach (e.g., the Tailorshop) could be administered along with an intelligence test battery and the relationship between performance on both tasks could be explored.
2.2.2 Complex Problem Solving and Intelligence

Dörner et al. (1983) used the Lohhausen system to explore the relationship between complex problem solving performance (Komplexes Problemlösen) and intelligence. This task is at the high end of the continuum for intransparency with over 2000 variables and is polytelic with multiple goals. Due to high polytely and subjective scoring techniques, the correlation between Lohhausen task performance and IQ was found to be zero. Similar findings have been reported for other computerised scenarios (for review, see Kluwe et al., 1991; Beckmann & Guthke, 1995). Two hypotheses have been put forward to explain the results: the different cognitive demands hypothesis (Dörner, 1986; Pulz-Osterloh, 1993; Rigas & Brehmer, 1999) and the low reliability hypothesis (Buchner, 1995; Funke, 1992, 1995). The different demands hypothesis argues that traditional IQ tests are ecologically less valid than complex computer simulated scenarios and the two tasks differ on degrees of complexity, connectivity, dynamics, polytely, and intransparency. The low reliability hypothesis argues that complex computer simulated scenarios lack reliability because they have unspecified goals, no optimal solution, and nonlinear relationships between the variables (Rigas, Carling, & Brehmer, 2002). The different demands hypothesis is generally accepted by complex problem solving researchers and has led to widespread discussion about the limitations of traditional intelligence tests (Funke, 2001).

The most widely accepted definition of complex problem solving was based on the theoretical approaches of both Dörner and Broadbent, and was provided by Frensch & Funke (1995, p. 18),
Complex problem solving occurs to overcome barriers between a given state and a desired goal state by means of behavioural and / or cognitive, multi-step activities. The given state, goal state, and barriers between given state and goal state are complex, change dynamically during problem solving, and are intransparent. The exact properties of the given state, goal state, and barriers are unknown to the solver at the outset. Complex problem solving implies the efficient interaction between a solver and the situational requirements of the task, and involves a solver’s cognitive, emotional, personal, and social abilities and knowledge.

Overall, participants acting in complex computer simulated scenarios were faced with many more variables than they were when presented with traditional cognitive abilities tests, including: complexity, connectivity, dynamic environments, intransparency, and sometimes polytely (Brehmer & Dörner, 1993; Brehmer, Leplat, & Rasmussen, 1991).

2.2.3 Explicit and Implicit Complex Problem Solving

Broadbent (1977) suggested that the computer simulations employed to investigate problem solving should not be too complex, in order to allow for the study of cognitive processes. New systems were designed that were amenable to experimental manipulation and control, because there was a well defined, underlying causal model, explaining interconnections amongst the system variables. These systems were still dynamic and complex, albeit less so than Dörner’s Lohhausen. Computer simulations
emerging from the Broadbent school had clearly defined goals and could be used to study cognitive processes (e.g., effect of feedback specificity and effect of time delay) on performance (Buchner & Funke, 1993; Funke, 1986, 1990, 1992; Funke & Buchner, 1992; Hubner, 1987, 1988, 1989; Kluwe, Misiak, & Haider, 1989; Ringelband, Misiak, & Kluwe, 1990; Thalmaier, 1979).

Broadbent’s research on different memory systems took an experimental approach and employed relatively simple computer simulated scenarios such as a simple transportation system or a simple sugar factory. These simulations had less than 10 variables and were used in studies of explicit and implicit problem solving. Explicit problem solving depends on the intended actions of the problem solver where connections between the variables are consciously learned and can be articulated by the participant. Explicit problem solving can be measured in post task questionnaires. In contrast, implicit problem solving occurs outside the realm of intention of the problem solver. Knowledge of the system is unconscious and cannot be articulated. However, implicit problem solving can be measured by performance in the complex computer simulated task (Berry & Broadbent, 1984; 1988; Broadbent, 1977; Broadbent & Aston, 1978; Broadbent, Fitzgerald, & Broadbent 1986; Hayes & Broadbent, 1988).

2.2.4 Theoretical Approach of the Thesis

In this thesis, the European tradition is adopted and complex computer simulated scenarios are employed to assess complex problem solving performance. Much of the European work has been summarised by Frensch & Funke (1995). So, just one decade
has passed since the European research (mostly conducted in Germany) in complex problem solving was first conveyed to English speaking researchers. In this volume, a chapter by Buchner (1995) identified the different theoretical approaches adopted by individual differences researchers, such as Dörner, and cognitive researchers, such as Broadbent. The distinction between the 2 types of complex problem solving research is not based on different task properties (i.e., complexity, polytely, etc.), but on different theoretical approaches. The Dörner school investigated relationships between complex problem solving and intelligence. The Broadbent school studied implicit learning in complex problem solving tasks. Common computer simulated scenarios, for example, the Furniture Factory (Wood & Bailey, 1985) could be employed for either of these approaches. All computer simulated tasks used in this thesis have moderate complexity, comprising around 20 to 30 interconnected variables and specified goals so that independent measures of complex problem solving performance can be extracted. The Furniture Factory task also features a review questionnaire and allows extraction of implicit and explicit learning data.

In the first part of this thesis, the theoretical approach initiated by Dörner will be adopted when investigating the relationships between complex problem solving task performance and specific components of intelligence. An exploratory approach was necessary for Dörner’s Lohhausen project due to the large number of variables that created a task too complex for an underlying causal mathematical model to be inferred by participants. In contrast, this thesis will employ complex problem solving tasks with well-defined underlying mathematical models describing the causal relationships between the relatively few interconnected variables. However, in the absence of prior research using specific cognitive abilities, hypotheses on what dimensions of ability are
expected to correlate with what aspects of performance on the simulation tasks cannot be made at this stage.

In the second part of this thesis, the theoretical approach initiated by Broadbent will be adopted. A distinction will be made between implicit and explicit problem solving. Implicit problem solving is characterised by passive learning of the rules of the complex system, which can improve performance but cannot be articulated. Whereas explicit problem solving is characterised by active learning (i.e., hypothesis testing of the complex system) which can also improve performance and can be verbalised. The relationship between implicit and explicit knowledge of the rules underlying the complex problem solving system and specific components of intelligence will be explored.

2.3 Naturalistic Decision Making

There are two types of research conducted in the area of complex problem solving, one is laboratory based research and the other is field research. Both approaches have their advantages and disadvantages. Laboratory research must strive for ecological validity so that results can be generalised to the real world (Funke, 2001). However, as evidenced by the history of complex problem solving research in the previous section, many assumptions of ecological validity (e.g., simple problem solving tasks) have later been refuted. Conversely, while field research, by its very definition has ecological validity, it is often criticised for being too complex and lacking experimental control. Thus, field researchers struggle to identify causal relationships in their interpretations.
Applied psychologists must bear in mind the limitations of both laboratory research and field research when they utilise the knowledge provided by research psychologists (Brehmer & Dörner, 1993).

2.3.1 Field Research vs. Laboratory Research

The dilemma that research psychologists face when designing computer simulations to assess complex problem solving is neatly described by Brehmer (2005, p. 77), ‘the best simulation of a cat is another cat’. In other words, laboratory research does not strive to perfectly simulate real world complex problems, because the computer simulations would be equally complex and intransparent as those problems observed in the field. In order to learn about complex problem solving behaviour, real world problems must be simplified, the difficult decision for laboratory research is what aspects of real world problems to include or exclude from the complex computer simulation. The most important characteristics to include in experiments assessing complex problem solving are an interaction between the problem solver and environment (Frensch & Funke, 1995). This interaction or circular relation can be observed only in dynamic situations whether they occur in the field or in the laboratory (Brehmer, 2005). The four features that characterise a dynamic situation according to Brehmer and Allard (1991) are that ‘it requires a series of decisions; the decisions are not independent; the state of the problem changes, both autonomously and as a consequence of the decision-maker’s actions; and the decisions must be made in real time’.
2.3.2 Naturalistic Decision Making vs. Complex Problem Solving

In contrast to complex problem solving research that is conducted in the laboratory, naturalistic decision making (NDM) research is conducted in the field. The first conference of naturalistic decision making took place about ten years ago (Klein, Calderwood, & Zsambok, 1993). So, it is a relatively new approach to research compared to complex problem solving research that emerged around 30 years ago (Wenke, Frensch, & Funke, 2005).

Recall from Chapter one, that the five features of complex problem solving in laboratory settings according to (Brehmer & Dörner, 1993; Dörner, 1980; Dörner & Kreuzig, 1983) are: (a) complexity, (b) connectivity, (c) dynamics, (d) intransparency (opaqueness), and (e) polytely.

In contrast, the eight features of naturalistic decision making, in field settings, according to Orasanu and Connolly (1993) and Zsambok (1997, p. 4) are: (a) Ill-structured problems, (b) Uncertain dynamic environments, (c) Shifting, ill-defined, or competing goals, (d) Action/feedback loops, (e) Time stress, (f) High stakes, (g) Multiple players, and (h) Organizational goals and norms.

The five feature taxonomy of complex problem solving features has become standard, in some form or another, in descriptions of complex problem solving research. Similarly, the eight feature taxonomy presented above is widely accepted as the core of the definition of naturalistic decision making research (Lipshitz, Klein, Orasanu, & Salas, 2001). By contrasting the two lists above characterising lab research and field
research respectively, it is apparent that complex problem solving studies have many defining features in common with naturalistic decision making studies. Complex problem solving and naturalistic decision making have in common their focus on real life tasks rather than artificial tasks. The employment of complex dynamic computer simulations has greatly improved the face validity and degree of fidelity of laboratory based research (Funke, 2001). Real life situations also consist of the five factors that define laboratory based complex problem solving tasks. Both can consist of complexity, action/feedback loops (connectivity), uncertain dynamic environments, ill-structured problems (intransparency/opaqueness), single or multiple goals (polytely), and time stress. Thus, the complex computerised scenarios employed in laboratory research can show many of the typical features of naturalistic decision making and can be used in naturalistic decision making research.

The differences between complex problem solving research and naturalistic decision making are relatively minor (Funke, 2001). The eight factor taxonomy of naturalistic decision making can be used to identify what is missing from traditional experimental research in decision making (Cohen, 1993). Complex problem solving research is directed to novices and the interplay of cognitive, motivational, and social components (Dörner, Kreuzig, Reither, & Stäudel, 1983). For example, a recent laboratory based study has investigated the role of emotions in action regulation for performance of the complex problem solving task FSYS 2.0 (Spering, Wagener, & Funke, 2005). From a complex problem solving perspective, decision making is one of many processes for action regulation. In contrast, naturalistic decision making research generally focuses on experts operating within their domains of expertise (Funke). However, this is not always the case and naturalistic decision making research can be conducted on novices
Overall, the main distinction is that complex problem solving research is conducted by observing performance on complex computer simulations, whereas naturalistic decision making is conducted by observing decision making behaviour in the field (Funke).

In a recent volume on naturalistic decision making by Montgomery, Lipshitz, and Brehmer (2005), several chapters present natural decision making research that have employed computer simulated scenarios identical to those used in complex problem solving research. For example, Networked Fire Chief, depicting forest fire environment (Omodei, Taranto, & Wearing, 1999), and Stocks and Flows, depicting the bathtub and cashflow task (Forrester, 1961; Sterman, 2000) were employed to explore individual decision making; whilst C³Fire, depicting team decision making (Granlund, 2002), was employed to explore team decision making.

Basically, the closer a computer simulation approximates real world problems, the more relevant it is for employment in naturalistic decision making studies (Montgomery, Lipshitz, & Brehmer, 2005). For example, Sterman and Sweeney (2005) describe the difficulties that people have in understanding action/feedback loops, particularly under delayed feedback conditions (time stress). Note that action/feedback loops are qualities of real world problems, they are identified as the fourth feature of Orasanu and Connolly’s (1993) taxonomy of naturalistic decision making, and they are also qualities of complex computer simulations defined by the second and third features of Dörner’s (1980) complex problem solving taxonomy. The naturalistic decision making approach is explored further in a study described in the next chapter, in which a stocks and flows task similar to that employed by Sterman and Sweeney (2005) was administered along
with a complex computer simulation and a battery of cognitive abilities tests. The relationships between the naturalistic decision making task, the complex problem solving task, and specific components of intelligence were explored.

2.3.3 Summary: Naturalistic Decision Making Research

In conclusion, at the extremes of both types of research, laboratory research and field research portray the problem solver very differently. In laboratory research, the problem solver’s behaviour is controlled by the environment, which can be manipulated experimentally. In field research, the problem solver’s behaviour is self regulated. Thus, the problem solver in the laboratory is observed rather differently to the problem solver in the field. Laboratory experiments tend to simplify real world problems, consequently some of the complexity of the real world situation is lost. The reduced complexity in laboratory research means that the behaviour observed is not self regulatory, as it is in field studies (Brehmer & Dörner, 1993). The advent of computers in the laboratory in the early 1970s allowed some of the complexity of the field to be simulated in the laboratory. Complex computer simulated scenarios do not reduce real world situations to the level of simplicity of classical psychology experiments; they are described as using ‘condensation’ rather than ‘reduction’ (Dörner, 1992, cited in Funke, 2001). Thus, computer simulations are a tool that can act as a bridge between laboratory research and field research.

In order to explore the relationships between problem solving competencies and intelligence, a number of criteria must be met. Firstly, complex problem solving and
intelligence must be defined and cannot overlap at theoretical or operational levels. Secondly, there must be a theoretical basis for the relationship to determine the direction of causality. Thirdly, the direction of causality must be supported by empirical evidence that may be provided by longitudinal research, manipulation of intelligence test instructions or task properties, or by control of a third variable that is confounding the relationship (Wenke, Frensch, & Funke, 2005). In the previous sections of this thesis, the definition of complex problem solving and its historical roots have been provided. The next two sections will provide a definition of intelligence and the history of intelligence testing in order to meet the first of the three criteria suggested by Wenke et al.
CHAPTER 3

Hierarchical theories of intelligence

Spearman (1904, 1927) proposed a unitary theory of a single process of intelligence termed Spearman’s $g$, or general intelligence, after noting that virtually all correlations among different clusters of cognitive abilities tests are positive (positive manifold). Spearman also developed factor analysis in order to interpret large correlational matrices. Thurstone (1931, 1938a) developed a method of factor analysis that yielded groupings of variables. Thus, the existence of both general and group factors was established. However, Thurstone noted that the general factor extracted from one test battery is not the same as the $g$ extracted from another battery.

Thurstone’s (1938b) evidence for primary mental abilities (PMAs) instigated a flurry of research that rapidly increased the number of PMAs and created the need for a model of intelligence to organise these group factors. Faceted theories (Guilford, 1967) and hierarchical models (Burt, 1949; Cattell, 1941; Cattell, 1943; Cattell & Horn, 1978; Vernon, 1950) emerged. The hierarchical model became the most widely recognised model of the structure of intelligence. It is based on factor analysis of large batteries of tests in order to define broad abilities (Cattell, 1971; Cattell, 1987; Cattell & Horn, 1978; Horn, 1980, 1988; Horn & Stankov, 1982). The factorial structure revealed by factor analysis will depend on the nature of the tests in the battery. Cattell’s approach used a large number of carefully chosen tests that yielded a number of second-order factors including the two most important mental ability factors, fluid ability (Gf) corresponding to abstract reasoning and crystallised ability (Gc) corresponding to
acculturated knowledge. Thus, the theory of fluid and crystallised intelligence (Gf-Gc theory) was established.

‘Evidence of studies…came forward again and again to indicate that a single common-factor theory was not adequate to explain the diversity of human intellectual capabilities’ (Horn, 1998, p. 58), leading to the current theory of several intelligences. There are two schools of thought regarding $g$ amongst those working within Gf-Gc framework, one considers $g$ a meaningless conglomerate of narrower abilities and is termed the truncated hierarchical theory (see Stankov, 2005). The other school accepts the existence of $g$ but research is mainly focussed on second-order factors (Stankov, 2002) that provide specific information about individuals’ strengths and weaknesses in intelligence rather than simple overall IQ (i.e., $g$).

In addition to Gf-Gc theory, intelligence researchers today propose alternate theories of the structure of intelligence. Some continue to argue for the existence of Spearman’s $g$ (Deary, 2002; Gottfredson, 1997, 1998, 2002; Jensen, 1998, 2002). A ‘monarchic’ general factor can be derived from the majority of assessment tools currently used (Gottfredson, 1997), and is analogous to an Intelligence Quotient. Other researchers have constructed ‘oligarchic’ abstract theories of multiple intelligences including musical, spatial, bodily-kinaesthetic and personal intelligence, (Gardner, 1983, 1998), creativity and practical intelligence (Sternberg, 1988, 1998, 2002). These theories are popular with the wider community and provided the foundation for growing areas of research such as emotional intelligence (Salovey & Mayer, 1989-1990). However, the main goal of the scientific study of human intelligence remains measurement (Mackintosh, 1998) and is grounded in Gf-Gc theory.
Carroll’s (1993) hierarchical theory consists of three levels of strata and is an integrative model based on 460 data sets found in the factor-analytic literature. Carroll’s meta-analysis employed exploratory techniques which are described as ‘more suitable than confirmatory techniques for initially identifying cognitive abilities and their structure (Carroll, 1993, p. viii)’. Hence it is named the Cattell-Horn-Carroll (CHC) theory of cognitive abilities. This model is a revised version of hierarchical theory with the $g$ factor, the first principal component, at the third-stratum or apex. There are 10 second-stratum factors immediately below $g$ (Carroll, 2003), roughly corresponding to the broad abilities in Gf-Gc theory, including Fluid Reasoning, Crystallised Intelligence, Short-Term Memory, Visual Processing, Auditory Processing, Long-Term Retrieval, Processing Speed, Reading and Writing Ability, Quantitative Knowledge, and Reaction Time/Decision Speed. Each of these broad abilities can be further classified into 70 narrower cognitive abilities at the first-stratum (McGrew, 2005) extended from Thurstone’s original PMAs.

The theory of fluid and crystallised intelligence (Gf-Gc theory) is recognised as the most empirically supported and theoretically sound description of the structure of human cognitive abilities (Ackerman & Heggestad, 1997; Carroll, 1993; Cattell & Horn, 1978; Flanagan, 2000; Gustafsson, 1984; Horn & Stankov, 1982; Snow, 1984; Stankov, 2000; Woodcock, 1990) and a more detailed account will be provided in the next section. Regardless of whether researchers accept the existence of $g$, proponents of Gf-Gc theory suggest that most or all of the broad abilities are important and should be the main focus of our theories about the structure of human cognitive abilities. However, the broad abilities with higher g-loadings (e.g., Gf) are deemed to be more
important than broad abilities with lower g-loadings (e.g., Ga). Intelligence tests employed in the present thesis, exploring the relationships between cognitive abilities and microworlds, reflect this view.

3.1 The Theory of Fluid and Crystallised Intelligence: A Review

The following sections will describe the rationale for adopting a particular theory for understanding intelligence when investigating factors that may affect microworld performance. Psychologists and researchers regard the second-stratum factors in the Cattell-Horn-Carroll (CHC) three-strata model, or the truncated two-strata model, as the essence of intelligence (Horn, 1998). The present section will give a brief review of the emergence of nine second-stratum factors of intelligence within the theory of fluid and crystallised intelligence (Gf-Gc theory), these factors will be defined, their guidance of the development of prominent test batteries will be described, and empirical evidence in support of separate indicators of intelligence will be presented.

All of the broad abilities in Gf-Gc theory are thought to be structurally equivalent. However, as a result of historical precedent, Gf and Gc have been the main focus of research. Gf tests are relatively, though not absolutely, ‘culture fair’ (Carroll, 1993; Mackintosh, 1998; Sternberg, 1989). Gf performance depends to a much smaller extent on formal education and acculturation than Gc performance (Cattell, 1987; Gustafsson, 1988; Horn & Stankov, 1982; Lohman, Pellegrino, Alderton, & Regian, 1987). The amount of acculturated knowledge required can be varied with respect to test content or operations required during the test. Gc depends on long-term memory store and organisation of information within that store whereas Gf depends on working memory
capacity (Kyllonen & Christal, 1990; Wittmann & Süß, 1999). Although it has been suggested that Gf is more genetically heritable than Gc, the evidence is controversial and inconclusive (Horn, 1998; Horn & Noll, 1993).

In addition to Gf and Gc, large-scale factor analytic studies verify the presence of several other broad second-order factors, including visual processing (Gv), auditory processing (Ga), short-term apprehension-retention (SAR), tertiary storage and retrieval (TSR), and processing speed (Gs) (Carroll, 1993; Fogarty & Stankov, 1988; Gustafsson, 1984; Lohman, Pellegrino, Alderton, & Regian, 1987; Stankov, 1988; Woodcock, 1990). The model has been elaborated with the proposal of two additional broad factors. These are quantitative knowledge (Gq: Horn, 1988) and correct decision speed (CDS: Horn & Hofer, 1992). Future research in Gf-Gc theory is likely to include the addition of new factors, for example, tactile-kinaesthetic (Roberts, Stankov, Pallier, & Dolph, 1997) and olfaction (Danthiir, Roberts, Pallier, & Stankov, 2001).

Definitions of the nine broad cognitive abilities according to (Horn & Noll, 1993) p. 173 are:

1. **Fluid reasoning (Gf)**, measured in tasks requiring inductive, deductive, conjunctive, and disjunctive reasoning to arrive at understanding relations among stimuli, comprehend implications, and draw inferences.

2. **Acculturation knowledge (Gc)**, measured in tasks indicating breadth and depth of the knowledge of the dominant culture.

3. **Quantitative knowledge (Gq)**, measured in tasks requiring understanding and application of the concepts and skills of mathematics.
4. Short-term apprehension-retention (SAR) also called short-term memory (Gsm), measured in a variety of tasks that mainly require one to maintain awareness of, and be able to recall, elements of immediate stimulation; that is, events of the last minute or so.

5. Fluency of Retrieval from Long-term Storage (TSR), also called long-term memory (Glr), measured in tasks that indicate consolidation for storage and mainly require retrieval, through association, of information stored minutes, hours, weeks, and years before.

6. Visual processing (Gv), measured in tasks involving visual closure and constancy, and fluency in “imaging” the way objects appear in space as they are rotated and flip-flopped in various ways.

7. Auditory processing (Ga), measured in tasks that involve perception of sound patterns under distraction or distortion, maintaining awareness of order and rhythm among sounds, and comprehending elements of groups of sounds, such as chords and the relations among such groups.

8. Processing speed (Gs), although involved in almost all intellectual tasks (Hertzog, 1989), measured most purely in rapid scanning and responding in intellectually simple tasks (in which almost all people would get the right answer if the task were not highly speeded).

9. Correct Decision speed (CDS), measured in quickness in providing answers in tasks that require one to think.

Although not always explicitly stated, these broad factors have played a prominent role in the development and interpretation of IQ tests and neuropsychological batteries (Alfonso, Flanagan, & Radwan, 2005). Tests based on Gf-Gc theory include the

There is a wide body of evidence in support of multiple and distinct indicators of intelligence (Carroll, 1993; Horn, 1998). This evidence supports the premise that the operationalisation of intelligence in Gf-Gc theory, described by the nine second-stratum factors listed above, is a true reflection of the construct of intelligence.

Structural evidence is based on the factor analysis of many studies of cognitive abilities batteries. Nine factors are described in terms of Cartesian coordinates that are rotated to simple structure. No set of eight factors can fully predict the variance of the ninth factor (Horn, 1998; Horn & Noll, 1993). This evidence of individual differences covariation can also be described in terms of a circular Radex model (Snow, Kyllonen, & Marshalek, 1984). The Radex model is very similar to Cattell’s hierarchical model although different techniques were employed. In the Radex model, multi-dimensional scaling was employed to illustrate the relationships between tests. Correlations between tests are represented by the distance between them in 2-dimensional space within the circular radex. Tests that are closer together are more highly correlated with each other than tests that are further apart. Tests closer to the centre of the circle are more highly correlated with the general factor of intelligence so that tests of fluid intelligence (e.g.,
Raven’s Progressive Matrices) are closer to the centre, while elementary cognitive tasks (e.g., reaction time tasks, perceptual/clerical speed, and forward digit span) are closer to the periphery (Snow et al., 1984). Thus, high complexity tasks are those towards the centre of the radex diagrams while low complexity tasks are towards the periphery. An important distinction must: between the use of “complexity” here and the use of “complexity” in Section 1.2.7 when discussed in relation to performances on computer simulations. The use of “complexity” in hierarchical theories of intelligence is used to distinguish between high and low g-loading tasks, while the use of “complexity” in simulations regards typical intelligence tests as low complexity tasks compared to more complex real-world situations (or approximations to them, as in simulations).

Developmental evidence indicates that separate cognitive abilities have different relationships to age. There are declines in the abilities of Gf, SAR, and Gs with age (wasting and deterioration of the brain), whereas Gc and TSR (learning and consolidation) remain steady or increase up to 60-65 years of age (they then decline a little, but not to the extent of decline in Gf, SAR, and Gs) (Horn, 1998; Horn & Noll, 1993). Neurocognitive evidence is remarkably similar to developmental evidence. Although there are declines in the abilities of Gc and TSR immediately following brain damage (stroke, surgery, blows to the head, brain tumors), these declines are largely reversed and abilities return to preinjury levels during the recovery period. In contrast, Gf, SAR, and Gs remain incapacitated (Horn; Horn & Noll).

Achievement evidence indicates that a conglomerate of carefully chosen and specific measures is the best predictor of academic and job performance (Campbell, 1990; Schmidt, Ones, & Hunter, 1992). For example, a test to predict occupational
achievement in electrical engineering would require inclusion of visualisation ability, numerical skill, and verbal comprehension (Horn, 1998). The best predictors of educational achievement are conglomerates of measures that indicate Gc, TSR, and Gq. In contrast, Gf, SAR, and Gs have much weaker predictive validity (Horn).

Gf-Gc theory was considered appropriate to guide the selection of cognitive abilities tests in the studies that follow. Predominantly, IQ or a general factor of intelligence has been employed in research exploring the relationships between intelligence and microworld performance (Dörner, Kreuzig, Reither, & Stäudel, 1983; Putz-Osterloh & Lüer, 1981; Strohschneider, 1986, 1991). Participants’ patterns of strengths and weaknesses in cognitive abilities can be derived from second-stratum broad abilities, not from a general factor. The more specific information provided by broad abilities allows a greater understanding of relationships between intelligence and microworlds. However, it is beyond the scope of this thesis to analyse relationships between microworlds and all nine broad abilities. Working memory capacity assessed in Gf tasks is believed to be more important than short-term memory (SAR) or long-term memory (TSR) in complex problem solving required in microworlds (Wittmann & Süß, 1999). Ga and CDS are not relevant in the studies that follow since the microworlds used do not contain an auditory component and complex decision making during microworld performance was not timed. Thus, the present thesis will investigate the relationships between performance in microworlds and five broad abilities comprising Gf, Gc, Gv, Gq, and Gs.
3.2 Complex Computer Simulated Scenarios: Some Examples

Dörner, Kreuzig, Reither, and Stäudel (1983) initially employed highly complex microworlds, to explore the relationships between complex problem solving and intelligence (e.g., the Lohhausen project). Later studies used less complex systems with well defined causal connections between the variables and clearer scoring protocols (e.g., Tailorshop). Berry and Broadbent (1984) used similar, yet generally less complex and more well defined computer simulations to investigate implicit learning in complex problem solving situations (e.g., a simple transportation system or a simple sugar factory). Ackerman and Cianciolo (2000) and Ackerman and Kanfer (1993) utilise complex computer simulations for their work on skill acquisition (e.g., air traffic control task). These simulations are characterised by relatively short term, tactical decision making under time pressure rather than long term, strategic decision making (Wagener, 2001). They also require little self-directed exploration to gather information. In this way they differ from complex problem solving tasks employed by Dörner et al. (1983) or Funke (1995). Thus, both cognitive and individual differences psychologists may employ simulations while seeking to understand behaviour in dynamic systems. Issues from the cognitive perspective (e.g., learning, goal specificity, and feedback) are beyond the scope of this thesis.

Computer simulations are also utilised by organisations for assessment and recruitment purposes. For example, McKinsey & Company recently employed computer simulation tools based on dynamic game theory to provide sophisticated support for negotiators (McKinsey, 2005). In addition, computer simulations may be designed for military
purposes. For example, America’s Army (AA) is a game that provides a virtual web-based environment where participants can explore army career options (America's Army, 2005). Simulations designed for recruitment or military purposes are beyond the scope of this thesis.

The present thesis takes an individual differences perspective. A summary of individual differences research exploring the relationship between complex problem solving and intelligence is provided in Appendix A. The list is organised by the simulations that were employed. Overall, the findings are inconsistent and often contradictory. Thus, highlighting the need for a comprehensive multivariate study exploring the relationship between complex problem solving and specific aspects of intelligence, which is the main aim of this study.

3.2.1 Summary: Complex Computer Simulated Scenarios

Many computer simulations and their associated publications are derived from German research. Thus, many of the studies investigating the relationship between complex problem solving competencies and intelligence have employed the Berlin Intelligence Structure Test (BIS: Jäger, Süß, & Beauducel, 1997). In contrast, this thesis employs cognitive abilities tests derived from the theory of fluid and crystallised intelligence (Horn, 1998) and explores the relationship between specific components of intelligence and complex problem solving competencies. Complex problem solving performance is assessed by computer simulations. This thesis employs two simulations that were designed in Germany: FSYS 2.0, Forestry Task (Wagener, 2001) was translated from
German to English for this thesis; and Tailorshop (Putz-Osterloh, 1981) was previously translated from German to English. Both FSYS 2.0 and Tailorshop were designed for individual differences research purposes. The third computer simulation employed in this thesis is the Furniture Factory (Wood & Bailey, 1985); this task was designed at an Australian management school for cognitive research purposes. All three computer simulations employed in this thesis are appropriate for complex problem solving research because they display the qualities required for complex problem solving tasks described in Section 1.2 of this thesis.

3.3 Overview of the Thesis

Computer simulations of real world environments are claimed to possess ecological validity because they approximate the complexity of real life situations more closely than traditional intelligence tests (Rigas, Carling, & Brehmer, 2002). While traditional intelligence tests strive to be ‘pure’ measures of a specific cognitive ability, micro-worlds are ‘impure’ tests of a conglomerate of psychological behaviours including complex problem solving, decision-making, and emotional functions (Brehmer, 2005). Computer simulations also increase interest and motivation and allow more convenient and adaptive testing than traditional intelligence tests (Kröner Plass Leutner, 2005; Washburn, 2003). Thus, simulations are promising new psychometric tools for training and assessment purposes and a proliferation of research is currently underway for their increased use in educational, corporate, and military settings.
Research with complex computer simulated tasks has involved both the experimental cognitive approach and the individual differences approach with each methodology addressing a different research question (Brehmer, 2005; Buchner, 1995). In the experimental cognitive approach, performance in different conditions of computer simulations is compared. For example, one group may be assigned to a micro-world with highly specific feedback while another group is assigned to a micro-world with limited feedback (Goodman & Wood, 2004). If participants in the former group outperform participants in the latter, this suggests that more specific feedback positively affects performance. In the individual differences approach, micro-world performance is explained by intelligence (Dörner, 1996). For example, participants are administered a micro-world task and a fluid intelligence test such as Raven’s Advanced Progressive Matrices (Rigas, Carling, & Brehmer, 2002). If there is a positive correlation between micro-world performance and fluid intelligence (Gf), then it is suggested that high Gf is important for micro-world performance.

Research following the experimental approach has provided stable results. However, research following the individual differences approach has produced inconsistent and often nonsignificant correlations between micro-world performance and intelligence (Rigas, Carling, & Brehmer, 2002). Attempts have been made to explain this lack of a relationship in terms of the different demands hypothesis (Rigas & Brehmer, 1999) and the low reliability hypothesis (Rigas et al.). The issue is further confused by more recent, albeit relatively few, findings of stable and positive correlations between micro-world performance and cognitive abilities (Gonzalez, Thomas, & Vanyukov, 2005; Kröner, Plass, & Leutner, 2005).
There are structural differences in computer simulated tasks with respect to the common characteristics they share such as complexity, connectivity, dynamics, intransparency (opaqueness), and polytely (Brehmer & Dörner, 1993). Complexity refers to the number of variables that must be controlled simultaneously by the user while making a decision, including possible trade offs between goals and side effects. Connectivity refers to the causal connections between the variables. Dynamics refers to the decision state of the user as a consequence of both their decisions and the environmental changes that are beyond their control. Intransparency (opaqueness) is defined as the characteristics of the micro-world that must be identified if hidden, or inferred if there are information delays that make present information obsolete. Polytely refers to the pursuit of multiple goals (Brehmer & Dörner). However, a single goal is more desirable for interpretation of results (Süß, Oberauer, & Wittmann, 2004). Variance in these task characteristics may lead to problems in defining what constitutes a complex problem solving task. On the other hand, these commonalities provide a framework for making generalisations regarding the similarities and differences between complex problem solving studies.

According to Brehmer (2005) most micro-world research is essentially investigating learning. This raises the question about what constitutes optimal performance. However, the vast majority of previous individual differences studies have used an overall score as a measure of micro-world performance. More recently, Kröner, Plass, and Leutner (2005) have suggested that micro-world performance consists of three components: rule identification, rule knowledge, and rule application. Scores on these measures can be combined with micro-world performance to give new scoring
techniques. Furthermore, Brehmer suggests focusing on ‘reasonable’ rather than ‘maximal’ performance.

In addition to scoring problems, research exploring the relationships between intelligence and micro-world performance has predominantly employed an IQ measure. This general intelligence score offers less specific information than that provided by broad cognitive abilities such as those within the theory of fluid and crystallised intelligence (Gf-Gc theory: Carroll, 1993; Horn & Noll, 1993; Stankov, 2000). Gf-Gc theory guided the selection of cognitive abilities tests in the present thesis. Relationships between broad abilities including Fluid reasoning (Gf), Acculturation knowledge (Gc), Visual processing (Gv), Quantitative knowledge (Gq), Processing speed (Gs), and three micro-worlds were explored.

Revised versions of the following three micro-worlds were employed in the present thesis:

1. The Furniture Factory management simulation (Berry & Broadbent, 1995; Goodman & Wood, 2004; Wood, Bandura, & Bailey, 1990; Wood, Atkins, & Bright, 1999; Wood & Bailey, 1985): Players manage a furniture factory and are responsible for allocating employees to jobs and motivating them to complete a series of weekly furniture orders in as short a time as possible.

2. The Forest task, FSYS 2.0 (Wagener & Conrad, 1996) was revised and translated from German into English for this thesis. Players manage a simulated company that plants, grows, and lumbers trees in its forests while controlling pests and soil conditions.
3. Tailorshop (Funke, 1991; Putz-Osterloh, 1981; Wittmann & Süß, 1999): In order to maximize the overall profit of the company, players manage a shirt factory by manipulating numbers such as machines, outlets, sales representatives.

Study 1 of this thesis explored the relationship of intelligence with both a computer simulated task, the Furniture Factory (Wood & Bailey, 1985), and a stocks and flows task, the Dynamic Forecasting Questionnaire (DFQ: Atkins et al., 2001) derived from Naturalistic Decision Making (NDM research). The effect of cover stories, gender, and individual differences in exploration strategies on complex problem solving performance was investigated.

Study 2 of this thesis explored the relationship of complex problem solving with intelligence, personality, interests, and biodata. Firstly, scoring issues in complex problem solving research were investigated. Secondly, the relationship of complex problem solving with both broad intelligence factors and specific components of intelligence was explored. Thirdly, the relationship between complex problem solving and non-cognitive variables (personality and vocational interests) was investigated. Finally, future directions in the area of complex problem solving research were identified.

Brehmer (2005) suggests that in order to understand what psychological demands micro-worlds actually make, research must concentrate on a limited number of micro-world tasks. This thesis will investigate and document the properties of three micro-worlds in order to find commonalities and differences among them, such as structural characteristics including complexity, connectivity, dynamics, and opaqueness (Brehmer
& Dörner, 1993). Relationships between micro-worlds and cognitive abilities will be further examined. Personality (5 factor model: Costa & McCrae, 1992), gender, and biodata will also be included in the analyses. Individual differences research in this area has led to inconsistent results that are difficult to interpret. Researchers must engage in a cumulative effort to understand the characteristics of computer simulations as research tools. This thesis will attempt to bring individual differences research in the area a step closer to obtaining stable results from which generalisations about complex problem solving tasks can be made.
CHAPTER 4
SPECIFIC COMPONENTS OF INTELLIGENCE AS PREDICTORS OF
COMPLEX PROBLEM SOLVING: STUDY 1

4.1 Introduction

The suggestion that ‘there is no significant correlation between scores on IQ tests and
performance in any complicated problem solving experiment’ (Dörner, 1996, p. 27),
prompted a huge movement in complex problem solving research with computer
simulations to find an explanation. The zero correlation finding was not consistent with
the suggestion that intelligence (assessed by Raven’s Advanced Progressive Matrices) is
the most important ability for real world complex problem solving (Gottfredson, 1997).
Researchers in the area were divided by proposals of two different explanations: The
different cognitive demands hypothesis (Dörner, 1983; Putz-Osterloh, 1993; Rigas &
Brehmer, 1999) and the low reliability hypothesis (Buchner, 1995; Funke, 1992, 1995).

4.1.1 Different demands hypothesis

In its strongest formulation, the different demands hypothesis predicts close to zero
correlations between traditional intelligence test scores (e.g., Raven’s Progressive
Matrices) and performance on complex computer simulated scenarios (Putz-Osterloh,
1993). The extreme view is that all traditional intelligence tests should be replaced by
computer simulations (Kröner, Plass, & Leutner, 2005). Complex computerised
scenarios are suggested to have higher ecological validity than traditional intelligence
tests. Computer simulations are characterised by complexity and connectivity (the number of variables within the system and the causal connections between them), Intransparency (the underlying rules of the complex system are based on mathematical algorithms, this underlying model is not visible to participants in the intransparent condition and certain aspects must be inferred), dynamics (the degree of change within the system that is computer automated and not under direct control of the participant), and polytely (the presence of multiple goals which can be contradictory and may necessitate a trade off in decision making). These five qualities of complex computer simulated scenarios are described in detail in Sections 1.2.1 to 1.2.5 of this thesis.

Computer simulations and traditional intelligence tests differ in respect to several task qualities. Traditional intelligence tests provide all the necessary information to the participant at the beginning of the test that is needed to solve the problem. In contrast, computer simulations provide only a portion of the required information to solve the problem at the beginning of the task. Traditional intelligence tests are characterised by one correct answer and one correct solution method. In contrast, computer simulations may have multiple goals or solutions and multiple paths to reach that solution. Computer simulation tasks and traditional intelligence tests are compared and contrasted in detail in Section 1.2.7 of this thesis. Based on the differences between complex problem solving tasks and traditional measures of IQ, Neisser (1976) suggested that traditional intelligence tests possess low ecological validity and they only assess a relatively small component of the set of abilities that are required for real life complex problem solving.
Extreme proponents of the different cognitive demands hypothesis suggest that complex computer simulated scenarios represent the future of intelligence testing and that one day all traditional intelligence tests should be replaced by computer simulations (Putz-Osterloh, 1993; Kröner, Plass, & Leutner, 2005). Some researchers are developing new computer simulations that are deliberately designed to overcome the problems associated with current computer simulations in terms of low reliability and low correlations with traditional intelligence tests (e.g., MULTIFLUX, Kröner, Plass, & Leutner, 2005). The correlation between intelligence (BIS-K) and MULTIFLUX performance was significant, $r = .65, p < .05$ (Kröner, Plass, & Leutner, 2005). This finding does not support the different cognitive demands hypothesis. The MULTIFLUX simulation is described in detail in Section 3.2 of this thesis.

Deliberately designing computer simulations to have high correlations with traditional intelligence tests may be disadvantageous because these new simulations could violate the basic principle that problem solving competence and intelligence must not overlap at theoretical or operational levels (Wenke, Frensch, & Funke, 2005). In other words, a computer simulation could be deliberately designed by manipulating its task properties (complexity, connectivity, dynamics, intransparency, and polytely) to closer approximate a traditional intelligence test. For example, reduction of the number of variables and the casual connections between them (reduction of complexity and connectivity), making the task more static (reduction of dynamics), explicitly providing the participant with the rules underlying the system (reduction of intransparency), and creating a single optimal solution with a single solution strategy (reduction of polytely).
This is an issue that will be further explored in this chapter by employing a naturalistic decision making (NDM) task assessing ability to understand stocks and flows. Relationships between traditional intelligence and a stocks and flows task, the Dynamic Forecasting Questionnaire (DFQ), will be explored, bearing in mind that these tasks must not overlap at operational or theoretical levels. The DFQ, and similar stocks and flows tasks (Sterman & Sweeney, 2005), are assumed to be complex problem solving tasks. The DFQ and MULTIFLUX tasks share similar task properties, such as low complexity (i.e., few variables compared to a computer simulation), and both were designed based on NDM theory (Sweeney & Sterman, 2000). NDM theory is described in detail in Section 2.3 of this thesis. The different demands hypothesis in its strongest formulation, prediction of zero correlation between complex problem solving and intelligence, is not expected to be supported in Study 1.

### 4.1.2 Low Reliability Hypothesis

The low reliability hypothesis (Buchner, 1995; Funke, 1992, 1995) suggests that performance measures derived from complex computer simulated scenarios lack reliability. Traditional intelligence tests possess excellent psychometric qualities and are perceived as the greatest contribution of individual differences research to psychology (Wittmann & Süß, 1999). In contrast, the low reliability of computer simulations is suggested to account for their low correlations with traditional intelligence test tasks. Buchner (1995) and Funke (1995) suggest that poor goal specificity, lack of a specified solution, multiple solution methods, and nonlinear relationships between the variables, are features of computer simulations that lower
their reliability. As a consequence of the low reliability of computer simulations, they cannot correlate with highly reliable traditional intelligence tests.

Furthermore, determining the reliability of complex problem solving tasks is not as simple a process as determining the reliability of traditional intelligence tests, which are characterised by discrete items. Computer simulations are characterised by repeated trials where the same set of decisions must be made each trial. There is evidence that practice on complex problem solving tasks leads to learning (Beckmann & Guthke, 1995; Berry & Broadbent, 1987). Thus, it would be illogical to assess test-retest reliability (Wenke, Frensch, & Funke, 2005).

There is also evidence that parallel-test reliability might be low due to the influence of semantic embedding and prior knowledge in the different cover stories of the two parallel computer simulations. ‘Cover story’ refers to the labelling of the input and output variables (Hesse, 1982; Putz-Osterloh, 1993; Stanley, Mathews, Buss, & Kotler-Cope, 1989). For example, the SUGAR FACTORY and an analogous computer simulation the COMPUTER PERSON (also known as PERSONAL INTERACTION) were administered to participants. In the version with the SUGAR FACTORY cover story, participants controlled the size of a workforce to achieve and maintain a given production level. In the version with the COMPUTER PERSON cover story, participants controlled the behaviour of a fictional person, Clegg, to improve the relationship between Clegg and his union chief. Participants achieved higher overall scores on the COMPUTER PERSON task than the SUGAR FACTORY task, even though the underlying mathematical algorithms of the tasks were identical (Berry &
Broadbent, 1984). These two computer simulations are described in more detail in Section 3.2.

Despite the problems associated with determining test retest and parallel-test reliability for computer simulations, some researchers have attempted to provide reliability estimates for these tasks. For example, test-retest reliability of the TAILORSHOP and POWERPLANT simulations were reported to be around .63 when performance was retested over one year (Süß, 2001) and aggregated scores were used across all trials. Reliability of performance scores for NEWFIRE and COLDSTORE computer simulations assessed by internal consistency (Cronbach’s alpha) were found to be .80 and .77 respectively (Rigas, Carling, & Brehmer, 2002). These findings do not support the low reliability hypothesis. Reliability of the complex problem solving tasks employed in this thesis will not be reported. As the studies were not longitudinal, test retest reliability may be confounded by practice effects. Parallel forms of the computer simulations (e.g., Furniture Factory and its parallel form, Coach) were not employed in the present research as these types of studies have been carried out previously (Pillinger, 2004) and the simulations were demonstrated to be sufficiently reliable for research purposes.

4.1.3 Brunswik Symmetry

The Mannheim Research Project (Süß, 1996; Wittmann & Süß, 1999) was developed to investigate the determinants of individual differences in complex problem solving performance. This research group suggested that the difficulties in establishing a
relationship between intelligence and problem solving competencies in previous research (Funke, 1995; Kluwe, Misiak, & Haider, 1991) is due to a lack of symmetry between the predictors, e.g. traditional intelligence tests, and the criteria measures (e.g., performance on complex computer simulated scenarios). In the majority of previous research, the predictor was an IQ score or a general factor of intelligence. The criterion measure in previous research was a single act criterion, calculated by overall performance during the last trial of a computer simulation of unknown reliability (Fishbein & Ajzen, 1975).

The relationship between complex problem solving and intelligence may be strengthened by using aggregated scores over several trials of a simulation rather than a single act criterion in the final trial (Wittmann & Süß, 1999). This aggregated scoring method will be adopted in the experiment presented in this chapter. However, scoring issues are not a major focus of this chapter and will be investigated and criteria measures investigated in more detail in Study 2.

In addition, the lack of a relationship between complex problem solving performance and intelligence associated with using a global IQ score as the predictor could possibly be overcome by employing cognitive abilities tests that tap into the different factors of intelligence, as described in the theory of fluid and crystallised intelligence (Gf-Gc theory: Horn & Cattell, 1966). Gf-Gc theory is explained in detail in Section 3.1. In this chapter, the relationship between performance on complex problem solving tasks and specific second order factors from Gf-Gc theory will be investigated. Relatively little research has investigated the relationship between specific components of intelligence and complex problem solving. However, some German research has been
conducted using specific intelligence factors measured by the Berlin Intelligence Structure (BIS: Jäger, Süß, & Beauducel, 1997).

4.1.4 Berlin Intelligence Structure Model

There has been relatively little research in exploring the relationship between complex problem solving and specific components of intelligence. As previously mentioned, the majority of research has focussed on global IQ measures. However, in Germany a hierarchical model of intelligence has been developed and employed in a small number of studies investigating complex problem solving (Wittmann & Süß, 1999). Chapter 3 of this thesis presents an overview of hierarchical theories of intelligence, which includes the theory of fluid and crystallised intelligence (Gf-Gc theory). More recently, the Berlin model of intelligence structure (BIS) (Jäger, 1982; Wittmann, 1988) was also based on hierarchical theories of intelligence.

Wittman and Süß (1999) suggested that some reservations remain about whether the BIS and Gf-Gc models are identical. A point of difference between the two models is that the BIS constructs are derived from experimental research with a focus on mean differences. In contrast, the Gf-Gc factors are derived from correlational research that deliberately removes the means from resulting constructs. In addition, the BIS is a facetted model whereas Gf-Gc theory is not (Wittmann & Süß). Despite these reservations, the BIS model and Gf-Gc theory are similar in terms of assumptions of hierarchy, multimodality, and multifactorial dependencies of intelligence test performances (Wittmann & Süß). The BIS consists of two facets, which could be
adapted to a three faceted model such as Guttman’s Radex model (see Section 3.1). There are also similarities (described in Section 3.1) between Gf-Gc theory and Guttman’s Radex model. However, the BIS model includes creativity as part of intelligence (Guilford, 1967), whereas creativity is not included in the Gf-Gc model.

The BIS model consists of four operations: Processing capacity, creativity, memory, and speed. It also consists of three contents: Verbal, figural, and numerical. The structure of the model is similar to Guilford’s (1967) Structure of Intellect (SOI) model, containing facets to organise ability. In the BIS model, four ‘operations’ are on one side of the square and three ‘contents’ are on the side, making 12 BIS cells (Süß & Beauducel, 2005). The general ‘g’ factor derived from the BIS is not biased compared with traditional g. Raven’s Advanced Progressive matrices (APM) is often used as a g measure. APM does not correlate with any BIS factors other than the BIS-K factor which is defined as processing capacity, capturing the ability to recognise relations and rules and to form logical inferences in figure series, number series, verbal analogies (Wittmann & Süß, 1999). Similarly, APM correlates with the Gf factor in Gf-Gc theory. There has been some concern over whether the results of studies employing the BIS can be related to studies employing cognitive abilities tests based on Gf-Gc theory. The comparison of the two models just described suggests that cognitive abilities measures derived from both models should be comparable. This assumption could be tested if the BIS was translated into English so that actual test items could be compared. However, a comparative study is beyond the scope of this thesis.

As far as the literature review indicates, this thesis represents the first time that the relationship between complex problem solving competencies and cognitive abilities
tests derived from Gf-Gc theory has been explored. Thus, results will be compared to previous research employing cognitive abilities based on the BIS model, which is not that dissimilar to Gf-Gc theory.

### 4.1.5 Specific Components of Intelligence as Predictors of Complex Problem Solving

In their review of previous complex problem solving research, Wenke, Frensch, and Funke (2005) suggested that specific components of intelligence are theoretically more interesting and are better predictors of complex problem solving performance than global IQ scores, such as Raven’s Progressive Matrices (Raven, Raven, & Court, 1993). Learning tests have also been successfully employed as predictors of complex problem solving (Beckmann & Guthke, 1995). Section 3.2 describes a study where learning tests are employed as predictors of performance on the CHERRYTREE simulation. However, the ‘componential’ approach of using learning tests as predictors is beyond the scope of this thesis, which concentrates on specific cognitive abilities tests as predictors of problem solving competencies. A comprehensive review of the results of previous studies employing complex problem solving tasks was provided in Section 3.2 and Appendix A and will now be discussed in relation to Study 1.

The Furniture Factory computer simulation (Wood & Bailey, 1985), is employed as a criterion measure of complex problem solving performance in the study presented in this chapter. The Furniture Factory has predominantly been employed in cognitive experiments investigating feedback specificity, self-efficacy, transfer, and learning
(Goodman, Wood, & Hendrickx, 2004), described in more detail in Section 3.2. Consistency of findings using the Furniture Factory task as a criterion measure will be discussed in comparison to previous results using other complex computerised scenarios (e.g., TAILORSHOP and NEWFIRE) as criteria measures.

The relationship between specific components of intelligence and complex problem solving was investigating in a study employing the computer simulation FEUER, which is a more complex version of NEWFIRE (Brehmer, 1987) (described in detail in Section 3.2). Significant correlations were observed between performance in FEUER and spatial ability (Gv) (Schoppek, 1991). Spatial ability (Gv) is described in Section 3.1 as measured in tasks involving visual closure and constancy, and fluency in “imaging” the way objects appear in space as they are rotated and flip-flopped in various ways. The relationship was explained by the fact that fire fighting is a spatio-temporal task (Brehmer, 1995).

An alternative explanation for the finding was that the broad visualisation factor (Gv), in complex Gv tasks, tends to load on the general ‘g’ factor of intelligence (Lohman, Pellegrino, Alderton, & Regian, 1987). If this were the case, then performance on NEWFIRE would be expected to have higher correlations with a measure of general intelligence (e.g., Raven’s Advanced Progressive Matrices, APM: Raven, Raven, & Court, 1993) than with spatial ability (Gv). However, the correlation between NEWFIRE and APM in a subsequent study (Rigas, Carling, & Brehmer, 2002), was lower, \( r = .25 \), than the strong correlation observed between NEWFIRE and Gv (Schoppek, 1991).
Another study (Wittmann & Süß, 1999) was conducted to investigate the relationships between complex problem solving competencies and working memory capacity (WMC) tasks. There was no correlation between WMC numerical, verbal, and switching and problem solving competencies. Results suggested that WMC spatial was the only WMC task to load on complex problem solving performance. Note that these studies were not designed using Gf-Gc theory to guide the selection of cognitive abilities tests. However, an attempt has been made to relate findings derived from single tests to Gf-Gc theory where possible.

Most of the previous research investigating the relationship between intelligence and complex problem solving competencies has produced a significant correlation by experimental manipulation of the computer simulation conditions. For example, a higher correlation between intelligence and complex problem solving performance in BIOLOGY LAB is observed when goal specificity is high. No corresponding correlation is observed in low goal specificity conditions (Vollmeyer, Burns, & Holyoak, 1996). Similarly, Strohschneider (1991) found significant correlations (up to $r = .59$) between almost all BIS factors and complex problem solving performance using the MORO task when goal specificity was high. The BIOLOGY LAB and MORO studies are described in more detail in Appendix A. In conditions of high transparency (i.e., participants were provided with a model explaining the underlying rules of the system), a significant relationship was found between intelligence and MULTIFLUX simulation performance, $r = .65$ (Kröner, Plass, & Leutner, 2005).

A review of the results from studies presented in Section 3.2, suggests that the relationship between intelligence and complex problem solving appears to be mediated
by task properties such as goal specificity (the opposite of polytely) (Vollmeyer, Burns, & Holyoak, 1996), feedback specificity and feedback delays (Goodman & Wood, 2004; Goodman & Wood, 2006; Goodman, Wood, & Hendrickx, 2004), complexity/connectivity (Kröner, Plass, & Leutner, 2005), and intransparency (Kröner, Plass, & Leutner, 2005; Putz-Osterloh & Lüer, 1981). Manipulation of task properties was not a direct aim of the studies presented in this thesis. However, findings from previous research, utilising manipulations of task properties, will be useful in guiding the conclusions drawn from the present experiments.

Wittman and Süß (1999) suggested that complex problem solving performance is related to specific components of intelligence (rather than global IQ). Based on their suggestion, it is hypothesised that complex problem solving performance, measured by Furniture Factory simulation and Dynamic Forecasting Questionnaire (DFQ) performance, will correlate with specific components of intelligence. Wittman and Süß (1999) employed the BIS model to assess specific components of intelligence. Their findings will be extended in Study 1 by employing cognitive abilities tests derived from the Gf-Gc model of intelligence.

4.1.6 Gender and Group Differences in Complex Problem Solving

Previous research has suggested that cover stories (semantics of the labelling of input and output variables) can influence overall complex problem solving performance. It seems plausible that knowledge of a specific domain would improve complex problem solving performance if the cover story of the computer simulation was consistent with
the knowledge base of the participant (Hesse, 1982; Putz-Osterloh, 1993; Stanley, Mathews, Buss, & Kotler-Cope, 1989).

Funke and Hussy (1984) conducted a study directly investigating gender effects of cover stories on complex problem solving. Participants were administered LUNAR LANDER (Thalmaier, 1979) where the main goal is to control the landing manoeuvre of a spacecraft on the surface of the moon. Participants were also administered COOKING, which is a parallel form of LUNAR LANDER, consisting of the same well-defined underlying mathematical algorithms. The only difference between the two computer simulations was the cover story. It was expected that males would outperform females on LUNAR LANDER. Conversely, it was expected that females would outperform males on COOKING. These hypothesised gender differences in complex problem solving performance were based on assumptions of gender differences in the semantic domains. However, results regarding gender differences were not significant. Males and females were roughly equivalent in their complex problem solving performance for either version of the task.

The finding of significant gender differences on the Tailorshop task, favouring males, prompted an open discussion amongst researchers to suggest causal explanations for the results (Wittmann & Süß, pp. 105-108). The following is a summary of that open discussion, thus there are no references other than the paper just mentioned. Alexander suggested that the cover story of a shirt manufacturing company was biased towards males because while females enjoy purchasing shirts, they are not particularly interested in the industrial concerns required for the manufacturing process. If Alexander was correct, we would expect males to outperform females on the Furniture Factory task.
also, because the cover story is similar to the Tailorshop, although the focus of the Furniture Factory is on the efficiency of staff, rather than the manufacture of goods.

Ackerman and Kanfer also suggested that the cover story may be influential in causing gender differences on the Tailorshop task. They suggested that a study could be conducted using parallel forms of tasks, where one task had the cover story of a Tailorshop and the other featured the cover story of a Dress shop. It was hypothesised that males would outperform females on the Tailorshop version of the task. Conversely, females would outperform males on the Dress shop version of the task. Again, these gender differences were based on assumptions of gender differences in the semantic domains of the cover stories. However, based on the findings of Funke and Hussy (1984), the semantic domains of the cover stories is unlikely to cause gender differences in complex problem solving competencies.

Wittmann and Süß (1999) conducted further research to find a causal explanation for the gender differences favouring males they observed in complex problem solving on the Tailorshop simulation. Wittmann and Süß (1999) administered the TAILORSHOP computer simulation along with questionnaires assessing economics knowledge (e.g., “Which of the following strategies serves best to decrease inflation rate?”) and an assessment of computer literacy. Results suggested that males had higher complex problem solving competencies than females. The main predictor variables that accounted for the gender difference were higher male than female scores in economics knowledge and computer literacy tests (Süß, Oberauer, & Wittmann, 2004; Wittmann & Süß, 1999). It was suggested that females were less competent in interacting with the computer possibly due to a lack of training. Subsequent training investigating training
interventions suggested that females benefited more than males if the trainer was also female (Wittmann & Süß, 1999).

Therefore, based on previous research (Funke & Hussy, 1984; Wittmann & Süß, 1999), no gender differences are expected for performance on the Dynamic Forecasting Questionnaire (DFQ) or for the Furniture Factory simulation employed in Study 1. Gender differences are not expected based on the cover stories of these tasks. However, gender differences may be expected due to gender differences in relevant skills.

Based on previous research investigating effects of cover stories described in this section, it is also expected that master of business administration (MBA) students would outperform undergraduate psychology students on a test of management skills such as the Furniture Factory task employed in Study 1 in this chapter.

4.1.7 Strategy in Complex Problem Solving

It has been suggested that both intelligence and knowledge are important predictors of complex problem solving performance (Wittmann & Süß, 1999). Complex problem solving knowledge consists of rule identification, rule knowledge, and rule application (Kröner, Plass, & Leutner, 2005). Crystallised intelligence (Gc) is defined in Section 3.1 as measured in tasks indicating breadth and depth of the knowledge of the dominant culture (Horn & Noll, 1993). Knowledge acquisition and application is dependent on a certain level of intelligence (Cattell, 1971) and this knowledge may be highly correlated with Gc. Rule identification refers to the exploration strategies that participants use to
understand the underlying rules of the system. Rule knowledge is the amount of explicit knowledge learned throughout simulation performance that can be articulated in a post-test questionnaire. Rule application is assessed by simulation performance where implicit and explicit knowledge of the causal connections between variables in the system are applied to decision making in the complex computerised scenario. In the present chapter, rule identification will be measured by exploration strategy. The relationship between cognitive abilities and the strategies employed by participants to discover the underlying rules of the Furniture Factory task will be investigated in Study 1.

Kröner, Plass, and Leutner (2005, p.352) employed the MULTIFLUX task to develop a model illustrating the relationships between global intelligence and three rule categories in complex tasks: Rule identification, rule knowledge, and rule application. Using the Furniture Factory game review questionnaire as an example, Kröner et al.’s rule knowledge factor can be extended to include three subfactors: Basic Facts/content, specific inferences about variables, and general information about decisions. The Furniture Factory game review questionnaire is included as Appendix B.

The Furniture Factory game review questionnaire lists the questions that could be employed to assess explicit knowledge in the Furniture Factory simulation. In this example, the rule knowledge factor from Kröner et al.’s (p. 352, 2005) model is divided into three subfactors. Firstly, questions 1 to 10 assess general information about decision making in the complex problem solving task. Secondly, questions 11 to 18 assess specific inferences about variables. Thirdly, questions 19 to 24 assess knowledge of the basic facts/content of the computer simulated scenario.
In addition to more detailed specification of knowledge factors, Kröner et al’s model can also be extended to include specific components of intelligence based on previous research by Wittman and Süß (1999), as shown in Figure 4.1.
Feedback is given to the participant as they explore the game leading to good performance space and bad performance space.

More learning in poor performance space and less learning in good performance space. Thus, expect poor performance space to have higher correlation with Intelligence.

**Assessed by Exploration Strategy**

**Assessed by Game Review Questionnaire**

**Assessed by Simulation Performance**

1. Mean Rule Knowledge A & Rule Application
2. Mean Rule Knowledge B & Rule Application
3. Mean Rule Knowledge C & Rule Application
4. Mean Total Rule Knowledge & Rule Application

**Rule Knowledge A**

- Basic Facts/Content
  - Q. 19-24

**Rule Knowledge B**

- Specific Inferences about Variables
  - Q. 11-18

**Rule Knowledge C**

- General Information about Decisions
  - Q. 1-10

Figure 4.1 Theoretical model of Furniture Factory complex problem solving competencies and specific components of intelligence.
Kröner et al. (2005) found that the relationship between rule application, assessed by MULTIFLUX performance, and global intelligence was significant only in a transparent version of the MULTIFLUX task. This path is displayed in Figure 4.1. It is hypothesised in Study 1 that this relationship would be significant in a nontransparent version of the Furniture Factory simulation if specific components of intelligence were employed as the criterion measure rather than simply a global IQ measure.

The relationship between cognitive abilities and the strategies employed by participants to discover the underlying rules of the Furniture Factory task will also be investigated in Study 1. Exploration strategy can reveal individual differences in complex problem solving performance even though overall performance scores may be equivalent. In a study by Spering, Wagener, and Funke (2005) participants were administered the Forestry system FSYS 2.0 under two experimental conditions: positive emotions and negative emotions. Overall complex problem solving performance was the same for both conditions. However, participants experiencing negative emotions were more focused on the seeking and use of information. Thus, manipulation of emotions resulted in different problem solving strategies, while overall problem solving performance scores remained unchanged.

In another study by Goodman, Wood, and Hendrickx (2004) participants were administered the Furniture Factory task under different conditions of feedback specificity. Those in the high feedback specificity condition tended to become overly dependent on feedback, resulting in less exploration and information seeking. Feedback specificity interacted with exploration strategy to affect learning. Less exploration of the dynamic system resulted in decreased learning of the rules underlying the system.
In addition, exploration strategies can be affected by experience in the domain of the cover story of the complex computer simulation (Krems, 1995).

Individual differences exist in the type of hypothesis testing that participants engage in when trying to discover the underlying rules of a complex system (Klayman & Ha, 1987, 1989). Participants engage in a sort of trial and error where they perform actions and learn from the resulting automatic feedback provided by the system. In addition to hypothesis testing, the process of identification can also be described using the two-space model (Simon & Lea, 1974). The two-space model separates the problem space into a rule space and an instance space. The rule space allows participants to generate hypotheses about the causal connections between system variables. All possible rules underlying the system are presented in the rule space. The instance space allows participants to plan and generate experiments. All possible states of the system are presented in the instance space (Simon & Lea). This two-space model was developed into an approach called scientific discovery as dual search (SDSS, Klahr & Dunbar, 1988).

Both the two-space model and the SDSS model assume the existence of an experiment space and a hypothesis space. The spaces interact with one another in a cycle, actions in the experiment space result in automatic feedback from which hypotheses can be developed; these hypotheses guide the participant to conduct another experiment, which leads to more hypotheses which lead to more experiments, and so on (Klahr & Dunbar, 1988). The two-space model and SDSS model do not suggest how to assess strategy during the process of complex problem solving. Exploration strategy has been operationalised for performance in the Furniture Factory computer simulation.
(Goodman, Wood, & Hendrickx, 2004), which is employed in the experiment in this chapter. There are three types of exploration strategy: Unsystematic exploration, systematic exploration, and change nothing.

Unsystematic exploration occurs when a participant engages in confounded experiments by changing two or more decisions at the same time (Goodman et al.). For example, if your car refuses to start, a mechanic could replace the battery, the starter motor, and the carburettor all at the same time. When the car starts, it will not be known what the problem actually was due to the three confounding variables. In addition, the solution will have been unnecessarily costly. In the Furniture Factory simulation, participants can engage in unsystematic exploration by making multiple decisions per trial.

Systematic exploration is suggested to be the optimal approach to hypothesis testing in real life and in complex problem solving tasks (Tschirgi, 1980; Vollmeyer, Burns, & Holyoak, 1996). The mechanic could have changed only one variable at a time, allowing unequivocal identification of the impact that the each change made on whether or not the car would start. Positive feedback would be the car starting and negative feedback would be the car not starting. In the Furniture Factory task, participants can engage in systematic exploration of the complex system by changing one decision per trial per employee. Positive feedback would be characterised by employees working more efficiently than the previous trial; and negative feedback would consist of employees working less efficiently than in the previous trial. The Furniture Factory computer simulation is described in detail in the Method section of this chapter.
The change nothing strategy could have been adopted by the mechanic if they had kept trying to start the car without any interventions. Similarly, in the Furniture Factory task, participants may retain their decisions from trial to trial and make no decision changes. The change nothing strategy tends to be adopted in conditions of positive feedback. If the goal state is being reached, there is no need to change strategies. An analogous situation would be profitable management of a company. The manager is said to be in a good performance space; therefore, there is no need to engage in hypothesis testing because the goal state of increasing capital is being achieved. It has been suggested that participants learn more in poor performance space, because they are engaging in hypothesis testing and learning the underlying rules of the system (R. E. Wood, personal communication, September 22, 2005). The three possible strategies in the Furniture Factory game: Systematic exploration, unsystematic exploration, and change nothing, define an exhaustive set of decision options with each option being mutually exclusive (Goodman, Wood, & Hendrickx, 2004). The relationship between exploration strategy on the Furniture Factory task and specific components of intelligence will be explored in Study 1.

4.1.8 Naturalistic Decision Making in Stocks and Flows Tasks

The experiment presented in this chapter represents the first time that specific components of intelligence have been used to predict performance on the Dynamic Forecasting Questionnaire (DFQ: Atkins, Newell, & Wood, 2001). The DFQ is employed in this chapter as a criterion measure of complex problem solving competency. Therefore, correlations between DFQ and specific components of
intelligence cannot be directly compared to previous research. However, consistency of our results with those found by naturalistic decision making (NDM) researchers (Sterman & Sweeney, 2005), who employed a similar stocks and flows tasks will be discussed.

The Sterman and Sweeney (2005) study employed two stocks and flows tasks: The Bathtub and Cash flow task. These dynamic tasks were used as criteria measures of complex problem solving performance. Cognitive abilities tests were not employed as predictor variables. The researchers inferred the level of intelligence of their participants by analysis of their university education. For example, the sample was comprised of students enrolled in the introductory system dynamics course at the MIT Sloan School of Management. ‘About three fourths were master of business administration students; the rest were in other master’s programs, PhD students, undergraduates, or students from graduate programs at other universities. More than one half had undergraduate degrees in engineering, computer science, mathematics, or the sciences; most of the rest had degrees in business or a social science (primarily economics). Fewer than 5% had degrees in the humanities. More than one third had a master’s, doctoral, or other advanced degree, most in technical fields.’ (Sterman & Sweeney, 2005, p. 71).

In the experiment presented in the current chapter, mathematics ability will not be inferred by whether students are enrolled in Arts, Science, Engineering, or Commerce. Cognitive abilities tests, including quantitative knowledge (Gq), will be explicitly measured by valid and reliable psychometric tests. Quantitative knowledge (Gq) is
measured in tasks requiring understanding and application of the concepts and skills of mathematics (Horn & Noll, 1993).

Results of the Sterman and Sweeney (2000; 2005) studies must be interpreted with caution to the lack of a psychometrically sound measure of cognitive abilities. The differences between the experiment in this chapter and the Sterman and Sweeney study are a good example of the differences between those researchers from an NDM background and those from an Individual Differences background. NDM researchers tend towards field studies, top down approach, and analysis of case studies. Individual differences researchers tend towards laboratory studies, bottom up approach, and analysis of large samples. Each approach had theoretical and methodological weakness and there is a need to synthesise results from both fields (Pretz & Sternberg, 2005).

Sterman and Sweeney (2005) employed two paper and pencil tasks that represented common dynamic systems from the real world: The Bathtub and Cash Flow. In the Bathtub task, the water level in the bathtub (stock) depends on how much water is coming in through the tap and going out through the drain (flow). In the Cash Flow task, the cash balance of a company (stock) has inflow and outflow as rates of receipts and expenditures respectively. Results suggested that participants from an elite business school, with extensive training in mathematics and science, could not understand stock and flow relations in highly simplified settings. Average performance in the most simple graphical integration task (measuring conservation of matter) was 77%. Average performance in a more complex graphical integration task (measuring understanding of the relation between the net flow into a stock and the slope of the stock trajectory) was just 48%. Participants had difficulty in understanding action / feedback loops under
conditions of time delay. Even though participants had successfully completed university courses in calculus, they could not understand basic principles of calculus such as: the slope of the stock is the net flow, and the change in the stock over an interval is the area enclosed by the net rate in that interval. It was suggested that poor performing participants, relied on a heuristic that matches the shape of the output of the system to the shape of the input. From the results of Sterman and Sweeney’s study, it is predicted that participants will have difficulty in the DFQ, stocks and flows task. Relationships between specific components of intelligence and complex problem solving performance on the DFQ task will be explored. Naturalistic Decision Making (NDM) is described in more detail in Section 2.3.2.

4.1.9 Overview of Predictions: Study 1

Historically, studies employing computer simulations have been unable to establish a relationship between intelligence and complex problem solving competencies (Dörner, 1996). The chequered history of results can be explained by a lack of symmetry (Brunswik symmetry) between predictor measures of intelligence test performance and criterion measures of computer simulation performance scores (Wittmann & Süß, 1999). The Brunswik lens model can be used to explain two major problems with previous research. Firstly, the predictor measure of intelligence has historically been an overall IQ score. To overcome the problems associated with employing a global IQ measure, it has been suggested that specific components of intelligence (e.g. Berlin Intelligence Structure, BIS factors) that are more symmetrical with complex problem solving performance should be employed. Secondly, the criterion measure of complex
problem solving performance has historically been the overall performance score in the
last trial of a computer simulation. To overcome the problem associated with a single
performance score, on a single trial of a computer simulation, it has been suggested that
aggregated scores across multiple trials should be calculated (Wittmann & Süß, 1999).
Scoring techniques for computer simulations will be considered in Study 1 of this
chapter and examined in more depth in Study 2 of subsequent chapters. By ensuring
that predictor and criterion measures employed in this study are compatible with
Brunswik symmetry, it is predicted that both the different demands hypothesis and low
reliability hypothesis will not be supported in Study 1. Thus, it was expected that there
would be a relationship between complex problem solving and intelligence.

Relatively little research has explored the relationship between complex problem
solving and specific components of intelligence. However, some German research has
been conducted using specific intelligence factors measured by the Berlin Intelligence
Structure (BIS) test (Jäger, Süß, & Beauducel, 1997). The specific components of
intelligence featured in the BIS model are roughly compatible with those characterised
by the Gf-Gc model. This thesis represents the first time that specific cognitive abilities
derived from Gf-Gc theory will be employed in the prediction of complex problem
solving performance. In Study 1, it is hypothesised that specific components of
intelligence will predict complex problem solving performance in both the Dynamic
Forecasting Questionnaire (DFQ) and the Furniture Factory simulation tasks.
Specifically, it was hypothesised that quantitative knowledge (Gq) would correlate with
performance on the DFQ due to the mathematical nature of the task. It was also
hypothesised that acculturation knowledge (Gc) would correlate with performance on
the Furniture Factory simulation due to the large amount of verbal instructions presented.

Previous research has suggested that cover stories (semantics of the labelling of input and output variables) can influence overall complex problem solving performance (Hesse, 1982; Putz-Osterloh, 1993; Stanley, et al., 1989). It is hypothesised that due to prior knowledge in the domain of management, MBA students’ complex problem solving performance scores on the Furniture Factory task will be significantly greater than Furniture Factory performance scores achieved by the undergraduate sample in Study 1. Prior studies have also investigated the influence of cover stories on gender differences in the performance of complex problem solving tasks. No gender differences were found based on assumptions of gender bias in the semantic domains of the cover stories (Funke & Hussy, 1984). In addition, previous findings of gender differences in the Tailorshop simulation, were suggested to be due to greater male than female knowledge of economics and computers (Süß, Oberauer, & Wittmann, 2005; Wittmann & Süß, 1999). Therefore, it was hypothesised that there would no gender differences for complex problem solving performance on the Furniture Factory simulation or the DFQ tasks. However, if a gender bias is found to exist, it is likely to be caused by gender differences in abilities rather than gender bias in the cover stories of the tasks.

Successful performance in complex problem solving tasks is a reflection of the extent to which participants have understood the causal connections between the variables, in the underlying model of the computer simulation. Exploration strategy indicates the type of hypothesis testing that is performed in real life and in complex computer simulated
scenarios (Tschirgi, 1980; Vollmeyer, Burns, & Holyoak, 1996). It was hypothesised that exploration strategy on the Furniture Factory task would correlate with fluid reasoning (Gf) as participants must understand relations among stimuli to perform well on the Furniture Factory simulation.

Finally, previous research has suggested that naturalistic decision making tasks such as stocks and flows systems are characterised by poor performance (Sterman & Sweeney, 2005). Participants seem to rely on heuristics from the real world that impede their ability to understand dynamic systems presented to them in an experiment. Based on Sterman and Sweeney’s findings, it is predicted that participants will have low performance scores on the DFQ, which is a stocks and flows task. The relationships between specific components of intelligence and complex problem solving on the DFQ task will also be explored. During the design of Study 1, it was assumed that the Furniture Factory task and the DFQ were complex problem solving tasks based on their similarities to tasks employed in previous complex problem solving (e.g., Tailorshop: Süß, Kersting, & Oberauer, 1991; 1993) and Naturalistic Decision Making research (e.g., stocks and flows tasks: Sterman & Sweeney, 2002; 2005) respectively. The premise suggested by Wenke, Frensch and Funke (2005) that complex problem solving tasks and intelligence tests must not overlap at theoretical or operational levels (i.e., that complex problem solving tasks are not computerised versions of traditional intelligence tests) will be evaluated for both the Furniture Factory and DFQ tasks.
4.2 Method

4.2.1 Participants

A total of 130 participants were involved in the study. The average age of the sample was 20.34 years with a standard deviation of 3.92 years. There were 82 females and 48 males. The mean age of the females was 20.57 years, with an age range from 17 to 41 years. The males had a mean age of 19.94 years, with an age range from 18 to 27 years. Participants with English as their first language represented 80% of the sample.

The sample comprised 86 first-year psychology students from the University of Sydney who were participating as part of a course requirement. In addition to course credit, first-year students received a detailed report outlining their results, with an emphasis on cognitive abilities and personality. The sample also comprised 44 second and third-year psychology students who responded to announcements in lectures. In return for their participation, third-year students received a report (more detailed than conventional participant feedback) outlining their results, with an emphasis on cognitive abilities.

All participants were questioned to ensure they had no anxiety about working with a Personal Computer (PC). In the case where an individual had no previous experience with working with a PC, they were excluded from the study. There were 2 such cases, both were aged over 50 years.
4.2.2 Materials

4.2.3 Cognitive Abilities Tests

Eight cognitive abilities tests measuring four factors within the theory of fluid and crystallised intelligence (Gf/Gc, (Cattell, 1943; Horn & Cattell, 1966) were administered to all participants. The cognitive domains assessed were Fluid Reasoning (Gf), Crystallised/acculturation Knowledge (Gc), Quantitative Knowledge (Gq), and Broad Visualisation (Gv). Swaps and Triplets tests were employed as Gf measures; Vocabulary Test and Proverbs Matching were employed as Gc measures; Numerical Operations and Financial Reasoning were employed as Gq measures; and Line Length and Letter Spotting were employed as Gv measures. For ease of recognition the tests are numbered as they appear in tables later in this thesis. All tests were in multiple-choice format (with the exception of Numerical Operations), they were computer administered, and scored automatically. The order of presentation of the tests for all participants was: Swaps, Vocabulary, Numerical Operations, Triplets, Proverbs Matching, Financial Reasoning, Line Length Test, Letter Spotting Test. Before each test, participants were presented with instructions on the particular task and at least one example. A time limit, determined from a pilot study, was placed on all tests. However, the time allowed was sufficient for approximately 95% of participants in the pilot study to complete the tests (under the instruction to work as quickly and accurately as possible). A detailed description of each test follows.
Fluid Reasoning: Swaps and Triplets tests were employed to measure Fluid Reasoning (Gf).

(1) Swaps

The Swaps task was developed by Stankov and associates (Stankov, 1999), it has good psychometric properties, and embodies sound principles from experimental cognitive psychology. The stimuli for the task were 3 framed pictures of equal size (e.g., Bird, Light bulb, Bottle, Eggs, Cards), which appeared on the screen in sets of 3 in various orders. For the 2 practice items, followed by 20 items, participants were instructed to mentally ‘swap’ the positions of the pictures. For example, in the level 1 ‘swaps’ condition, participants were presented with three pictures in the following order: Bottle, Eggs, Cards. They were then asked what the order of the pictures would be if you ‘swap’ 1 and 3. Participants were presented with 6 alternatives and were requested to use the mouse to click on the alternative that was their answer. The answer to the example is: Cards, Eggs, Bottle. In the practice questions, participants are shown the pictures physically moving into the new order. However, in the actual test items, the pictures do not physically move but must be mentally ‘swapped’ by the participant. Complexity was manipulated by altering the number of times that the pictures were to be rearranged and included level 1, level 2, level 3, and level 4 swaps. For example, a level 4 swaps condition would ask the participants what the new order of the pictures will be if you: Swap 1 and 3, then swap 1 and 2, then swap 1 and 3, then swap 2 and 3. The task was computer administered and use of paper and pencil was not allowed in order to increase the load on working memory. There was a generous 3-minute time
limit for each item and participants were instructed to work as fast and accurately as possible.

(2) Triplets

The Triplets task was developed by Stankov and associates (Stankov, 1999), it has good psychometric properties and embodies sound principles from experimental cognitive psychology. For each of the 42 items plus 3 practice items, participants were instructed to make decisions about sets of 3 numbers (e.g., 7, 3, 1). In the first block of 10 items, participants were instructed to remember the following rule: ‘Click on YES if the first number is the highest and the last number is the lowest. If this is not the case, click on NO.’ In the second block of 32 items, participants were instructed to remember the following rule: ‘Click on YES if the first number is the highest and the last number is the lowest OR if the first number is the lowest and the middle number is the highest. Click on NO if neither of these situations are true.’ Each item contained a different set of 3 numbers although the rule remained the same for each of 2 blocks of items. The task was computer administered and use of paper and pencil was not allowed in order to increase the load on working memory. There was a generous 1-minute time limit for each item and participants were instructed to work as fast and accurately as possible.
Crystallised/acculturation Knowledge: Vocabulary and Proverbs Matching tests were employed to measure Crystallised Knowledge (Gc).

(3) Vocabulary Test

The Vocabulary test is a common synonyms vocabulary test (Frensch, Ekstrom, & Price, 1963) and is similar to the Vocabulary sub-test of the WAIS-III (Wechsler, 1997), a computerised version was employed (Stankov, 1999). For the 2 practice items followed by 30 questions, participants were presented with a word at the top of the screen (e.g., help) and were asked to use the mouse to click on the word or phrase from 4 alternatives presented below (e.g., sympathise, assist, hinder, lift up) that was closest in meaning to the word at the top of the screen (e.g., the synonym). There was a generous 1-minute time limit for each item and participants were instructed to work as fast and carefully as possible.

(4) Proverbs Matching

The Proverbs Matching test was developed by Suzanne Morony under the direction of Lazar Stankov and was computer administered (Stankov, 1999). This task was included as a Gc marker as it is relies on acculturated knowledge. The participant is presented with a proverb, such as ‘Birds of a feather flock together’, and is asked to choose from amongst 5 alternatives, a second proverb with the most similar meaning or relevant message as the original proverb. For example: Tell me what company you keep and I will tell you who you are; Opposites attract; There is little friendship in the world and
least of all between equals; To check an elephant, inspect its tail; Shared joy is doubled joy. There were 2 practice items followed by 17 questions. After the practice items, participants were given an explanation of the answer. In the above example, the explanation provided was: “Birds of a feather flock together” means that similar people tend to engage in similar activities. Therefore, the correct answer was “Tell me what company you keep and I will tell you who you are” as this again says that people who are together are similar. There was a generous 1 minute time limit per item and participants were instructed to work as fast and carefully as possible.

**Quantitative Knowledge:** Numerical Operations and Financial Reasoning tests were employed to measure Quantitative Knowledge (Gq).

(5) **Numerical Operations**

The Numerical Operations test is a common measure of the facility to carry out elementary arithmetic operations and is similar to the Numerical Operations sub-test of the WAIS-III (Wechsler, 1997). A computerised version of the task was employed (Stankov, 1999). ‘Problems requiring selection of appropriate mathematical operations’ (Carroll, 1993, p. 241) will have a high loading on the Gq factor (Carroll’s RQ factor). There were 2 practice items followed by 25 questions. Participants were instructed to carry out simple numerical calculations.
For example, in some items they were asked to type into a box the result of their calculations.

E.g. $121 \div 11 = \underline{}$

In some items participants had to indicate a numerical operation. For example, which sign is needed, (+ - × ÷) to satisfy the equation:

E.g. $5 \underline{} 10 = 15$

Thus, participants needed to either type in their answer or select a numerical operation with the mouse button. There was a generous 1-minute time limit for each item and participants were instructed to work as quickly and accurately as they could.

(6) Financial Reasoning

The Financial Reasoning test was originally developed by Lazar Stankov as part of a selection battery for entry-level jobs in financial institutions and was computer administered in this study (Stankov, 1999). ‘Mathematical reasoning problems such as word problems (solving verbally stated mathematical problems)’ (Carroll, 1993, p. 241) will have a high loading on the Gq factor (Carroll’s RQ factor). There was 1 practice item followed by 12 test items. Participants were instructed to solve financial problems and to choose their answer from amongst 5 alternatives that appeared on screen underneath the question. To the right of each question and the corresponding answer
choices, a calculator appeared on screen to assist participants. They were instructed to press the buttons on the calculator using the mouse, in a similar manner to a normal calculator. An example is, ‘Linda deposited $1000 into a savings account. A year later she closed the account. After the deduction of $6 service charge, she received $1084. What was the annual interest rate for the savings account?’ Alternative answers were: 7.5%; 8%; 9%; 9.75%; and 10%. There was a 1 minute time limit on each item and participants were instructed to work as fast and carefully as they could.

**Broad Visualisation:** Line Length and Letter Spotting tests were employed to measure Broad Visualisation (Gv).

(7) **Line Length**

The presentation of stimuli in the form of lines on a computer screen as a measure of the ability to make fine visual discriminations is well reported in the literature (Balakrishnan & Ratcliff, 1966; Baranski & Petrusic, 1994; Olsson & Winman, 1996). Carroll includes the line length test in abilities in the domain of visual perception, ‘length estimation is the ability to compare the length of lines or distances’ (Carroll, 1993, p. 308). The test employed here was derived from (Stankov & Crawford, 1996) and was also computer administered (Stankov, 1999). The test consisted of the simultaneous presentation of 5 vertical, non-aligned lines of between 50 and 60 pixels in length spaced 100 pixels apart. In each trial four lines were the same lengths whilst the fifth was 1, 2, 3, 4, or 5 pixels longer. The position of the longer line was randomised and participants were required to use the mouse to click on the ‘response button’ that represented the serial position of the longer line. The test consisted of 2 practice items.
followed by 20 test items (i.e., 4 items of each length difference). There was no time limit for completion although participants were instructed to work as fast and carefully as they could. [Note: A conversion factor of 30 pixels equals 1 cm is appropriate for the computer equipment employed in this study].

(8) Letter Spotting

The Letter Spotting test (Stankov, 1999) is a new computerised measure of perceptual speed. Carroll (1993, p. 310) includes the letter spotting test ‘Finding A’s’ in abilities in the domain of visual perception, characterized by the task of finding in a mass of distracting material a given configuration which is borne in mind during the search’ (Carroll, 1993, p. 308). During the Letter Spotting task, a series of letters were flashed on the screen (e.g., ‘dfgtdjtdq’) and participants were instructed to look for a particular letter (e.g., ‘d’) and count the number of times they could spot it. For each trial, the different letter strings were 10 letters in length and were flashed 10 times. There were 50 trials in total plus 2 practice trials. Trials were either all upper case or all lower case. Participants were instructed to use the mouse to choose from 5 alternative answers: None; 1; 2; 3; or 4. The speed at which the letter strings were flashed varied between 0.3 and 0.5 seconds. Participants were requested to work as fast and carefully as they could.
4.2.4 Complex Problem Solving Tasks

4.2.5 Dynamic Forecasting Questionnaire

Participants completed the Dynamic Forecasting Questionnaire (DFQ) developed by Paul Atkins (Atkins, Newell, & Wood, 2001) to measure understanding of system dynamics. The DFQ was developed at the Australian Graduate School of Management (AGSM) using a Masters of Business Administration (MBA) student sample. Stocks and flows offer a diagrammatic representation of variables within a dynamic system (Forrester, 1961). Previous research (Diehl & Sterman, 1995; Paich & Sterman, 1993) suggests that prediction of changes in dynamic systems is a difficult task. The DFQ was designed to be a tool for use by educators, researchers and business people to measure the abilities of participants to discern critical variables from complex scenarios and to discern flows from stocks. In contrast to an individual differences approach, the DFQ was developed as part of an ongoing experimental cognitive psychology research program at the management school that investigated performance on complex tasks (Atkins, Wood, & Rutgers, 2002; Wood, Atkins, & Tabernero, 2000). Reliability estimates based on Cronbach’s alpha, $\alpha = .74$, suggest the DFQ has sound psychometric properties.

The main objective for the participant is to estimate the behaviour of dynamic systems following a change in the state of the system. A simple example of a dynamic system is a bathtub where the water level in the bathtub (stock) depends on how much water is coming in through the tap and going out through the drain (flow). Another example of a dynamic system is an inventory (stock) that has inflow and outflow as rates of orders.
and sales respectively. More complex dynamic systems involve feedback loops such as a population with inflow and outflow as births and deaths respectively. The number of births and deaths depend on the population (stock). All the problems within the DFQ are examples of non-linear growth (births) and decay (deaths) that depend on the population (Atkins, Newell, & Wood, 2001) as shown in Figure 4.2.

The rate of births and deaths depends upon how many people there are in the population. This can be represented in the following way:

\[
\text{Births} = 10\% \times \text{Population} \\
\text{Deaths} = 10\% \times \text{Population}
\]

Population = 1000 people

What this picture shows is that the rate of births and the rate of deaths both depend on the size of the population. The diagram shows that the population starts at 1000 people. In this situation, 100 people would be born each year (10\% \times 1000 = 100) and 100 people would die (10\% \times 1000 = 100). Therefore, the population would stay constant at 1000 people.

**Figure 4.2.** DFQ Task. Graphical integration task with population growth cover story. DFQ questions were based on the dynamic system shown.

**Example DFQ Question**

What will happen to the population if, in year 5, the rate of births increases from 10\% to 20\% of the population as a result of a new fertility drug being introduced?
Explanation of correct answer:

The correct answer is option (d). This time the growth is exponential rather than linear. The more births there are the greater the population and the greater the population the greater the rate of new births. The rate of growth of the population is dampened slightly by the fact that, as population increases, deaths also increase. However, the number of deaths per unit population is still less than the number of births per unit population so exponential growth occurs.

The DFQ is a 15-item (plus 2 practice questions) paper and pencil questionnaire. Each item has 8 multiple-choice options. During the 10-minute practice phase, participants learned how dynamic systems could be represented in pictures. They were given 2
practice questions followed by detailed explanations of the correct answers. Participants were encouraged to consult with the experimenter if any questions arose. Next, they were administered the 20-minute test. The test was divided into 2 sections. The first section consisted of 5 questions that required prediction of population changes to the same system shown in Figure 4.2. The second section was based on a more complex system where the population was divided into both healthy and sick populations. Participants were asked 5 questions that required prediction of the spread of sickness and what impact it would have on both healthy and sick populations. Thus, in the second section 10 answers were required (both healthy and sick population predictions) for the 5 questions. Thus, participants were required to give 15 responses in total. As suggested by Atkins et al. (2001), total scores were calculated by summing the total correct out of the 15 answers.

4.2.6 Furniture Factory

The Furniture Factory is a business simulation that has been extensively used in the past in cognitive experimental research where performance in different computer simulations, or different versions of the same computer simulation, is compared, following the Broadbent tradition as described in Section 2.2.1. Past research has employed different versions of the Furniture Factory to investigate the effects of feedback specificity on exploration learning opportunities and learning (Goodman et al., 2004; Goodman & Wood, 2004; Goodman & Wood, 2005); learning and transfer (Pillinger, 2004) and the effects of task complexity and goal setting on self-regulation, exploration and performance (Wood, Atkins, & Bright, 1999; Wood, Bandura, &
Bailey, 1990). This thesis presents the first time the Furniture Factory task has been administered in experiments taking the individual differences approach, where an attempt is made to explain complex problem solving performance by intelligence (Dörner, 1996). The Furniture Factory task meets the criteria for a complex problem solving task, outlined in Section 1.2.6 and consists of complexity, connectivity, dynamic environment, and intrasparency (in the absence of polytely).

The Furniture Factory is a business simulation designed to investigate the effect of different kinds of management information on decision-making performance. All instructions, including how to enter responses, were presented on participants’ individual computer monitors to ensure that all participants received the same information. Players of this computer simulation adopt the role of a furniture factory’s production order manager who is presented with a weekly order of furniture items along with a listing of available employees for that week. The manager is responsible for allocating employees to jobs and motivating them so that they complete the weekly special order in as short a time as possible.

![Figure 4.4. Screen image from Week 1 of the Furniture Factory computer Simulation.](image)
The weekly production orders include manufacturing new furniture items, for example, chairs, repairing old ones, and restoring antiques. The participants manage their employees for a total of 18 production orders, with each order representing a performance trial in the simulation. Each trial consisted of a different production order and participants took between 1 to 1 ½ hours to complete all 18 trials. Production orders are broken down into 5 specific jobs such as assembling the timber, upholstering the furniture, and preparing the products for shipment. Participants are given estimated times indicating how long each specific job usually takes. These times are based on the performance of managers and MBA students on the simulation, whose scores were used to norm the task. Participants were also presented with the overall time that a completed order should take, this was calculated by simply summing the times estimated for each of the 5 specific jobs required for the work order. The main aim of the participant is to ensure that the different weekly orders are produced as efficiently as possible. Efficient production is achieved by minimising the amount of time employees spend on their allocated job in comparison to estimated times.

In addition to instructions outlining their role and main aims as the Furniture Factory manager, participants were also given information regarding the skills and work-related preferences of each worker and the descriptions and requirements of each job. This information was presented during the instruction phase and was also provided as a reference sheet (see Appendix C) that the participant could refer to at any time during the game.

Consistent with other computer simulated tasks, the Furniture Factory is based on underlying mathematical algorithms representing decision rules that the participant must
infer as they learn from their actions across trials. Participants received no information regarding the mathematical model that determined individual employee performance, such as the number of hours taken to complete their assigned job. See Wood and Bailey (1985) for the mathematical model used to calculate the hours taken to complete the assigned furniture order on each trial.

After receiving the weekly production order, the first task of the participant was to allocate each of the five employees to one of the five jobs. A different production order was provided each week, however, the 5 jobs required to produce the order remained the same across the 18 simulated weeks. Similarly, the 5 employees available on the roster remained constant throughout the game. The jobs that were to be completed each week vary in their nature. Some are interesting, some are dull, and some require attention to detail. The employees on the roster also vary in their skills, experience, motivational level, preference for routine or challenging assignments, and standards of work quality. Following is a profile description of one of the employees:

Hilary has been with the company for only a few years. She is a seamstress and dressmaker by training, but was made redundant when her previous employer got into financial difficulties. She takes pride in her work, both in her furniture covers and the dressmaking and embroidery that occupy her evenings.

The participant’s managerial task was to allocate workers from a roster to the different jobs to complete the production order efficiently. By correctly matching employees to job requirements, participants could achieve a higher overall level of organisational performance than if employees were poorly suited to jobs. The same employee could
not be allocated to more than one job and no employee could be retrenched. However, employees could be reassigned to a different, and potentially more suitable job, if the manager decided the job-person-fit could be improved.

After workers were allocated to jobs, the second task of the participant managers was to set each worker a production target or goal from a set of options based on goal theory (Locke & Latham, 1990) that is: 1) completion in 75% of estimated time (a hard goal), 2) the estimated hours 100% of estimated time (a moderate goal), 3) 125% of estimated time (an easy goal), 4) an instruction to ‘do your best’ and 5) ‘no set goal’. Goal assignments influenced an employee’s performance according to the mathematical model underlying the simulation whereby continual setting of unattainable goals would lead to diminished motivation and slower job performance.

Once the production goals were set, the participant manager received a performance report on how well their group had performed on the most recent production order, this was the actual time that their work team took to complete their assigned jobs. Overall performance was simply a sum of the actual times taken to perform each of the 5 jobs that week. These actual times are endogenous variables that are computed by the simulation. The participant could compare individual employees’ actual job completion times to the estimated (standard) times and use this information to adjust their decisions so as to improve their work team’s performance. Note that the preset performance standard was set at a level that was difficult to accomplish. An example of a performance report for a week is: “Your department produced the special order in 143% of estimated time.” In this instance, the work team is not working efficiently, indicating poor performance. The performance report was presented in writing on the
computer screen for each performance trial and remained on the screen until it was updated for the next trial. In week 1, a mathematical model (Wood & Bailey, 1985) was used to calculate the hours taken to complete a production order on the basis of the adequacy of job assignment and production targets. From the second week onwards, actual job completion times are also influenced by two additional motivational factors: ‘feedback’ and ‘rewards’ assigned during the previous week. The options for these two decisions are described next.

After receiving feedback on the job completion times of each employee for that weekly order, the third task of the participant manager was to assign one of four feedback levels of specificity to each employee. The feedback options available were: 1. Tell the employee how their time taken to complete the job compares to the estimated time (outcome feedback). 2. Discuss the employee’s working methods with them but do not specify how their time taken to complete the job compares to the estimated time (process feedback). 3. Tell the employee how their time taken to complete the job compares to the estimated time and discuss their working methods with them (outcome and process feedback). 4. Do not give the employee any feedback at all (no feedback).

Finally, the participant managers fourth decision was to assign a reward to each employee for that week’s job performance. They could not distribute major rewards such as promotions or salary increases. However, there were 3 minor rewards that could be assigned: 1. Compliment the employee privately on their good performance (moderate reward). 2. Publicly recognise the employee by posting a memo in the tearoom acknowledging the employee’s contribution for that week (high reward). 3. No reward provided.
In summary, in order to manage the workgroup in the efficient production of orders, the participant made four decisions, corresponding to four exogenous variables, for each worker on each decision trial: 1. Job Assignment; 2. setting production targets; (after receiving a report on the performance of individual employees and the work group) 3. providing feedback; and 4. distributing rewards. They could choose to retain or alter any or all of their decisions across the 18 trials. Job assignment and goals influence the performance of employees in the week in which they are set, whereas feedback and rewards influence employee efficiency in the following week. Level of organisational complexity of the Furniture Factory simulation can be adjusted by altering the numbers of trials, employees, jobs, and degree of match between employees’ skills and job requirements (Wood, Bandura, & Bailey, 1990). The version of the Furniture Factory completed by participants in the current experiments was of moderate complexity with 20 exogenous decisions per trial (4 decisions for each of the 5 employees) and 5 endogenous variables per trial (job completion times for each of the 5 employees), giving a total of 25 variables.

4.2.7 Decision Rules for Furniture Factory Simulation

Decision rules for the four decision types according to (Goodman & Wood, 2006) are:

1. Employee job allocation: Assign each employee to a job based on the match between job and employee characteristics.

2. Goal: Give the high goal initially. Give the moderate goal after an employee performs very poorly for 2 consecutive weeks (<=20% worse than standard).
Give the high goal after an employee performs very well (>=standard). Giving no goal or the low goal is never optimal.

3. Feedback: Give outcome plus process feedback initially. Give only outcome feedback after an employee performs well (>= standard) for 3 consecutive weeks. Giving no feedback or only process feedback is never optimal.

4. Reward: Give an employee no reward for poor performance (< standard). Give the moderate reward when performance is close to standard (standard <= performance < 5% better than standard). Give the high reward for good performance >= 5% better than standard).

These decision rules represent the underlying mathematical algorithms of the system that explain connections between the variables. The greater the participants’ understanding of these causal connections, the more likely they are to achieve high complex problem solving performance in the Furniture Factory simulation.

4.2.8 Furniture Factory Measures: Performance

The standard hours for each order were the same for all participants (estimated score). The simulation model automatically calculated the number of production hours for each trial (actual score) on the basis of the participant’s job allocations and selection of motivational factors using the underlying mathematical algorithms of the system (Wood & Bailey, 1985). Performance scores were computed for each trial; by taking the actual number of hours that employees took to complete the job, divided by the number of hours that had been estimated for the job (these estimations were based on MBA students’ results in previous studies (Wood & Bailey). The formula used was:
Performance = (2 – (total actual score cycle / total estimated score cycle)) * 100

The 2- simply provides a figure in a positive direction for ease of reporting. As suggested by the test developer (Wood & Bailey, 1985), three blocks of trials were created for the purpose of summarising and managing the complex problem solving performance data. Block 1 included the performance data for trials 1 to 6, block 2 included the performance data for trials 7 to 12, and block 3 included the performance data for trials 13 to 18. Performance scores were averaged across the three blocks of 6 trials each.

4.2.9 Furniture Factory Measures: Strategy

No measures of strategy were built into the Furniture Factory system at the time this study was conducted. Therefore, a measure of strategy was designed for Study 1, using SPSS SYNTAX. A strategy score was calculated for each of the four decision categories: Employee job allocation, Goal, Feedback, and Reward. For example, if the employee job allocation in trial 1 was not equal to the employee job allocation in trial 2, then the participant had changed their decision between trials and was given a score of 1. In contrast, if the employee job allocation in trial 1 was equal to the employee job allocation in trial 2, then the participant had adopted a ‘no change’ strategy and was given a score of 0. This is a binary scoring system, 1 = ‘change’ and 0 = ‘change nothing’. Section 4.1.7 described these strategy types in more detail. Strategy scores were calculated for the 20 exogenous variables (manipulated directly by the participant
manager) over 18 trials of the simulation. Strategy scores were summed for each of the 5 decisions (e.g. 5 employees to allocate to 5 jobs) within the 4 decision types (e.g. Employee job allocation) per trial. Thus, for each participant, 4 strategy scores were calculated: Strategy (Employee job allocation); Strategy (Goal); Strategy (Feedback); and Strategy (Reward).

4.3 Procedure

Participants completed the battery over two 1 ½ sessions separated by 1 week. Prior to undertaking the tests, participants were informed that the study was confidential, were given a brief rationale for the experiment, and were asked to give their informed consent to participate. Prior to commencing each test, time was allowed for the participants to read the instructions specific to each test and to take themselves through the practice questions. Participants were encouraged to alert the experimenter if they had any questions regarding the instructions or any difficulty with the practice items, so that additional assistance could be provided. Participants were instructed to answer every question, even if they had to guess. The three sections of the battery comprised the eight cognitive abilities tests (60 minutes), the Dynamic Forecasting Questionnaire (DFQ, 30 minutes) and the Furniture Factory (90 minutes). Total testing time for the whole battery was 3 hours, although all time limits were generous and not exceeded in approximately 95% of cases. Participants were encouraged, if necessary, to take a break between tests, rather than during tests. All tests were presented on Dell Pentium 4 computers, with the exception of the DFQ, which is a paper and pencil test. Scrap paper was provided for participants to make notes for the Furniture Factory and DFQ tasks.
Scrap paper for note taking and the use of calculators were prohibited for the cognitive abilities tests. The order of presentation of cognitive abilities tests was described in Section 4.2.3. The order of the three tasks (cognitive abilities battery, DFQ, and Furniture Factory) administered in each test session, was counter balanced across groups of participants, to minimise order effects. There were up to eight individuals present during any one test session. Participants were seated apart to facilitate administration and privacy of responding. At the end of testing, participants were given a debrief explaining the main aims of the experiment and when they would be contacted with their results.

4.3.1 Statistical Analyses

All statistical analyses in this study were performed using subprograms from the Statistical Package for Social Scientists (SPSS, 2003) computer package. The initial aim of the data analysis was to establish the reliability and validity of the data obtained for each of the measures included in this study. Where possible, this was done by comparing data obtained in this study with that obtained in previous validation studies. The second stage of data analysis employed correlational and regression techniques to investigate the nature of the relationship between cognitive abilities and complex problem solving.
4.4 Results and Discussion

The measures of cognitive abilities and complex problem solving that are analysed here are described in detail in Section 4.2.2. Prior to main analyses, data screening procedures were executed on the variables. Specifically, variables were checked for accuracy of data entry. SPSS FREQUENCIES was used for evaluations of the assumptions of multivariate analyses. With the use of a $p < .001$ criterion for Mahalanobis distance, no outliers among cases were identified. No cases had missing data and no suppressor variables were found, $N = 130$. Examination of the correlation matrix reveals no evidence of multicollinearity and singularity among the independent variables. Residual scatterplots revealed that assumptions of normality, linearity, homoscedasticity and independence of residuals were met. Overall, the distributions of variables fit the assumptions of multivariate analysis. Confirmatory statistical techniques are not appropriate for this thesis, which takes an exploratory approach.

4.4.1 Descriptive Statistics for Cognitive Abilities Variables

Data screening procedures included the usual checks for normality, linearity, and multicollinearity. No problems were detected. The scores of the eight cognitive abilities tests were analysed separately and also scores on pairs of tests (e.g., Swaps and Triplets) were combined to give mean scores compatible with the intelligence factors from Gf-Gc theory, (e.g., Fluid Reasoning, Gf). The means, standard deviations, and Cronbach alpha reliability estimates for each of the 8 cognitive abilities tests and the 4 respective factor scores (Gf, Gc, Gq, and Gv) are reported in Table 4.1. The cronbach alpha estimates are based on the test manual for the cognitive abilities test battery.
(Stankov, 1999) because the test is computer scored, item level data was unavailable to calculate reliability estimates specifically for this sample. The normative sample, described in the test manual (Stankov, 1999,) is similar in terms of gender, age, and education to the university student sample in the present study.

Table 4.1

Summary Statistics for Cognitive Abilities Tests

<table>
<thead>
<tr>
<th>Cognitive abilities</th>
<th>α</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Swaps (Gf)</td>
<td>.97</td>
<td>81.98</td>
<td>14.82</td>
</tr>
<tr>
<td>2. Triplets (Gf)</td>
<td>.88</td>
<td>86.67</td>
<td>14.26</td>
</tr>
<tr>
<td>Fluid Reasoning (Gf)</td>
<td></td>
<td>83.87</td>
<td>12.16</td>
</tr>
<tr>
<td>3. Vocabulary Test (Gc)</td>
<td>.80</td>
<td>74.36</td>
<td>14.42</td>
</tr>
<tr>
<td>4. Proverbs Matching (Gc)</td>
<td>.70</td>
<td>59.31</td>
<td>15.72</td>
</tr>
<tr>
<td>Crystallised Knowledge (Gc)</td>
<td></td>
<td>66.50</td>
<td>14.24</td>
</tr>
<tr>
<td>5. Numerical Operations (Gq)</td>
<td>.95</td>
<td>83.10</td>
<td>13.30</td>
</tr>
<tr>
<td>6. Financial Reasoning (Gq)</td>
<td>.80</td>
<td>56.19</td>
<td>18.90</td>
</tr>
<tr>
<td>Quantitative Knowledge (Gq)</td>
<td></td>
<td>69.98</td>
<td>13.93</td>
</tr>
<tr>
<td>7. Line Length (Gv)</td>
<td>.68</td>
<td>65.63</td>
<td>15.40</td>
</tr>
<tr>
<td>8. Letter Spotting (Gv)</td>
<td>.74</td>
<td>56.58</td>
<td>10.71</td>
</tr>
<tr>
<td>Broad Visualisation (Gv)</td>
<td></td>
<td>61.98</td>
<td>10.72</td>
</tr>
</tbody>
</table>

N = 130

Note. Cronbach’s alpha reliability estimates are not derived from this sample.

It can be seen from Table 4.1 that internal consistency reliability estimates (alpha) for the different cognitive abilities tests ranged from .68 for Line Length (Gv) to .97 for Swaps (Gf). The Fluid reasoning tasks, Swaps (Gf) and Triplets (Gf), have relatively high mean scores compared with the other cognitive abilities tests. Although a ceiling effect is not evident, these tests were regarded as less difficult than the other cognitive abilities tests.
4.4.2 Descriptive Statistics for Complex Problem Solving Variables

For interest sake, internal consistency (Cronbach’s alpha) for the Furniture Factory was calculated, for interest sake, $\alpha = .99$. However, internal consistency reliability estimates are not an appropriate estimate of reliability for computer simulations, which consist of repeated trials of the same set of decisions. In contrast to traditional cognitive abilities tests, the items in computer simulation tasks are not independent. Test retest reliability and parallel forms reliability are possible with computer simulations, however, these methods (and their associated problems) were discussed in detail in Section 4.1.2.

The means, standard deviations, and Cronbach alpha estimates for the 2 complex problem solving tasks, the Furniture Factory simulation and the Dynamic Forecasting Questionnaire (DFQ), are reported in Table 4.2. The Furniture Factory performance scores are divided into 3 blocks with the average scores of 6 trials of the simulation per block. Furniture Factory Block 1 is the mean score of trials 1 to 6, Furniture Factory Block 2 is the mean score of trials 7 to 12, and Furniture Factory Block 3 is the mean score of trials 13 to 18.

Table 4.2

Summary Statistics for Complex Problem Solving Tasks

<table>
<thead>
<tr>
<th>CPS Tasks</th>
<th>$\alpha$</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furniture Factory Block 1</td>
<td></td>
<td>96.21</td>
<td>15.01</td>
</tr>
<tr>
<td>Furniture Factory Block 2</td>
<td></td>
<td>98.33</td>
<td>23.68</td>
</tr>
<tr>
<td>Furniture Factory Block 3</td>
<td></td>
<td>98.74</td>
<td>26.03</td>
</tr>
<tr>
<td>Dynamic Forecasting</td>
<td>.74</td>
<td>5.89</td>
<td>3.17</td>
</tr>
</tbody>
</table>

$N = 130$
It can be seen from Table 4.2 that the internal consistency reliability estimate (alpha) for the Dynamic Forecasting Questionnaire (DFQ) is .74, indicating moderate internal consistency. The mean score for the DFQ indicates that average performance was around 39.27% accuracy. A 2-tailed t-test for independent samples indicated that DFQ performance for females ($M = 5.55$, $SD = 2.86$) was not significantly different from DFQ performance for males ($M = 6.48$, $SD = 3.60$), $t(81.52) = 1.53$, $p > .05$. Mean scores indicate that both males and females found the test difficult with males scoring 43% accuracy and females scoring 37% accuracy. These results are consistent with the findings of Sterman and Sweeney (2005) who reported that participants from an elite business school had difficulty, 48% accuracy, on a complex graphical integration task (measuring understanding of the relation between the net flow into a stock and the slope of the stock trajectory). Thus, the prediction in Section 4.1.9 (Predictions of study 1), that participants would have low performance scores on the DFQ, was supported.

Both the DFQ (employed in Study 1) and the Bathtub/Cashflow tasks (employed by Sterman and Sweeney) were developed based on Naturalistic Decision Making (NDM) research. The mean scores on the Furniture Factory task indicate that participants’ performance improved from initial trials of the simulation, Block 1, $M = 96.21$, to the final trials of the simulation, Block 3, $M = 98.74$.

4.4.3 Correlations between Cognitive Abilities and Furniture Factory Trials

Table 4.3 presents the correlations between performance scores on 18 individual trials of the Furniture Factory simulation, corresponding to 18 simulated weeks, and specific components of intelligence, Gf, Gc, Gq, and Gv.
Table 4.3

Correlation Matrix of Cognitive Abilities and Furniture Factory Trials 1 to 18

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Gf</td>
<td>.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Gc</td>
<td>.26</td>
<td>.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Gq</td>
<td>.23</td>
<td>.14</td>
<td>.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Gv</td>
<td>.25</td>
<td>.11</td>
<td>.18</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>FF1</td>
<td>.03</td>
<td>.30</td>
<td>.18</td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>FF2</td>
<td>.00</td>
<td>.25</td>
<td>.11</td>
<td>.18</td>
<td>.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>FF3</td>
<td>.08</td>
<td>.32</td>
<td>.32</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>FF4</td>
<td>.05</td>
<td>.28</td>
<td>.30</td>
<td>.30</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>FF5</td>
<td>.07</td>
<td>.32</td>
<td>.30</td>
<td>.30</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>FF6</td>
<td>.08</td>
<td>.22</td>
<td>.22</td>
<td>.22</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
<td>.32</td>
</tr>
<tr>
<td>11.</td>
<td>FF7</td>
<td>.14</td>
<td>.19</td>
<td>.00</td>
<td>.08</td>
<td>.36</td>
<td>.57</td>
<td>.69</td>
<td>.68</td>
<td>.73</td>
<td>.84</td>
</tr>
<tr>
<td>12.</td>
<td>FF8</td>
<td>.10</td>
<td>.25</td>
<td>.10</td>
<td>.07</td>
<td>.44</td>
<td>.61</td>
<td>.66</td>
<td>.77</td>
<td>.83</td>
<td>.92</td>
</tr>
<tr>
<td>13.</td>
<td>FF9</td>
<td>.01</td>
<td>.21</td>
<td>.12</td>
<td>.07</td>
<td>.43</td>
<td>.63</td>
<td>.67</td>
<td>.76</td>
<td>.76</td>
<td>.87</td>
</tr>
<tr>
<td>14.</td>
<td>FF10</td>
<td>.02</td>
<td>.19</td>
<td>.07</td>
<td>.06</td>
<td>.45</td>
<td>.67</td>
<td>.67</td>
<td>.72</td>
<td>.79</td>
<td>.86</td>
</tr>
<tr>
<td>15.</td>
<td>FF11</td>
<td>.06</td>
<td>.17</td>
<td>.06</td>
<td>.06</td>
<td>.42</td>
<td>.60</td>
<td>.70</td>
<td>.69</td>
<td>.73</td>
<td>.83</td>
</tr>
<tr>
<td>16.</td>
<td>FF12</td>
<td>.07</td>
<td>.21</td>
<td>.08</td>
<td>.09</td>
<td>.41</td>
<td>.57</td>
<td>.65</td>
<td>.72</td>
<td>.76</td>
<td>.86</td>
</tr>
<tr>
<td>17.</td>
<td>FF13</td>
<td>.12</td>
<td>.24</td>
<td>.14</td>
<td>.07</td>
<td>.38</td>
<td>.52</td>
<td>.64</td>
<td>.70</td>
<td>.71</td>
<td>.85</td>
</tr>
<tr>
<td>18.</td>
<td>FF14</td>
<td>.08</td>
<td>.23</td>
<td>.01</td>
<td>.04</td>
<td>.37</td>
<td>.52</td>
<td>.67</td>
<td>.66</td>
<td>.74</td>
<td>.85</td>
</tr>
<tr>
<td>19.</td>
<td>FF15</td>
<td>.14</td>
<td>.24</td>
<td>.09</td>
<td>.09</td>
<td>.44</td>
<td>.58</td>
<td>.65</td>
<td>.70</td>
<td>.76</td>
<td>.87</td>
</tr>
<tr>
<td>20.</td>
<td>FF16</td>
<td>.12</td>
<td>.19</td>
<td>.08</td>
<td>.08</td>
<td>.38</td>
<td>.56</td>
<td>.66</td>
<td>.69</td>
<td>.76</td>
<td>.84</td>
</tr>
<tr>
<td>21.</td>
<td>FF17</td>
<td>.12</td>
<td>.21</td>
<td>.12</td>
<td>.08</td>
<td>.42</td>
<td>.56</td>
<td>.63</td>
<td>.67</td>
<td>.73</td>
<td>.83</td>
</tr>
<tr>
<td>22.</td>
<td>FF18</td>
<td>.12</td>
<td>.27</td>
<td>.08</td>
<td>.09</td>
<td>.44</td>
<td>.56</td>
<td>.66</td>
<td>.69</td>
<td>.76</td>
<td>.84</td>
</tr>
</tbody>
</table>

Table 4.3 continued below

<table>
<thead>
<tr>
<th></th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.</td>
<td>FF8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>FF9</td>
<td>.84</td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>FF10</td>
<td>.91</td>
<td>.88</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>FF11</td>
<td>.88</td>
<td>.90</td>
<td>.90</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>FF12</td>
<td>.91</td>
<td>.82</td>
<td>.90</td>
<td>.89</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>FF13</td>
<td>.89</td>
<td>.89</td>
<td>.90</td>
<td>.89</td>
<td>.90</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>FF14</td>
<td>.87</td>
<td>.80</td>
<td>.88</td>
<td>.86</td>
<td>.92</td>
<td>.94</td>
<td>.94</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>FF15</td>
<td>.91</td>
<td>.85</td>
<td>.82</td>
<td>.91</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>FF16</td>
<td>.84</td>
<td>.86</td>
<td>.87</td>
<td>.93</td>
<td>.87</td>
<td>.85</td>
<td>.90</td>
<td>.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>FF17</td>
<td>.87</td>
<td>.91</td>
<td>.90</td>
<td>.95</td>
<td>.88</td>
<td>.87</td>
<td>.90</td>
<td>.93</td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>FF18</td>
<td>.88</td>
<td>.85</td>
<td>.90</td>
<td>.90</td>
<td>.91</td>
<td>.90</td>
<td>.94</td>
<td>.95</td>
<td>.94</td>
<td>.94</td>
</tr>
</tbody>
</table>

N = 130
Note: Measures are underlined where p < .05
As shown in Table 4.3, there were no significant correlations between fluid reasoning (Gf) and Furniture Factory performance on any trials. This finding is inconsistent with the suggestion that general intelligence is the most important ability for real world complex problem solving (Gottfredson, 1997). However, this finding is consistent with Dörner’s (1996, p. 27) suggestion that ‘there is no significant correlation between scores on IQ tests and performance in any complicated problem solving experiment’. Recall that global IQ scores are traditionally measured by Gf tasks such as Raven’s Progressive Matrices. Findings similar to the lack of a correlation between complex problem solving and Gf, described here, are likely to have led Dörner and others to making such strong statements as the different cognitive demands hypothesis. In contrast to Dörner’s suggestion of zero correlations between complex problem solving and intelligence, Table 4.3 reports that there are significant correlations between complex problem solving and specific components of intelligence. For example, crystallised knowledge (Gc) significantly correlates with 16 out of 18 trials of the Furniture Factory task. The highest correlation is between Gc and complex problem solving performance on week 5 (trial 5) of the Furniture Factory ($r = .32, p < .05$).

In addition, there are significant correlations between broad visualisation (Gv) and Furniture Factory performance week 1 ($r = .30, p < .05$) and week 2 ($r = .18, p < .05$). The finding of a relationship between complex problem solving performance and Gv is consistent with the results of previous studies described in Section 4.1.5. For example, significant correlations were found between performance in FEUER and Gv (Schoppek, 1991). It has been suggested that Gv tends to load on the general factor (g) of intelligence, and that the observed correlation between Gv and complex problem
solving (CPS) performance in Schoppek’s study was in fact a correlation between g and CPS (Lohman, Pellegrino, Alderton, & Regian, 1987). If this were the case, then the observed correlation in Study 1, between Gf and CPS, should have been higher than the correlation between Gv and CPS. However, this was not the case and we can conclude that broad visualisation ability (Gv) is a predictor of complex problem solving performance during initial trials of the Furniture Factory simulation. This result is consistent with previous findings (Rigas, Carling, & Brehmer, 2002), where the correlation between NEWFIRE and Gf was lower than the correlation between NEWFIRE and Gv. This result is also consistent with the finding that spatial working memory capacity (WMC) is the only WMC task to load on complex problem solving performance (Wittmann & Süß, 1999). WMC spatial is assessed by the Letter Spotting (Gv) task employed in Study 1. Finally, Table 4.3 also shows a moderate correlation (r = .18, p < .05) between quantitative knowledge (Gq) and week 1 (trial 1) Furniture Factory performance.

Overall, as predicted in Section 4.1.9 (Predictions of study 1), the correlations between performance on the Furniture Factory task and specific components of intelligence (Gc, Gv, and Gq) do not support the different demands hypothesis because a relationship has been established. These findings do support the use of Brunswik symmetry in this area of research (Wittmann & Süß, 1999), which suggests employment of specific cognitive abilities tasks and consideration of all trials of simulations, rather than searching for a relationship between a global IQ score and a single performance score based on the last trial of a simulation.
4.4.4 Correlations between Cognitive Abilities and Complex Problem Solving Tasks

Correlations among summary performance scores on the Furniture Factory complex problem solving task; the 4 cognitive abilities factors (Gf, Gc, Gq, and Gv); and gender are presented in Table 4.4. Gender scores are binary, male = 1, female = 0. Therefore, negative correlations with gender, indicate a gender bias in favour of females. While positive correlations with gender, indicate a gender bias in favour of males.

Table 4.4

**Correlation Matrix of Cognitive Abilities and Complex Problem Solving Tasks (the Furniture Factory and the Dynamic Forecasting Questionnaire)**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gf</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Gc</td>
<td>.34</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Gq</td>
<td>.26</td>
<td>.15</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gv</td>
<td>.23</td>
<td>.14</td>
<td>.28</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. FF Block 1</td>
<td>.07</td>
<td>.30</td>
<td>.08</td>
<td>.14</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. FF Block 2</td>
<td>.07</td>
<td>.21</td>
<td>.08</td>
<td>.08</td>
<td>.91</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. FF Block 3</td>
<td>.12</td>
<td>.24</td>
<td>.09</td>
<td>.08</td>
<td>.87</td>
<td>.97</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>8. DFQ</td>
<td>.21</td>
<td>.19</td>
<td>.43</td>
<td>-.01</td>
<td>.10</td>
<td>.10</td>
<td>.08</td>
<td>1.00</td>
</tr>
<tr>
<td>9. Gender</td>
<td>-.01</td>
<td>-.13</td>
<td>.26</td>
<td>-.17</td>
<td>-.17</td>
<td>-.09</td>
<td>-.07</td>
<td>.14</td>
</tr>
</tbody>
</table>

_N = 130_

**Note:** Measures are underlined where _p < .05_

It can be seen from Table 4.4 that summary score of Furniture Factory performance are significantly correlated with crystallised knowledge (Gc) across all 3 Blocks of trials. The correlation between Furniture Factory performance Block 1 (simulated weeks 1 to 6) and Gc is significant (r = .30, _p < .05_). The corresponding correlations for Block 2 and Block 3 were also significant, (r = .21, _p < .05_) and (r = .24, _p < .05_) respectively. These findings are consistent with the suggestion that knowledge acquisition and
application, in complex problem solving tasks, is dependent on a certain level of intelligence (Cattell, 1971), this knowledge may be highly correlated with Gc. The correlation between Furniture Factory performance and Gc may also be mediated by task properties. For example, the Furniture Factory task involves a vast amount of verbal instructions including verbal descriptions of employees who must be matched to appropriate jobs (Appendix C).

Correlations between summary performance scores (Block 1 to Block 3) on the Furniture Factory and Gf, Gq, and Gv, were not significant. However, recall from Table 4.3, that a trial-by-trial analysis of Furniture Factory performance, revealed significant correlations between initial trials of the simulation, with Gq and Gv. This again highlights the importance of scoring for complex computerised scenarios and is consistent with previous scoring suggestions based on Brunswik symmetry (Wittmann & Süß, 1999). Single overall performance scores based on the final trial of a simulation are not sufficiently detailed to allow investigation of relationships between complex problem solving competencies and intelligence.

A 2-tailed t-test for independent samples indicated that females (M = 98.21, SD = 14.59) had significantly higher performance scores on FF Block 1 than males (M = 92.79, SD = 15.24), t(128) = -2.01, p < .05. A 2-tailed t-test for independent samples indicated that FF Block 2 performance for females (M = 100.01, SD = 22.65) was not significantly different from FF Block 2 performance for males (M = 95.47, SD = 25.33), t(128) = -1.06, p > .05. A 2-tailed t-test for independent samples indicated that FF Block 3 performance for females (M = 100.19, SD = 25.21) was not significantly different from FF Block 3 performance for males (M = 96.27, SD = 27.46), t(128) = -
.83 , p > .05. Thus, females outperformed males on FF Block 1. However, no significant gender differences were observed for FF Blocks 2 and 3.

Gender differences in Furniture Factory performance are unlikely to be explained in terms of a gender bias in the semantic details (company management) of the cover story (Funke & Hussy, 1984). In the 1940s, both Terman and Wechsler found that females outscored males on some verbal tests. A meta-analysis of some 150 studies of verbal abilities (Hyde & Linn, 1988) found that females outscored males by 3.5 IQ points (in studies prior to 1973), and females outscored males by 1.5 IQ points in subsequent studies. Although the gender gap in verbal competencies appears to have decreased, there is still evidence that young girls learn to talk sooner than boys, and have acquired a larger vocabulary by the age of 2 or 3 (Fenson et al. 1994). However, the correlation between gender and Gc presented in Table 4.2.6, is not significant (r = -.13, p > .05). Although this correlation is not significant, it is negative which implies female superiority on verbal tasks. However, no conclusions can be drawn from this non significant correlation. Perhaps the gender difference, favouring females, on initial trials of the Furniture Factory simulation is mediated by an interaction of superior female verbal abilities and some other skills that have not been measured in this study. For example, the gender difference, favouring males, on Tailorshop performance was suggested to be mediated by male superiority of economics knowledge and computer literacy (Süß, Oberauer, & Wittmann, 2005; Wittmann & Süß, 1999). These findings support the hypothesis, in Section 4.1.9 (Predictions of study 1), that gender differences on complex problem solving tasks were more likely to be due to gender differences in skills, rather than due to gender bias in the cover stories of the computer simulations.
Correlations among summary performance scores on the 2 complex problem solving tasks, the Furniture Factory and the Dynamic Forecasting Questionnaire; the 4 cognitive abilities factors (Gf, Gc, Gq, and Gv); and gender are presented in Table 4.4.

Table 4.4 also shows that the two complex problem solving tasks, the Furniture Factory and the Dynamic Forecasting Questionnaire (DFQ) are not significantly correlated, for Furniture Factory Block 1 ($r = .10, p > .05$), for Block 2 ($r = .10, p > .05$), and for Block 3 ($r = .08, p > .05$). This lack of a correlation between the two complex problem solving tasks may be due to a lack of symmetry between the DFQ, which is a paper and pencil task, and the Furniture Factory, which is a computer simulation. Shared variance amongst complex problem solving tasks will be further explored in Study 2.

As shown in Table 4.4, there are significant correlations between complex problem solving performance on the DFQ and specific components of intelligence, namely, fluid reasoning (Gf) ($r = .21, p < .05$), crystallised knowledge (Gc) ($r = .19, p < .05$), and quantitative knowledge (Gq) ($r = .43, p < .05$). No significant correlations were observed between DFQ performance and broad visualisation (Gv), ($r = -.01, p > .05$). The hypothesis in Section 4.1.9 (Predictions of study 1) that there would be a correlation between quantitative knowledge (Gq) and performance on the DFQ was supported. The hypothesis that acculturation knowledge (Gc) would correlate with performance on the Furniture Factory (FF) simulation was also supported. Overall, Gq represents the cognitive abilities factor that is the strongest predictor of DFQ performance and Gc represents the cognitive abilities factor that is the strongest predictor of FF performance.
4.4.5 Regression of Dynamic Forecasting Questionnaire Performance on Cognitive Abilities Factors

A stepwise regression was performed, with Dynamic Forecasting Questionnaire (DFQ) performance as the dependent variable and cognitive abilities factors as the independent variables. Independent variables included: Fluid Reasoning (Gf), Crystallised Knowledge (Gc), Quantitative Knowledge (Gq), and Broad Visualisation (Gv). Table 4.5 displays R, R², and adjusted R² after simultaneous entry of the four cognitive abilities tests.

Table 4.5
Regression of Dynamic Forecasting Questionnaire Performance on Cognitive Abilities Factors

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>R</th>
<th>R²</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative Knowledge (Gq)</td>
<td>.46</td>
<td>.21</td>
<td>.21</td>
</tr>
</tbody>
</table>

N = 130

The regression shown in Table 4.5 shows that the only the independent variable, Quantitative Knowledge (Gq), has been entered into the equation. The other independent variables: Fluid Reasoning (Gf), Crystallised Knowledge (Gc), and Broad Visualisation (Gv), failed to meet the selection criteria. R² was significantly different from zero with Gq in the equation, R² = .21 (adjusted R² = .21), F(1,122) = 32.71, p < .01. Therefore, Quantitative Knowledge (Gq) explains 21% of the variability in Dynamic Forecasting Questionnaire performance. Whereas, Fluid Reasoning (Gf), Crystallised Knowledge (Gc), and Broad Visualisation (Gv), do not account for any
significant proportion of the variability in DFQ performance, over and above what Gq explains.

Overall, it appears that there may be some overlap between traditional measures of Gq and complex problem solving performance on the DFQ. This finding may violate the premise described in Section 4.1.9 (Predictions of study 1) that complex problem solving tasks and intelligence tests must not overlap at theoretical or operational levels (Wenke, Frensch, & Funke, 2005).

4.4.6 Gender Differences

No gender effects were observed for DFQ performance. The correlation between gender and DFQ performance was not significant ($r = .14, p > .05$). A 2-tailed t-test for independent samples indicated that DFQ performance for females ($M = 5.55, SD = 2.86$) was not significantly different from DFQ performance for males ($M = 6.48, SD = 3.60$), $t(81.52) = 1.53, p > .05$. Despite the fact that males had significantly higher Gq scores than females ($r = .26, p < .05$) and Gq accounts for 21% of the variance in DFQ performance.

Previous research suggests that male superiority on mathematics is only at extremes of the distribution (Tyler, 1965). Males are over represented among high achievers in numerical reasoning and mathematics. As a consequence, there are more males than females with particularly high scores in mathematics although the mean scores are similar (Tyler). Thus, the finding of higher male than female performance scores on the
Gq tasks employed in Study 1, may have been due to outliers in the sample, and does not translate to gender differences in DFQ task performance.

4.4.7 Group Differences

In addition to gender differences, group differences were also investigated. A t-test was calculated to investigate group differences on the Furniture Factory task. Complex problem solving competencies, assessed by Furniture Factory performance, was compared for the undergraduate student sample in Study 1 of this thesis and for the MBA student scores previously collected (Wood & Bailey, 1985) to norm the Furniture Factory task. The following t-test is calculated using raw performance scores because the adjusted mean performance scores, reported in Table 4.2 earlier, are based on previous studies using an MBA student sample (Wood & Bailey, 1985), as described earlier in Section 4.2.8. A 2-tailed, one sample, t-test indicated that Furniture Factory, raw score performance, Block 3, was not significantly different for the undergraduate psychology student sample (M = 117.58, SD = 30.22) than for the MBA student sample (M = 116.3833, SD = 38.23), t(129) = .45, p > .05. Thus, the hypothesis (from Section 4.1.9) that MBA students would outperform undergraduate students on the Furniture Factory task was not supported.

This finding is inconsistent with previous suggestions that knowledge of a specific domain would improve complex problem solving performance if the cover story of the computer simulation were consistent with the knowledge base of the participant (Hesse, 1982; Putz-Osterloh, 1993; Stanley, Mathews, Buss, & Kotler-Cope, 1989). The results of Study 1 suggest that, more than just the semantic domain of the cover story of the
computer simulation may mediate group differences. It has also been suggested that prior knowledge in the domain of the complex task may hinder exploration of the task in order to learn the underlying rules. Participants with prior knowledge tend to engage in less exploration and consequently, learn fewer of the underlying causal connections between the variables in complex tasks (Goodman, Wood, & Hendrickx, 2004; Krems, 1995). Perhaps the advantage of prior management knowledge that MBA students possess is counteracted by the fact that they engage in decreased exploration strategies compared to their undergraduate student counterparts, although there are many other possibilities for further investigation.

4.4.8 Exploration Strategy

Exploration strategy in the Furniture Factory task was investigated in Study 1 and results are presented in Table 4.6. Decision rules underlying the Furniture Factory simulation were presented above in Section 4.2.7. Game review questions are presented in Appendix B. The calculation of strategy scores was described in Section 4.2.9.

Table 4.6

Correlation Matrix of Furniture Factory Performance and Furniture Factory Strategy

<table>
<thead>
<tr>
<th>Furniture Factory Performance</th>
<th>Strategy (Employee)</th>
<th>Strategy (Goals)</th>
<th>Strategy (Feedback)</th>
<th>Strategy (Rewards)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF Block 1</td>
<td>-.83</td>
<td>-.53</td>
<td>-.53</td>
<td>-.33</td>
</tr>
<tr>
<td>FF Block 2</td>
<td>-.90</td>
<td>-.52</td>
<td>-.52</td>
<td>-.29</td>
</tr>
<tr>
<td>FF Block 3</td>
<td>-.93</td>
<td>-.52</td>
<td>-.48</td>
<td>-.29</td>
</tr>
</tbody>
</table>

N = 130

Note: Measures are underlined where p < .05

It can be seen from Table 4.6 that all correlations between Furniture Factory performance and the ‘change nothing’ strategy (described in the last paragraph
Section 4.1.7) are significant, ranging from the lowest correlation ($r = -0.29$, $p < .05$) for the decisions regarding allocation of rewards to employees, in Furniture Factory Blocks 2 and 3, to the highest correlation ($r = -0.93$, $p < .05$) for decisions regarding allocation of jobs to employees in Furniture Factory Block 3.

Causal connections between the variables are represented in a mathematical model underlying the Furniture Factory system. As described by the decision rules presented in Section 4.2.7, the main aim of the Furniture Factory task is employee job allocation. Participant managers are required to assign each employee to a job based on the match between job and employee characteristics. Table 4.6 suggests that adopting a ‘change nothing’ strategy between trials, regarding the allocation of employees to jobs (i.e., allocate the same employee, the same job, each trial), is roughly equivalent to achieving successful performance on the Furniture Factory simulation. This finding is consistent with the significant correlation described earlier between crystallised knowledge and Furniture Factory performance. Recall that the Furniture Factory game is characterised by large amounts of verbal instructions, including verbal descriptions of employees and jobs (Appendix C).

The correlation matrix for accuracy of participants’ responses to the 24 game review questionnaires (Appendix B) and Furniture Factory performance was also examined. This relatively large correlation matrix is not presented; instead, pertinent results will be described. The Furniture Factory game review question that has the highest correlation with Furniture Factory performance was game review question 10 ($r = 0.39$, $p < .05$). Game review question 10 states: ‘It is a good idea to move people regularly between jobs at the Furniture Factory.’ The correct answer is: ‘False.’ This significant correlation also suggests that participants who had learnt the most important underlying
rule of the simulation (i.e., to keep people in the same jobs across the simulated weeks) consequently adopted a ‘change nothing’ strategy. The ‘change nothing’ strategy led to high complex problem solving performance on the Furniture Factory task. The finding that the ‘change nothing’ strategy has an almost perfect correlation with successful performance on the Furniture Factory task raises some problems for interpretation of results. Furniture Factory performance is confounded by exploration strategy on the same task. Therefore, we are unable to discern whether there is support for the hypothesis (presented in Section 4.1.9) that exploration strategy on the Furniture Factory task would correlate with specific components of intelligence. On the Furniture Factory task, the relationship between exploration strategy and specific components of intelligence is roughly the same, as the relationship between complex problem solving performance and specific components of intelligence. Elaboration of Study 1 findings, limitations of the present study, and suggestions for future research will now be presented.

4.5 Conclusion: Study 1

The results from Study 1 provide further support for the extension of Kröner, Plass, and Leutner’s (2005) model illustrated in Figure 4.1. The path from intelligence to knowledge application (assessed by performance on a complex task) was only significant in the transparent version of MULTIFLUX, in the Kröner et al. study. However, by extending their model to include specific components of intelligence, the results of Study 1 suggest that the paths from specific cognitive abilities to complex problem solving performance are significant, even under the absence of manipulating
qualities of complex problem solving tasks, such as complexity, connectivity, intranparency, dynamics, and polytely. For example, crystallised knowledge (Gc) is a predictor of Furniture Factory simulation performance. In addition, quantitative knowledge (Gq) is a predictor of performance on the Dynamic Forecasting Questionnaire (DFQ).

It is possible to construct new complex tasks in order to have strong correlations with intelligence, for example MULTIFLUX (Kröner, Plass, & Leutner, 2005). It is also possible to manipulate the qualities that define complex tasks, such as transparency (Kröner et al., 2005) or goal specificity (Burns & Vollmeyer, 2002; Vollmeyer, Burns, & Holyoak, 1996), in order to increase the correlation between intelligence and complex problem solving. However, there are serious problems that can arise by employing either of these two approaches. Firstly, deliberately designing a complex problem solving task in order to increase correlations with intelligence may violate the premise that complex problem solving tasks and traditional intelligence tests must not overlap at theoretical or operational levels (Wenke, Frensch, & Funke, 2005). In Study 1, quantitative knowledge (Gq) accounted for 21% of the variance in performance on the DFQ task. If the DFQ task was, in fact, simply a Gq test, then Wenke et al.’s premise would be violated. Consequently, the DFQ could not be considered a true complex problem solving task. In addition, by manipulating the qualities of a complex task, such as transparency, in order to increase correlations with intelligence, researchers are, in fact, reducing the differences between a complex task and a traditional intelligence test. Consequently, the complex task could lose the very 5 qualities that define it.
The correlation between the two complex tasks employed in Study 1, the Furniture Factory and the Dynamic Forecasting Questionnaire, was not significant. Study 2 of this thesis will employ three complex computerised scenarios, the Furniture Factory, the Forestry system (FSYS 2.0), and the Tailorshop, in order to further investigate shared variance amongst complex problem solving tasks.

The specific components of intelligence employed as predictors of complex problem solving in Study 1, were based on Gf-Gc theory. Results from our study are compatible with previous research which employed specific cognitive abilities tasks based on the Berlin Intelligence Structure (BIS) model. For example, Study 1 found significant correlations between a broad visualisation (Gv) factor, derived from Gf-Gc theory, and Furniture Factory performance. Similarly, previous researchers have found significant correlations between a Gv factor, derived from the BIS model, and complex problem solving performance on the Tailorshop task (Wittmann & Süß, 1999). In Study 2 of this thesis, further comparisons will be made between previous research, employing tests from the BIS (Süß, Oberauer, & Wittmann, 2005; Wittmann & Süß, 1999) and the current research, which employs tests from Gf-Gc theory. The number of cognitive abilities tests will be increased in Study 2, so that factor analytic techniques can be employed to further explore the relationship between intelligence and complex problem solving ability.

The finding of significant correlations between specific components of intelligence and complex problem solving competencies in Study 1 provides further support for the application of the Brunswik lens model to complex problem solving research (Wittmann & Süß, 1999). There is more symmetry between complex problem solving and specific
cognitive abilities, than with a global IQ score alone. Similarly, aggregated scoring techniques were employed to assess complex problem solving performance, Previous research has tended to rely on a single performance score, on a single last trial of a simulation. In study 1, the correlation between broad visualisation (Gv) and complex problem solving performance was only apparent in the initial trials of the simulation. This relationship would have gone unnoticed if an overall performance score were used, instead of trial-by-trial data analysis. In addition to the employment of specific cognitive abilities tests as predictors, aggregation of scores across multiple trials of a complex computerised scenario, as criteria measures, leads to increased symmetry between predictor and criterion. The affect of different scoring techniques, on the relationship between intelligence and complex problem solving, will be further explored in Study 2.

The observed correlation between performance on the DFQ stocks and flows task and Gq in Study 1, is inconsistent with Sterman and Sweeney’s (2005) findings that maths ability, assessed by level of university education in science subjects, did not predict performance on the Bathtub and Cashflow stocks and flows tasks. It is also quite ironic that paper and pencil stocks and flows tasks are employed in Naturalistic Decision Making (NDM) research, which by its very definition has a field study focus and claims to be real world orientated. However, the DFQ task was in fact more similar to a traditional cognitive ability test (Gq) than to a complex problem solving task such as the Furniture Factory. This highlights the need to compare tasks and synthesise results from both individual differences and NDM fields in complex problem solving research.
Despite the differences in predictor measures between the Sterman and Sweeney study and Study 1 of this thesis, participants in both studies experienced great difficulty in the criterion measure of performance on the stocks and flows tasks. Advocates of the naturalistic decision making (NDM) approach have suggested that complex problem solving in the real world (e.g., filling up a bath tub) is an easy task because of the evolution of heuristics that are ecologically rational and adapted to the real world environment (Gigerenzer et al., 1999). However, in lab settings participants struggle to understand stocks and flows tasks because their heuristics are context specific and cannot be transferred to laboratory situations, such as the Dynamic Forecasting Questionnaire. Even though the logical structure underlying stocks and flows situations is the same for both filling a bathtub and predicting death rates in the DFQ. The difficulty with the DFQ task experienced by participants in Study 1 is consistent with the evolutionary perspective. It is not necessary to understand stocks and flows when filling a bathtub in the real world. When the water reaches the top of the bathtub, there is a simple, first order negative feedback process to turn off the tap (Sterman & Sweeney, 2005). The stocks and flows tasks employed in both Study 1 and Sterman and Sweeney’s study, were static paper and pencil tasks. These tasks did not allow participants to use a ‘wait and see’ strategy before taking an action. This ‘wait and see’ strategy is very effect for real world complex problem solving (Diehl & Sterman, 1995). Thus, the static nature of the stocks and flows tasks led to reduced ecological validity. Study 2 of this thesis, will abandon the use of paper and pencil complex problem solving tasks. Dynamic computerised complex tasks will be employed in order to assess complex problem solving under conditions of increased ecological validity.
Complex problem solving research is a relatively new area that was initiated in the early 1970s (Wenke, Frensch, & Funke, 2005). As such, there has been very little research conducted regarding the influence of person characteristics, such as age and gender, in complex problem solving performance. The samples have been largely homogenous, which makes person characteristics, such as age effects, difficult to investigate. Some, albeit very little, previous research has been conducted to explore gender differences on complex problem solving tasks. A gender difference, favouring males, was found on the Tailorshop computer simulation (Süß, Oberauer, & Wittmann, 2005; Wittmann & Süß, 1999). Although not a main aim of this thesis, gender differences on complex problem solving tasks were also explored in Study 1. Results suggest that females significantly outperform males on the Furniture Factory computer simulation. Some researchers have proposed that the future of intelligence testing will utilise computer simulations which have greater ecological validity than traditional intelligence tests (Kröner, Plass, & Leutner, 2005). If this pursuit is to be taken seriously, then appropriate tables will need to be developed in order to norm participants’ scores to take gender and age effects into consideration. These considerations must be kept in mind, whether these tools are to be used for assessment, recruitment, or training purposes. Gender differences on complex problem solving tasks will be further explored in Study 2. In addition, the effect of personality and career interest on complex problem solving performance will be investigated.

The cover story of a complex problem solving task should be highly relevant for assessment if designing complex computer simulated scenarios for training or recruitment purposes (Brehmer & Dörner, 1993). However, it can cause problems in studies of complex problem solving performance, as a general ability, where the cover
story should be of little intrinsic interest. The comparison of MBA students with undergraduate students, on the Furniture Factory task in Study 1 revealed no difference between complex performance scores for the two groups. Results could simply be interpreted as suggesting that the cover story does not affect performance on the Furniture Factory task. However, there are several conflicting reasons that confuse the issue of group differences based on the semantic domain of the cover story. Firstly, MBA students would have possessed superior knowledge in the semantic domain of the cover story (i.e., management) than the undergraduate students. This prior knowledge should have been advantageous to MBA student performance on the complex task (Hesse, 1982; Putz-Osterloh, 1993; Stanley, Mathews, Buss, & Kotler-Cope, 1989). Secondly, previous findings suggest that participants who rely more on prior knowledge (i.e., MBA students) are less likely to engage in exploration strategies to discover the underlying rules of the complex system (Goodman, Wood, & Hendrickx, 2004; Krems, 1995). In the results section it was suggested that the first and second point just raised, might counteract one another, leading to no difference between MBA and undergraduate student performance. However, further analysis of results of Study 1 revealed that a ‘no change’ strategy, which involves minimal task exploration, was advantageous for Furniture Factory performance. These conflicting explanations of causation for group differences (including gender differences) highlight the fact that a ‘cover story’ explanation is not as straightforward as suggested by previous researchers. Many elements of complex problem solving performance must be taken into consideration simultaneously when attempting to find causal explanations.

The finding that the ‘change nothing’ strategy has an almost perfect correlation with successful performance on the Furniture Factory task raises some problems for
interpretation of results. For example, there are problems differentiating participants with high performance on the Furniture Factory task, due to their knowledge of the underlying rules of the game, from successful participants who discovered the ‘no change’ loophole by accident, or those who were simply responding in a ‘donkey vote’ type fashion. First of all results on the Furniture Factory task of the present sample are comparable to the results achieved by MBA students. Secondly, mean performance scores on the cognitive abilities tests also indicate that the sample in Study 1 was indeed motivated to perform their best on all tasks and did not engage in a ‘donkey vote’ response strategy. The fact that some participants may have discovered the ‘change nothing’ strategy as a loophole, rather than truly understanding the underlying rules of the simulation remains a concern. Unfortunately, there is no way of identifying those participants who discovered the ‘loophole’ in the Furniture Factory task. The presence of loopholes can be attributed to the mathematical algorithms underlying complex tasks. Although exploiting loopholes may be an advantageous strategy in real world complex problem solving, it is not a condition that psychometricians want in computer simulations (Mané & Donchin, 1989), where the focus is assessment of complex problem solving competencies (i.e., learning the underlying rules of the task). ‘Loopholes’ are rarely discussed in research employing computer simulations. However, as mentioned in Section 3.2, participants discovered a loophole in the SPACE FORTRESS simulation. In this previous research, the participants disregarded the task instructions and found a loophole that they could exploit to win the game by means other than which the task was designed (Mané & Donchin, 1989). Study 2 of this thesis will investigate the presence of ‘loopholes’ in other computer simulated tasks in an attempt find out what proportion of participants exploit loopholes in computer simulations.
CHAPTER 5
INTRODUCTION: STUDY 2

5 Introduction

The gap between field research and laboratory research has always been a problem in psychology. With the advent of computers into the laboratory, computer simulated tasks allowed the observation of complex problem solving performance in the laboratory with a higher degree of ecological validity than ever before. Intelligence testing has the slowest evolution rate of all major technologies (Sternberg & Kaufman, 1996) due to sound psychometric properties including high predictive validity of traditional intelligence tests, the belief they are satisfactory in their current form, the high financial investment required for development of new tests, and the validation of new tests against traditional ones. Studies 1 and 2 of this thesis address the latter issue. A question that has plagued researchers since the inception of computer simulations into the laboratory is: What personal qualities (abilities, personality) determine success on simulations? So far, research in this area has been rather limited (Rigas & Brehmer, 1999).

Much of the previous research reviewed in this chapter regarding the relationship of complex problem solving in computer simulated tasks with intelligence and personality was, or still remains, unpublished and was discovered via correspondence with the respective (mostly German) researchers after data collection for Study 2 had taken place. The design of Study 2 was purely exploratory and was not guided by previous findings, which are generally inconsistent. Given the lack of research in this area, the
aim of Study 2 was not to extend previous findings. Rather, the focus of the study was
to bring individual differences research in the area a step closer to obtaining stable
results from which generalisations about complex problem solving tasks can be made.
In addition, Study 2 involved the development of a new consistent scoring technique,
which made comparison between performance scores on different computer simulations
possible. Factor analytic techniques were employed to analyse computer simulation
performance, there is no evidence of application of factor analytic techniques to
simulation performance in previous research. Study 2 also involved translation, from
German to English, of the Forestry System (Wagener & Conrad, 1996) computer
simulation.

5.1 Scoring Issues in Complex Problem Solving Research: Brunswik Symmetry

Validity of various predictors in order of highest to lowest predictive validity for job
performance are: cognitive ability composite, job tryout, biographical inventory,
reference check, experience, interview, rating of training success, academic
achievement, education, interest inventory, age (Hunter & Hunter, 1984). The best
predictors of performance are cognitive abilities, more specifically, hierarchical models
of intelligence (Carroll, 1993), even better than an actual job tryout. The greater the
symmetry between predictors and criteria, the higher the predictability of the
psychometric test (Wittmann & Süß, 1999).
Wittmann and Süß (1999) define Brunswik symmetry as:

Placing hierarchical models on one side of the lens, gestalt principles force researchers to look at the criterion side for symmetrical hierarchical models as well…we postulate that the true latent structures of psychological constructs have to be symmetrical. We coined this term Brunswik symmetry. Every level of generality in the predictor model has its symmetrical level of generality at the criterion side, and vice versa. If we could only know and adequately assess the respective components, all correlations in pairs between symmetrical levels would be perfect (i.e., one). (p. 79)

Brunswik symmetry can be applied to previous research regarding predictive validity of psychometric tests. For example, (Mischel, 1968) conducted a literature review investigating the prediction of behaviour from broad personality constructs and found an average coefficient around .30. These coefficients cast doubt on the predictive validity of personality for explaining human behaviour. However, (Epstein, 1980) suggested that the coefficients, reported in Mischel’s review, were confounded by the fact that the criteria of human behaviour was assessed by single acts of low reliability.

Aggregation of repeated single act criteria increases the proportion of systematic variance that can be predicted, leading to more reliable and broad criteria (Fishbein & Ajzen, 1975; Wittmann & Süß, 1999). In the real world, managerial job performance is unlikely to be assessed by a single act; it is more likely to be assessed by company performance over the entire financial year. The same logic can be applied to complex problem solving performance. Aggregated performance across multiple trials of a
computer simulation leads to broader and more reliable criteria that can be better predicted by broad constructs such as cognitive abilities and personality. Thus, in Study 2, aggregated scoring techniques were employed in addition to the final trial performance scores that have been predominantly used in past research (Dörner, 1996). In accordance with Brunswik symmetry, in Study 2 it is expected that correlations between intelligence and complex problem solving will be increased by the use of specific cognitive abilities tests and aggregated computer simulation scores, rather than general or factor scores of intelligence and final trial computer performance simulation scores.

5.1.1 Goal achievement scores

In addition to final trial scores and aggregated scores over multiple trials of a simulation, a new computer simulation performance score (goal achievement) was developed and employed to investigate the relationship between intelligence and complex problem solving. Goal achievement scores indicate the average percent of trials on which participants achieved the goal of the task. Development of these scores are consistent for each of the three computer simulations employed in Study 2 (Furniture Factory, Tailorshop, and Forestry System) and facilitate direct comparison between performance on different computer simulations. Goal achievement scores also offer a way to measure the gap (barrier) between the problem state of the task and the goal state that the participant is trying to achieve. The gap definition of complex problem solving was described in detail in Section 1.1.1.
The gap between the problem state and the goal state is different across computer simulated tasks. For example, in the Furniture Factory task, the participant-manager finds the company in a good state, with most employees easy to motivate and a couple who present quite a challenge. Participants must try to exceed performance scores set by a normative sample, thus, the Furniture Factory task is characterised as having a relatively highly achievable goal with a fairly small gap. The participant-manager of the Forestry System task finds the company in a moderate state, with a large sum of money in the bank account, offset by the threat of pests and poor soil quality, for which pesticides and fertilizers must be purchased. The Forestry System task is characterised by having a moderate gap. In the Tailorshop task, the participant-manager initially finds the company in a terrible state and must work hard to prevent financial ruin. Thus, the Tailorshop task is characterised by having a difficult goal with a relatively large gap between the problem state of the task and the goal state the participant is trying to achieve. Further details regarding the differences between the 3 simulations will be provided in the Method section in Chapter 6.

Goal achievement scores were developed and calculated for each of the three simulations. It was hypothesised that goal achievement will be highest for computer simulations characterised by smaller gaps. Conversely, goal achievement will be lowest for simulations in which the gap is greater. The computer simulations employed in Study 2, in order of increasing gap, are: Furniture Factory, Forestry System, and Tailorshop. Thus, it was expected that goal achievement scores would be high for the Furniture Factory, moderate for the Forestry System, and low for Tailorshop performance. In addition, the relationship between intelligence and complex problem
solving was explored using three scoring techniques (aggregated, final trial, and goal achievement) to measure computer simulation performance.

5.1.2 Factor Analysis of Complex Problem Solving Performance

Complex problem solving performance was assessed by performance scores on the three computer simulations. Following the logic of Brunswik symmetry, factor analyses of computer simulation performance scores was conducted. Factor scores of computer simulated performance across all trials may be an alternative aggregated scoring method than simply averaging performance across trials to calculate mean scores. It is hypothesised that factor scores of computer simulation performance may reveal the different ‘phases’ of performance throughout the task. For example, there may be a ‘rule identification’ phase followed by a ‘rule application’ phase on more complex tasks such as the Forestry System, which would support the model of complex problem solving performance presented in Figure 4.1. It is expected that trials in which participants engage in ‘rule application’ will correlate with cognitive abilities. In contrast, trials in which participants do not engage in ‘rule application’ will not correlate with cognitive abilities.
5.1.3 **Relationships of Complex Problem Solving Performance across Three Computer Simulations**

There has been little prior research investigating the relationship between performance on different computer simulations. A significant correlation was observed between Tailorshop and PowerPlant, $r = .48, p < .05$ (Süß, 2001). PowerPlant was described in Section 3.2). It is hypothesised that the three simulations employed in Study 2 will share common variance, as all three should assess complex problem solving.

5.1.4 **Loopholes**

Results from analyses of Furniture Factory performance in Study 1 indicated the presence of a ‘loophole’ in the computer simulated task. Recall that all computer simulations are based on mathematical algorithms that define the causal connections between the variables. Some participants achieve high simulation performance scores through exploiting loopholes in the model underlying the simulation. In these cases, performance scores do not reflect ‘rule application’ (see Figure 4.1). Participants who exploit loopholes have not engaged in ‘rule identification’ and ‘rule knowledge’ aspects of complex problem solving. They have not understood the underlying rules of the task in order to achieve high computer simulation performance scores, for example in the Space Fortress task (Mané & Donchin, 1989). The Forestry System task employed in Study 2, features a scoring protocol (SKAPCOR) that identifies participants who achieve high performance scores (high total assets) by means other than which the task was designed. The test developer (Wagener & Conrad, 1996) does not explicitly
describe SCAPCOR as a loophole detector. However, in translating this task from German to English for Study 2, it became apparent that the score could be used for this purpose. The proportion of participants who exploit loopholes will be investigated in Study 2 because this topic has received virtually no research attention and may offer an additional explanation for the lack of a relationship between intelligence and complex problem solving in prior research.

5.2 Relationship of Complex Problem Solving and Intelligence

5.2.1 Relationship between Intelligence and Complex Problem Solving Performance on Three Computer Simulations

The relationship between Tailorshop performance and intelligence was investigated in transparent and intransparent conditions (Putz-Osterloh & Lüer, 1981). In the transparent condition, participants were provided with a diagram depicting the connection between the variables. Conversely, in the nontransparent condition, participants were not provided with the model and had to try to infer the rules themselves. A relationship between complex problem solving and IQ was only found in the transparent condition. Results were interpreted as supporting the different cognitive demands hypothesis. That is, intelligence tests are transparent and will not correlate with more realistic and intransparent complex problem solving tasks. There is a general consensus that empirical findings do not support a link between general intelligence and complex problem solving (Wenke, Frensch, & Funke, 2005).
In subsequent studies, guided by Brunswik symmetry, the relationship between Tailorshop performance and intelligence was again tested with more a specific intelligence test (the Berlin Intelligence Structure model, BIS-4, Jäger, 1997) and using an aggregated complex performance score, inclusive of all trials of the Tailorshop task. These studies did find a correlation between the Tailorshop task and specific cognitive abilities even under intransparent conditions. Correlations between Tailorshop performance and the BIS-K factor (processing capacity, capturing the ability to recognise relations and rules and to form logical inferences in figure series, number series, verbal analogies) was significant, $r = .47$, $p < .05$ (Süß, Kersting, & Oberauer, 1993). Substantial correlations were also found for the BIS-K factor and complex problem solving performance on Lunar Lander (Hussy, 1989) and Multiflux computer simulations, $r = .65$, $p < .05$ (Kröner, Plass, & Leutner, 2005). Similarly, the correlation between Forestry System performance and the BIS-K and the BIS-KV was significant, $r = .34$, $p < .05$, and $r = .52$, $p < .05$ ($N = 68$), respectively (Wagener, 2001). Interestingly, the correlation increased from .34 when the BIS level 2, operations, factor scores were used to .52 when the more specific, BIS level 3, facets score was employed (Wagener).

Overall, conclusions from these studies suggest that correlations between complex problem solving and intelligence can be found when tests of factor scores of intelligence are employed. It is hypothesised in Study 2, that there will be a correlation between intelligence factors (e.g., Gf, Gc, Gv) and complex problem solving. There is some concern over whether BIS results would be compatible with results derived from studies employing Gf-Gc theory tests of cognitive abilities (Süß & Beauducel, 2005). Results from Study 2, which is based on Gf-Gc theory, will be compared to results from
previous studies employing the BIS. Very little research has explored the correlation between specific cognitive abilities tests, as opposed to the factors on which they load. Thus, the relationship between sixteen specific cognitive abilities tests and complex problem solving performance on three computer simulations was also explored.

5.2.2 Predictors of Complex Problem Solving Performance

In study 1, the Dynamic Forecasting Questionnaire (DFQ) was employed as a static (paper and pencil) complex problem solving task and the Furniture Factory simulation was employed as a dynamic problem solving task. Unfortunately, analysis of results suggested that the DFQ stocks and flows task, which was similar to that employed in Naturalistic Decision Making (NDM) research (Sterman & Sweeney, 2005), was in fact measuring quantitative knowledge (Gq). Thus, comparison between the DFQ and the Furniture Factory related to dynamic features was not possible. In Study 2, the relationship between a static criterion of academic success (University Admissions Index, UAI) and a static fluid reasoning, Gf, task (Raven’s Progressive Matrices) will be explored. The relationship between static UAI scores and dynamic complex problem solving performance (computer simulated tasks) will also be explored. It was predicted that there would be a significant correlation between the static variables (UAI and Gf). Conversely, it was also predicted that the correlation between the static UAI scores and dynamic complex problem solving performance would not be significant.
5.3  Relationship between Complex Problem Solving and Non-Cognitive Variables

5.3.1  Personality

Little research has been conducted to explore the relationship between personality and complex problem solving. In one study, the relationship was explored between the NEO five factor model of personality (German version) and complex problem solving on the Forestry System (Wagener, 2001). Overall, correlations between personality and Forestry System performance were lower than the researcher had expected. However, employing a sample of 183 participants, significant correlations were found between Forestry System performance and Neuroticism ($r = -.16, p < .05$) and Extraversion ($r = .14, p < .05$). In Study 2 of this thesis, the relationship between the five factors of personality (Costa & McCrae, 1992) and complex problem solving performance in three computer simulations will be explored.

5.3.2  Interests

The relationship between complex problem solving and interests was also explored in Study 2. Holland’s (1959, 1973) hexagonal model of career interests includes: realistic, artistic, investigative, social, enterprising, and conventional). Previous research suggests a relationship between realistic, investigative, and artistic vocational interest and intelligence (Ackerman & Heggestad, 1997). In terms of Ackerman’s Process, Personality, Interests and intelligence as Knowledge (PPIK) theory, realistic and investigative interests correlate with intelligence as process
factors (reasoning, mathematics, spatial) and abilities related to both process and knowledge (mechanical). Artistic interests are correlated with intelligence as knowledge factors (verbal). Thus, following on from the expectation that there will be a relationship between complex problem solving and intelligence, and that intelligence is related to interests, it was hypothesised that complex problem solving will also be related to interests. Specifically, it was hypothesised that interests that correlate with intelligence (realistic, investigative, and artistic) would correlate with complex problem solving performance (Furniture Factory, Tailorshop, and Forestry System performance).

5.4 Summary of Key Predictions

Firstly, scoring issues were addressed. The relationship between intelligence and complex problem solving was explored using three scoring techniques (aggregated, final trial, and goal achievement) to measure computer simulation performance. It was expected that correlations between intelligence and complex problem solving would be increased by the use of specific cognitive abilities tests and aggregated computer simulation scores, rather than general or factor scores of intelligence and final trial computer performance simulation scores. In order to identify different ‘phases’ of simulation performance, factor scores of computer simulation performance across trials were calculated. It was expected that trials in which participants engage in ‘rule application’ would correlate with cognitive abilities. In contrast, trials in which participants do not engage in ‘rule application’ would not correlate with cognitive abilities. In addition, goal achievement scores were developed and calculated for each
of the three simulations. It was hypothesised that goal achievement will be highest for computer simulations characterised by smaller gaps. Conversely, goal achievement will be lowest for simulations in which the gap is greater. Additionally, it was hypothesised that the three simulations employed in Study 2 would share common variance, as all three should assess complex problem solving. The proportion of participants who exploit loopholes was also measured.

Secondly, the relationship between intelligence (using tests from Gf-Gc theory) and complex problem solving (assessed by performance scores on the Furniture Factory, Tailorshop, and Forestry System computer simulations) was explored. A relationship between intelligence factors and complex problem solving performance on the three computer simulations was investigated. The correlation between sixteen specific cognitive abilities tests and complex problem solving performance was also explored and was expected to be higher than that for factor scores of intelligence and complex problem solving performance. Specifically, it was hypothesised that performance on ‘impure’ cognitive abilities tests (i.e., tests that have dual factor loadings) such as esoteric analogies and critical reasoning would have higher correlations with complex problem solving than ‘pure’ cognitive abilities tests (i.e., tests that have single factor loadings). In addition, it was predicted that there would be a significant correlation between the static variables (UAI and Gf). Conversely, it was also predicted that the correlation between the static UAI scores and dynamic complex problem solving performance would not be significant.

Thirdly, the relationship between complex problem solving and non-cognitive variables was investigated. It was hypothesised that interests that correlate with intelligence...
(realistic, investigative, and artistic) would correlate with complex problem solving performance (Furniture Factory, Tailorshop, and Forestry System performance). The relationship between the five factors of personality (Costa & McCrae, 1992) and complex problem solving performance in three computer simulations was also explored. It was hypothesised that there would be a negative correlation between the personality trait neuroticism and complex problem solving performance.
CHAPTER 6

METHOD: STUDY 2

6.1 Participants

A total of 296 participants were involved in the study. The average age of the sample was 21.02 years with a standard deviation of 6.18 years. There were 189 females and 107 males. The mean age of the females was 20.83 years, with an age range from 17 to 58 years. The males had a mean age of 21.37 years, with an age range from 17 to 47 years. Participants with English as their first language represented 79% of the sample.

157 were first-year psychology students from the University of Sydney who were participating as part of a course requirement. 87 were second and third-year psychology students who responded to a poster campaign in lectures and 52 were members of the general public who responded to poster campaigns at the Paddington RSL club and the Centre for Continuing Education, the University of Sydney. In return for their participation, all participants received a detailed report (more detailed than conventional participant feedback) outlining their results, with an emphasis on cognitive abilities and personality.

Participants were questioned to ensure they had no anxiety about working with a Personal Computer (PC). In the case where an individual had no previous experience
with Microsoft Windows at all, they were excluded from the study. There were 6 such cases; all were aged over 50 years.

6.2 Materials

6.2.1 Cognitive Abilities Tests: Computerised

The computer administered cognitive abilities tests employed in Study 2 were identical to those employed in Study 1. These tests are described in Section 4.2.3. Swaps and Triplets tests were employed to measure Fluid Reasoning (Gf); Vocabulary and Proverbs Matching were employed as Crystallised Knowledge (Gc) measures; Numerical Operations and Financial Reasoning were employed as Quantitative Knowledge (Gq) measures; and Line Length and Letter Spotting were employed as Broad Visualisation (Gv) measures (Stankov, 1999).

In addition to the cognitive abilities tests employed in Study 1, Speed of test taking (Gs) for each of the eight tests, listed above, was measured (Stankov), and an additional Gv measure, Paper Folding, was included.

Speed of Test Taking

Much of the recent work on intelligence is directed at establishing the empirical link between fluid intelligence and speed of mental operations. It is important to keep in mind the fact that speed is not a unitary construct – there are several types of mental speed and each one of these may have a different relationship to intelligence. While the
measurement of mental speed is important from both the theoretical and practical points of view, mental speed is not the basic process of intelligence (Stankov & Roberts, 1997). The present battery measures one of the mental speed constructs – the speed of test taking. In an organisational setting, speed of test taking may be useful for the section of personnel for jobs requiring a quick decision under time pressure. It is useful to compare people with respect to speed-accuracy trade-off (i.e., whether those who are quick are also prone to committing more errors). Finally, it is also known that speed of mental processing is closely linked to age. Older people are slower than younger people, though there are considerable individual differences in the aging process (Stankov, 1999).

**Paper Folding**

The Paper Folding task is considered to assess spatial skills including a working memory component (Kyllonen, Lohman, & Snow, 1984; Salthouse, Mitchell, Skovronek, & Babcock, 1989). It was included in the present study as a marker of broad visualization ability (Gv). In this test, participants were instructed to imagine the folding and unfolding of pieces of paper. In each problem in the test, figures displayed a square piece of paper folded from one to three times, the punching of a hole through all thicknesses of the folded paper, and a pattern of circles indicating the locations of the punched holes in the unfolded paper. Examples of the three conditions of the spatial Paper Folding task are displayed in Figure 6.1.
Figure 6.1 illustrates the outcome of each fold. However, computer administration of the task allowed more realistic displays of the fold lines including different coloured lines for the paper outline, internal folds, and the most recent fold. Items with 2 stages of folding and 3 stages of folding are intermingled throughout the test with a single 1 fold item given in the middle of the test. Participants were asked to decide whether the pattern of holes in the final display was consistent with the pattern that would result from the earlier sequence of folds and punch locations. The Paper Folding task has 19 items, plus one example that displayed the correct answer and one practice that indicated whether the response was correct or incorrect. Answers were chosen from amongst 5 alternative figures that showed where the holes were when the paper is completely unfolded. The test was not timed. However, participants were instructed to work as quickly and carefully as they could. Total test time was approximately 20 minutes.

6.2.2 Cognitive Abilities Tests: Paper and Pencil

In addition to the nine computerised cognitive abilities test, six paper and pencil cognitive abilities tests were also presented to the participants as a test booklet. The six
paper and pencil tests included: General Knowledge, Letter Series, Esoteric Analogies, Raven’s Advanced Progressive Matrices, Problem Solving, and Critical Reasoning. Each of these tests will now be described in turn.

**General Knowledge Test**

The General Knowledge test was taken from the Gf/Gc Quickie Battery (Stankov, 1997). Participants were presented with 24 true or false questions that assess their knowledge of history, geography, current events, science, and technology. As such, the content of this test is similar to that of the Information sub-test of the WAIS-III (Wechsler, 1997).

An example is, ‘Homer wrote both the *Iliad* and the *Odyssey.*’

**Letter Series Test**

The Letter Series test is a typical Thurstonian series completion task, which operates as a marker for the Induction primary factor and at the second-order helps define a Gf factor. Carroll (1993) described such tasks as requiring participants to “notice what rule is exemplified in the progression of the series, and show induction by supplying (the) element that continues the series” (p. 211). Participants were presented with a series of letters (e.g., A, C, E, G,?) and were required to supply the next letter that continued the series. There were 12 items and 2 practice items, presented in an open-ended format.
**Esoteric Analogies Test**

In this task, participants were required to determine the relation between two words and to select, from among five alternatives (e.g., the word having the same relation to a third word (e.g., Light is to Dark as Happy is to?)). There were 12 items, and one practice item.

**Raven’s Advanced Progressive Matrices (RAPM) Test**

This task was a mixture of the Standard and Advanced versions of the Raven’s Progressive Matrices (Raven, 1938; Raven, Court, & Raven, 1979; Raven, Raven, & Court, 1993) containing 10 items that has been employed in previous studies in the individual differences laboratory at the University of Sydney (e.g. Stankov & Crawford, 1996a, 1996b).

![Example item from the Raven’s Progressive Matrices task.](image)

**Figure 6.2.** Example item from the Raven’s Progressive Matrices task.
As shown in Figure 6.2, participants were presented with a 3 X 3 matrix of symbols, with the bottom-right symbol missing. Underneath the matrix were eight other symbols, one of which was the correct solution to completion of the matrix. Participants were instructed to determine which of the eight symbols solved the problem.

**Problem Solving**

Participants completed 15 Problem Solving items taken from the quantitative section of the Graduate Management Admission Test (GMAT) prepared for the Graduate Management Admission Council by Educational Testing Service (ETS, 1997). Problem Solving questions are designed to test basic mathematical skills, understanding of elementary mathematical concepts, and the ability to reason quantitatively and solve quantitative problems. Problems were presented in paper and pencil format and participants were instructed to solve each problem. Then indicate the best of the 5 multiple-choice answers given.

An example is, ‘Two oil cans, X and Y, are right circular cylinders, and the height and the radius of Y are each twice those of X. If the oil in can X, which is filled to capacity, sells for $2, then at the same rate, how much does the oil in can Y sell for if Y is filled to only half its capacity?’ Alternative answers are: (a) $1; (b) $2; (c) $3; (d) $4; and (e) $8. Participants were given a generous 25-minute time limit and were instructed to work as quickly and accurately as they could.
Critical Reasoning. Participants completed 15 Critical Reasoning items taken from the Verbal section of the Graduate Management Admission Test (GMAT) prepared for the Graduate Management Admission Council by Educational Testing Service (ETS, 1997). Critical Reasoning questions are designed to test the reasoning skills involved in making arguments, evaluating arguments, and formulating or evaluating a plan of action. Questions were presented in paper and pencil format and participants were instructed to solve each question, and then select the best of the 5 multiple-choice answers given.

An example is, ‘Most consumers do not get much use out of the sports equipment they purchase. For example, seventeen percent of the adults in the United States own jogging shoes, but only forty-five percent of the owners jog more than once a year, and only seventeen percent jog more than once a week.’

‘Which of the following, if true, casts most doubt on the claim that most consumers get little use out of the sports equipment they purchase?’

Alternative answers are: (a) Joggers are most susceptible to sports injuries during the first six months in which they jog; (b) Joggers often exaggerate the frequency with which they jog in surveys designed to elicit such information; (c) Many consumers purchase jogging shoes for use in activities other than jogging; (d) Consumers who take up jogging often purchase an athletic shoe that can be used in other sports; and (e) Joggers who jog more than once a week are often active participants in other sports as well.
Participants were given a generous 25-minute time limit and were instructed to work as quickly and accurately as they could.

6.2.3 Personality

Personality was assessed by the 60-item Openness Conscientiousness Extraversion Agreeableness Neuroticism Index Condensed (OCEANIC) self-report questionnaire (Roberts, 2000). The OCEANIC has 5 scales used to assess each of the 5 factors of personality (Five Factor Model of personality: Costa & McCrae, 1992). These are: (a) Openness (e.g. ‘I am philosophical.’); (b) Conscientiousness (e.g. ‘I am organised.’); (c) Extraversion (e.g. ‘I like parties, where there are a lot of people.’); (d) Agreeableness (e.g. ‘I try to be kind to everyone I know.’); and (e) Neuroticism (e.g. ‘I worry more than most people.’). Participants were asked to use a 6-point Likert scale to rate their first impression of how well each statement describes the way they think or feel, where 1 = never and 6 = always. To control for acquiescence response, 7 sentences were worded in a negative direction. Overall scores for each of the 5 factors of personality, tapped by the different statements, were calculated by summing across the 12 items for each of the 5 scales after reversing some items for consistency of direction of expression.

6.2.4 Interests

The Interest Determination, Exploration and Assessment System (IDEAS: Johansson, 1990) was employed to measure career interest. Participants rated each of the 128 items
on a 5-point scale, where 0 = ‘dislike the activity very much’ and 4 = ‘like the activity very much’. Overall career interest scores were calculated by summing across the 8 items in each of the 16 subscales. These were: (a) mechanical/fixing (e.g. ‘Work with small hand tools’); (b) protective services (e.g. ‘Drill in military company (march in formation’); (c) nature/outdoors (e.g. ‘Plant your own garden’); (d) mathematics (e.g. ‘Study algebra’); (e) science (e.g. ‘Work in a research laboratory’); (f) medical (e.g. ‘Be part of a medical operating team’); (g) creative arts (e.g. ‘Visit art galleries’); (h) writing (e.g. ‘Write a novel’); (i) community service (e.g. ‘Do volunteer work with a community group’); (j) educating (e.g. ‘Teach swimming’); (k) child care (e.g. ‘Be a nursery school helper’); (l) public speaking (e.g. ‘Lead group discussions’); (m) business (e.g. ‘Interview people for a job’); (n) sales (e.g. ‘Sell something to a customer’); (o) office practices (e.g. ‘Do office work, such as typing or filing’); and (p) food service (e.g. ‘Study home economics’).

6.2.5 Complex Problem Solving Tasks

Three computer simulations were employed to assess complex problem solving competencies: The Furniture Factory (Goodman & Wood, 2004; Wood & Bailey, 1985); The Tailorshop (Putz-Osterloh, 1981; Süß, 1996); and the Forestry System (FSYS 2.0: Wagener & Conrad, 1996). The Furniture Factory task was the same computer simulation that was employed in Study 1, and was previously described in Section 4.2.6. The Tailorshop and Forestry computer simulations will be described next.
6.2.5.1 **Tailorshop: The Task**

The Tailorshop simulation used in the present studies is a slightly modified version of the business simulation originally developed by Dietrich Dörner and first used in a study by Putz-Osterloh (1981). The Tailorshop has been used extensively as a research tool for investigating individual differences in complex problem solving performance, most notably the Mannheim Research Project (Süß, 1996; Wittmann & Süß, 1999). The Tailorshop task meets the criteria for a complex problem solving task, outlined in Section 1.2.6 and consists of complexity, connectivity, dynamic environment, and intrasparency (in the absence of polytely).

The participant is assigned the role of the manager of a Tailorshop company for 12 simulated months; this corresponds to 12 trials within the computer simulation. Before making any managerial decisions, the participant is informed that the company is in financial trouble and it will go bankrupt without their intervention. Participant managers are also informed about the main objective of the game, which is to use their managerial skills to increase the total assets of the company and thus save it from financial ruin. The total assets consist of the money that is deposited into the company’s bank account (the balance), the value of the machines, the property value of the sales outlets, the value of raw materials, and the shirts in stock. A participant reference sheet (Appendix D) is provided describing important information about the company and the various intervention options that are available to the participant managers. Participants are instructed to read the reference sheet before commencing the practice trial. This sheet can then be referred to at time throughout the game.
The first part of the participant reference sheet provides written information to the participant manager from the simulated board of directors of the company. It contains specifications about prices and the organization of the company. Participants are instructed to pay close attention to information regarding ‘Investment costs and company expenditure’ and the ‘Organization of the company’.

The information concerning investment costs and company expenditure provides important details such as: The raw material prices vary according to the market conditions; Last year the average price was $5 per unit of raw material; If an employee is fired, you must pay them a month’s salary as termination compensation; and the bank supplies you with generous but limited credit, if you are deemed to have a low credit rating and can no longer receive credit for a purchase, e.g. buying more machines or establishing new sales outlets, you will be informed on screen. You must then try, without generating more costs or making new investments, not to lose profits. Further unlimited credit can be received for the current costs, such as loans, social expenditure, and income-related expenses, as long as there are no new purchases made.

The information concerning the organization of the company provide important information such as: One employee is required to properly operate every shirt-producing machine regardless of whether it is a larger machine capable of producing 100 shirts per month or a smaller machine with a 50 shirts per month capacity. The participant is also alerted to the fact that a sales representative is not counted as an employee of the company. As external consultants, sales representatives charge the company a monthly fee of $500.
The second part of the reference sheet outlines all the important variables in the Tailorshop simulation that are under the control of the participant manager and those variables that are calculated by the underlying mathematical algorithms of the game. The system has 24 variables all together, 12 of them are exogenous variables and can be manipulated directly by the participant manager. The 12 exogenous variables are: raw material, shirt price, advertising, selling positions, sales representatives, 50 shirts/month machines, 100 shirts/month machines, maintenance $, worker for 50 shirts/month machines, workers for 100 shirts/month machines, wages $, social costs $. By clicking on the exogenous variable of interest to highlight it, e.g. ‘Sales Representatives’, the participant is able to access information regarding that variable e.g. ‘This month you have employed 2 sales representatives. How many sales representatives (1 to 3) would you like to employ next month.’ Figure 6.3 displays a screen image from the Tailorshop simulation.

Figure 6.3. Screen image from the Tailorshop computer simulation.
The system also has 12 endogenous variables that are computed by the simulation after the participant manager has decided on their interventions for a trial. The 12 endogenous variables are: Total assets, bank balance, raw material price, demand, shirts in stock, shirts sold, production, production loss, machine damage, machine efficiency, workload efficiency, work motivation. These endogenous variables are calculated by the underlying mathematical algorithms of the game, which correspond to rules of play. These rules, reflecting the causal connection between the variables, are not explicitly provided to the participant, but must be inferred throughout playing the simulation by observing the effects of one’s decisions as the participant manager.

After introducing participants to the task and allowing them to read through the important information provided by the participant reference sheet, the participants were given a practice run of managing the Tailorshop company for two standardised trials corresponding to two simulated months. During the first month of practice, the experimenter guided the participants through some example interventions showing the participant managers the various actions that are possible for each month of organizational management. During the second practice month, the participants were given more autonomy and were instructed to take themselves through various standardised company interventions without specific guidance from the experimenter. Directions for the second practice month are provided in Appendix E. Participants were able to ask the experimenter for help if they experienced any problems while taking themselves through the second practice month. They were assured that they would not be held accountable for the consequences of their decisions during training that were guided by a compulsory set of instructions. The 2-month practice phase took approximately 20 minutes. Throughout the practice phase, the participants learned how
to navigate around the program interface using the various computer menus provided; more detail on system navigation will be described next.

The information displayed on the computer screen, i.e. the program interface, is divided into two parts. The exogenous variables, which can be manipulated directly by the participant manager, are listed on the left hand side of the screen and the endogenous variables, which are automatically calculated by the simulation in the following month, are listed on the right hand side of the screen. Participant managers can buy raw materials, change the price of the shirts, modify the expenditure on marketing and advertising, establish or close new sales outlets, buy or sell machines and many others. As described earlier, the participant is able to access more information regarding each of these variables by highlighting them using the mouse. Participants use this extra information to guide their decisions and may make as many or few interventions as they wish each month. All decisions are indicated as ‘Plans’ on the screen and calculations will not be carried out until the participant manager is satisfied that all the necessary decisions have been made and they wish to proceed to the next month. This allows the participant to modify any or all of their decisions before proceeding to the next month.

Once they have decided on their complete set of interventions for the month, participants press the button for “execute the plans and close the month”. The participant is then informed that 1 month has passed. The screen displays the new status of the company, as a result of the participant manager’s actions in the previous month as well as the possible interventions for the current month. There is also an option to review sets of decisions made in the previous month in order to understand how those
prior decisions affected the current status of the company, however, decisions made in the previous month can no longer be modified.

In the top row of the program interface, participant managers are shown in which simulated month (1-12) they are currently situated. Information regarding the highest possible value of total assets that their interventions as the company manager could achieve by the end of the 12-month period is also displayed. The real time remaining (in minutes) for the 12 simulated months was displayed at the bottom of the screen. Participants were given 1 hour to complete the 12 trials. However, this time limit was generous and all participants were able to complete the game within the limit.

In summary, the Tailorshop simulation was introduced to the participants with detailed written instructions and two standardised practice trials guided by the experimenter. The training phase took approximately 20 minutes. Participants then managed the Tailorshop for 12 simulated months, their main objective was to increase the total assets of the company and minimise losses; the actual simulation took approximately 1 hour.

6.2.5.2 Tailorshop: Measures

German SPSS variable names, for the various scores that can be extracted from the Tailorshop data, have been retained for ease of comparison to previous research. Measurement of problem solving performance was based on the scoring protocol used in previous research by Heinz-Martin Süß and colleagues (Süß, 1996, 2001; Süß, Oberauer, & Wittmann, 2005; Wittmann & Süß, 1999). The participant’s performance in the simulated organization was measured in terms of the final status of the total assets
of the Tailorshop company. The total assets in the final trial of the game (month 12) is termed ‘xgekap 12’.

Another measure termed ‘xtrend’ gives the total number of months with gain, i.e. the value of the company’s total assets in the current month ‘xgekap month x’ is higher than the value of total assets in the previous month ‘xgekap month (x-1)’. If the total assets had increased since the previous month, e.g. (xgekap 12) – (xgekap 11) > 0, then the participant is said to have ‘won’ the trial and is assigned a score of 1. If the total assets had decreased since the previous month, e.g. (xgekap 12) – (xgekap 11) < 0, then the participant is said to have ‘lost’ the trial and is assigned a score of 0. The score ‘xtrend’ is the sum of all ‘won’ and ‘lost’ trials and is a total score out of 12 months, thus ranging from 0 to 12.

The ‘xtrend’ score reduces participant’s performance to a binary score of 0 or 1 and thus some variance in the raw scores is lost by using this procedure to average performance over all 12 trials of the simulation. Another score reflecting average performance throughout the game could be calculated by taking the mean of the raw scores of total assets at the end of each month. This mean score is termed ‘xgekap mean’.

Süß (2001) suggested a new criterion measure termed ‘xplg’. This score takes into consideration the number of shirts sold ‘xverkauf’ and the profit margin per shirt ‘xgspa’. As the distribution of the raw scores deviated considerably from a normal distribution, Süß transformed them into ranks (Z scores). The new criterion measure is simply a sum of the z-scores of ‘xverkauf’ and ‘xgspa’.
6.2.5.3 Forestry System (FSYS 2.0)

6.2.5.4 Forestry System: Translation

The Forestry system, FSYS 2.0 (Wagener & Conrad, 1996) was specifically translated from German to English for Study 2 of this thesis. This study represents the first time the Forestry task has ever been administered in English. As mentioned in Chapter 2, the vast majority of individual differences research exploring the relationship between intelligence factors and microworld performance has been conducted in Germany. Thus, many microworld tasks are designed for German speaking samples. Individual difference researchers must engage in a cumulative effort to understand the characteristics of microworlds as research tools. Translating commonly used German microworld tasks, for administration to English only speaking samples, is an important first step towards obtaining stable replications of results from which generalisations about micro-worlds can be made. There were 5 components of the Forest task to be translated: The program interface; test manual; participant introductory reference sheet; information text; and scoring protocol. Important steps in the translation process for each of these will now be described in turn.

The Forest task’s program interface is the visual display that the participant observes on screen. The test developer Dietrich Wagener conducted an initial translation of the program interface.

The FSYS 2.0 test manual (Appendix F) contains information regarding system software and hardware requirements, components, installation, and running a test
session. The author, assisted by a second translator, conducted an initial translation of the test manual. The lack of familiarity of the assistant translator with computers caused difficulties and a more computer proficient translator was employed to continue assistance with the translation. Finally, German speaking and English speaking computer technicians were employed to assist with translation of the most technical sections. Many highly technical terms could not be found in the German to English dictionary, nor could they be translated by AltaVista’s Babel Fish translating program (AltaVista, 2004). At every stage, translated text was submitted to the original test developer for clarification of meaning and approval.

The Participant introductory reference sheet (Appendix G) was designed to familiarise the participants with the simulation. The aim of the translating process (carried out by the author) was not a literal translation of the German text into English. The translation was designed to convey the same information as the original in an easy to read style with the main aim to maintain consistency with the program interface (described earlier) and information text that will be described next.

The test manual and participant introductory information are included as appendices, as their translation, from German to English, formed a substantial component of this thesis and this information is not available elsewhere. The technical documentation for the other two microworlds employed in the present research can be accessed by contacting the test developers.

The information text is a store of information within the system, which participants can open to learn about almost everything of significance for the management of the
company. The present researcher carried out translation of the information text in a consistent way, according to the words used for key concepts in the program interface and the participant introduction. A literal translation was not required, with an emphasis on conveying the same information as the original and consistency with other parts of the computer program. Some words could not be translated because they were fictional entities such as trees that did not exist in the real world. This lowered the impact of prior knowledge on game performance because the trees themselves, their rate of growth, susceptibility to pests, etc. were not based on any real world trees. The author used creative license to devise new fictional names for trees, pests, fertilizers, pesticides, etc. ensuring that they did not correspond to reality. Consideration was also given to the length of words because this can cause problems in the program. The monetary values provided in the simulation’s bank account were not changed from euros to dollars, using the current exchange rate, because the amounts used in the scenario are high enough that almost no participant would notice the subjective difference in this area. Maintaining consistency also facilitated comparison of data between the German version and English version of the Forestry system (FSYS 2.0).

Finally, the test developer translated the scoring protocol after translation of the program interface; test manual; participant introductory information; and information text were complete. A pilot study was conducted where the English version of the Forest test was administered to five PhD students, judged to have high verbal skills. The participants played the game three times each and commented on all the texts used. All suggestions were given to the test developer for approval, and if accepted, were implemented in the final version of the task. The license server is located at the
University of Mannheim and a computer technician from the University of Sydney assisted with resolving installation problems and setting up licensing via the Internet.

6.2.5.5 **Forestry System: The Task**

The Forest task (FSYS 2.0: Wagener & Conrad, 1996) is a business simulation that was developed for psychometric measurement. It has predominantly been used in individual differences research (Wagener, 2001; Wagener & Conrad, 1996) to assess complex problem solving behaviour and more recently in experimental research investigating the role of emotions in complex problem-solving (Spering, Wagener, & Funke, 2005). The design of the user interface was based on a theoretical model (Dörner, 1986) that describes the demands of complex problem solving. The Forest task meets the criteria for a complex problem solving task, outlined in Section 1.2.6 and consists of complexity, connectivity, dynamic environment, and intransparency (in the absence of polytely).

Participants were given 10 minutes to read through the introductory information that was provided as a participant reference sheet and thus could be referred to at any time throughout the game. This described the participant’s role as the manager of a forestry company in a fictitious future whose main objective was to increase the total assets of the company. The participant made decisions regarding planting, growing, and cutting down trees. In order to grow trees that were suitable for harvesting, the participant manager had to control pests and care for the quality of the soil. The total assets
calculated themselves automatically from the value of five forests (depending on tree quantity and the quality of the soil) and the status of the bank account.

After reading through the introductory information, participants were then taken through a group practice session. The experiment supervisor read a script that can be found in the Forest Test Manual (Appendix F) while the participants followed along on their computer screens. This ensured that all participants received the same information. After the 20-minute instruction period, participants were given a maximum of 90 minutes to independently work through the 50 simulated months with the opportunity to ask the experimenter questions at any time. Figure 6.4 displays a screen image from the Forestry system (FSYS 2.0).

Figure 6.4. Screen image from Month 1 of the Forestry system (FSYS 2.0).
In order to achieve the long-term objective of increasing the total assets of the company after 50 simulated months, the participants were able to make a number of decisions. These decisions were made monthly by using the onscreen menus that were colour coded into 2 groups: green (sections of the company) and yellow (actions). Participants could click on any of the 7 green buttons to switch between the various sections of the company (Forests 1 to 5, Finances, and Information Text). Participants could click on any of the 3 yellow coded buttons in order to initiate an ‘action’, such as planting or cutting down trees. The button ‘statistics’ allowed for analysis of previously simulated months. Once decisions for the month had been made, the button ‘advance’ progresses the game to the next month. Each of these decision types (menus) moves the participant between different computer screen displays that will now be described in turn.

Sections of the company:

Forests 1 to 5: There are five forests in which trees can be planted, grown, and cut down to generate profit. The harvested trees are automatically sold, and this is the only source of income in the simulation. The five forests are all the same size, however the soil conditions can vary and are affected by the increase or loss of trees. The forests are always a monoculture, so in every forest there is only one type of tree planted. The growth of the trees begins with the planting of seedlings. Later, more trees can be planted, but not before the previous generation of trees in the current forest is completely cut down. In the each of the five ‘Forest’ sections of the game, onscreen information available to the participant includes tree variety; tree population; age of tree population; mean thickness of timber; sales revenue %; infestation of 3 types of pests
(number of pests per kilogram of leaves and an indication of low, medium, or high infestation); and 5 types of minerals in the soil (grams per cubic metre).

Finances: All costs (wages, fertilizer, seedlings, and pesticides) are automatically withdrawn from the participant managers bank account. Similarly, all income from timber sales is automatically deposited into the bank account. The ‘Total Business’ section of the game contains onscreen information regarding the net worth of each of the five forests separately; fixed costs; material costs; revenue from timber sales; bank balance; and total assets. This section provides important feedback for the participant because they can assess, at the end of each month, how well they are progressing towards the main objective of the game that is to increase the total assets of the business.

Information Text: In the ‘Information Text’ section of the company, participant’s can learn almost everything of significance for the management of the company. The computer screen display in this section is divided into 2 menus; the ‘category’ menu can be used to select specific information from the ‘information’ menu. There are 4 major categories of information with several subcategories in each: 1. ‘Finances’ provides information on fixed costs/overheads, material costs, value of forest plots, workers; 2. ‘Pests’ provides information on pests and pest control; 3. ‘Soil’ provides a description of the four types of fertilizers; and 4. ‘Trees’ provides information on the eight types of trees, their mineral requirements, susceptibility to pests, and pest control. The information department provides all the facts required for each month’s organisational decision. The workers can only work on one task in one forest per month. So, the
participant must prioritise the most important task then make decisions based on the appropriate category from the information text.

Actions: The participant manager can allocate 4 types of tasks to the workers: plant seedlings, cut down trees, pest control, and fertilize. In order to issue a work order, the participant must first select the forest in which the work team will work in the present month, prioritise the most important action for the present month, check the information text (optional), and then activate the ‘action’ button. The ‘action’ computer display screen provides information regarding the status of the current work order e.g. “At present your workers are unengaged.” Some work orders take several months to complete and the participant has the option to “stop the work order” and move workers onto another task before the present task is complete e.g. Forest 3 is 78% harvested. Thus, tradeoffs must be made.

Statistics: Participants are advised to always bear all forests in mind. They can use the ‘statistics’ screen in order to better assess the trends. The graphic presentation provides a very quick overview and gives the participant feedback relating to all aspects of the company. Values that can be displayed include: age of the tree population, mean thickness of the timber, sales revenue, infestation by each type of pest, and levels of each type of mineral in the soil. Participants can choose to view the status of the company as a graphic presentation or as a presentation of exact numbers in a tabular format. Thus, the participant can accurately analyse the development of the individual forests and the status of the company at any time.
Advance: After completing the monthly decisions, the participant manager used the ‘advance’ button to progress to the next month until all 50 months were completed.

In summary, in order to increase the total assets of the forestry company, the participant manager decided which of the five forests was the priority for that particular month. Then they chose one of four exogenous variables (plant trees, cut down trees, pest control, fertilize) for the work group to carry out. The ‘information’ and ‘statistics’ screens could be accessed at any time to assist with these decisions. The ‘finance’ screen displayed endogenous information that was calculated by the system. The ‘finance’ screen provided participants with feedback regarding their progression towards the main objective of increasing the total assets of the company across the 50 simulated months of management tenure.

6.2.5.6 Forestry System: Measures

The measurement of complex problem solving performance on the Forestry system was exactly like that employed by Wagener and colleagues (Wagener, 2001; Wagener & Wittmann, 2002).

SKAP: The participants’ organisational performance was measured in terms of the final status of the total assets of the forestry company. This scale is directly equivalent to the main objective of the task to maximise total assets (based on value of the land, trees, and the bank balance combined). A higher total value implies better participant performance. The Forestry overall performance score is equal to the final status of the
total assets of the company in month 50, which is the final trial of the game, dollar figures range up to tens of millions. These dollar values are converted to a scale of 0 to 100. A value of 50 raw points indicates that the assets of the company at the end of the task are just as high as they were at the beginning when the participant manager took over the company. Thus, participants with a score greater than 50 have made an economic gain. This variable is referred to as ‘SKAP’ in research by Wagener and colleagues (Wagener, 2001; Wagener & Conrad, 1996; Wagener & Wittmann, 2002). The system automatically calculates SKAP scores in the output file. In cases that are outliers, the system automatically replaces the SKAP score with the participant number. Therefore, outliers must be removed from the sample, as SKAP scores are missing. The output provides a single SKAP score for the final trial of the computer simulation. FSYS month 50 raw scores are not replaced with participant numbers and can also be employed as a measure of SKAP in the final trial of the computer simulation.

SKAPCOR: SKAPCOR is a version of SKAP where scores are re-calculated for those participants who have engaged in major operational mistakes (SKAPCOR > 100). The SKAPCOR score distinguishes between those participants who have engaged in a high level of problem solving performance and those who have found a ‘loophole’ in the task. Participants exploiting the ‘loophole’ engage in risk taking behaviour that create short term profit for the company. However, the consequences of their decisions will lead to financial ruin. These participants have not played the role of the manager, as they were instructed to do. In these cases, high SKAP scores do not reflect superior problem solving behaviour. These participants must be removed from the sample as they are outliers and their scores do not reflect high problem solving performance but
rather problematic decision making that reflects poor managerial skills. The output provides a single SKAPCOR score for the final trial of the computer simulation.

6.3 Development of a Consistent Complex Problem Solving Score: Goal Achievement

In order to compare complex problem solving performance across the three computer simulated tasks (Furniture Factory, Tailorshop, and Forestry System), it was necessary to devise a new consistent scoring measure. ‘Goal achievement’ reflects problem solving competencies and is common to all three computer simulations. The concept of ‘goal achievement’ performance scores was based on ‘xtrend’ scores for the Tailor Shop simulation (Süß, 1996) where total assets in the current trial are subtracted from total assets in the previous trial. If the company has made a profit, then the participant is said to have ‘won’ that particular trial and a score of 1 is given. Conversely, if the total assets in the current trial are less than the total assets in the previous trial, the company is running at a loss, and the participant is said to have ‘lost’ and is given a score of 0 for that particular trial. At the end of the simulation the sum of all ‘won’ and ‘lost’ trials is calculated. This measure of goal achievement was extended to calculate new performance scores for the Furniture Factory and Forestry system tasks. Calculations of the ‘goal achievement’ scores for the Furniture Factory and Forestry system will be described next.
6.3.1  **Goal Achievement score: Furniture Factory**

The Furniture Factory instructions assigned participants a goal to motivate their employees to work as efficiently as possible. An overall measure of ‘goal achievement’ was calculated based on the formula:

\[ \text{Outcome} = (2 - \frac{\text{total actual score cycle}}{\text{total estimated score cycle}}) \times 100 \]

If a participant’s actual score were equal to the estimated score, then the Outcome score would be 100. If a participant’s actual score was less than the estimated score, i.e. they managed to motivate their employees to work more efficiently than the estimated time, then the Outcome score would be greater than 100. In contrast, if the participant’s actual score was greater than the estimated score, i.e. their employees worked more slowly than the estimated time, then the Outcome score would be less than 100. In summary:

\[ \text{Outcome} > 100, \text{ you win. Win = 1.} \]
\[ \text{Outcome} < 100, \text{ you lose. Lose = 0.} \]

A measure of ‘goal achievement’ was calculated by summing the number of weeks that a participant ‘won’. Thus, a score of 18 represents a participant who managed to motivate their employees to work more efficiently than the estimated time for all 18 simulated weeks of the game. They achieved the goal they were instructed to achieve during the Furniture Factory task. Goal achievement scores range from 0 to 18.
6.3.2 Goal Achievement score: Forestry System

The Forestry system instructions assigned participants a goal to increase the total assets of the company. An overall measure of ‘goal achievement’ was calculated based on increases or decreases in the total assets of the company in each of the 50 simulated months.

Firstly, outliers (those participants who had exploited a ‘loophole’ in the system) were removed from the sample. There were 10 such cases. Secondly, the total assets in each month, was subtracted from the subsequent month. For example, Total assets in month 1, was subtracted from total assets in month 2. Total assets in month 2, was subtracted from total assets in month 3, and so on, for each of the 50 simulated months. If there was an increment in the total assets since the previous month, the difference between the months was positive and the participant had achieved the goal for that month. If the goal was achieved, the participant had a ‘win’ and was assigned a score of 1.

(FSYS month 2 − FSYS month 1) > 0, you win. Win = 1.
(FSYS month 3 − FSYS month 2) > 0, you win. Win = 1.

In contrast, if there was a decrease in the total assets since the previous month, the difference between the months was negative and the participant failed to achieve the goal for that month. If the goal was not achieved, the participant had a ‘lose’ and was assigned a score of 0.

(FSYS month 2 − FSYS month 1) < 0, you lose. Lose = 0.
A measure of ‘goal achievement’ was calculated by summing the number of months that a participant ‘won’. Thus, a score of 50 represents a participant who managed to increase the total assets of the company across all 50 simulated months of the game. They achieved the goal they were instructed to achieve during the Forestry system task. Goal achievement scores range from 0 to 50.

For all three computer simulations, it is acknowledged that the ‘goal achievement’ scoring procedure will reduce some of the variance captured by the raw scores. Both raw scores and goal achievement scores will be included in data analyses.

### 6.4 Procedure

Participants completed the test battery over 3 days, for 2 and ½ hours each day, sessions were 1 week apart. Prior to undertaking the tests, participants were informed that the study was confidential, were given a brief rationale for the experiment, and were asked to give their informed consent to participate. Before commencing each test, time was allowed for the participants to read the instructions specific to each tests and to take themselves through the practice questions. Participants were encouraged to alert the experimenter, if they had any questions regarding the instructions, or any difficulty with the practice items, so that additional assistance could be provided. Participants were instructed to answer every question, even if they had to guess. The seven sections of the test battery comprised the nine computerised cognitive abilities tests, six paper and
pencil cognitive abilities tests, three computer simulations, one personality test, and one career interests inventory. The order of presentation of the seven sections was counterbalanced across the three test sessions, to minimise order effects. No more than one complex computer simulated scenarios was presented in a single session.

The nine computerised cognitive abilities tests comprised the eight original tests employed in Study 1, and presented in the same order (60 minutes), plus the Paper Folding task (20 minutes). The six paper and pencil cognitive abilities tests were presented in the following order: Letter Series (5 minutes), Analogies (5 minutes), Problem Solving (22 minutes), Critical Reasoning (22 minutes), Raven’s Advanced Progressive Matrices (20 minutes), and General Knowledge (5 minutes). The three complex computer simulated scenarios included the Furniture Factory, employed in Study 1 (90 minutes), the Tailorshop (60 minutes), and the Forestry system (90 minutes). There were no time constraints for the completion of the OCEANIC personality questionnaire and the IDEAS career interest inventory, with most participants completing the questionnaires within 15 minutes and 10 minutes respectively. Total testing time for the entire test battery was 7 and ½ hours, although all time limits were generous and not exceeded in 95% of cases.

Participants were encouraged, if necessary, to take a break between tests, rather than during tests. All computerised tests were presented on Dell Pentium 4 computers. Scrap paper was provided for participant to use for scratch work, on the paper and pencil cognitive abilities tests, and the three complex computer simulated scenarios. Participants were not allowed to use paper for scratch work, or calculators, for any of the nine computerised cognitive abilities tasks, as this would diminish the load on
working memory during task performance. There were up to eight individuals present during any one, test session. Participants were seated apart to facilitate administration and privacy of responding. At the end of the third test session, part pants were provided with a debrief explaining the main aims of the experiment and when they would be contacted with their results.

6.4.1 Statistical Analyses

All statistical analyses in this study were performed using subprograms from the Statistical Package for Social Scientists (SPSS, 2003) computer package. The initial aim of the data analysis, was to establish the reliability and validity of the data obtained for each of the measures included in this study. Where possible, this was done by comparing data obtained in this study with that obtained in previous validation studies. The second stage of data analysis, employed correlational techniques and exploratory factor analyses, to investigate the underlying factor structure of the whole test battery. The main aim was to establish the nature of the relationship between cognitive abilities and complex problem solving on computer simulations tasks. The third stage of data analysis, employed regression techniques to investigate the amount of variance in complex problem solving performance accounted for by specific components of intelligence. The relationship of personality, career interests, and biodata with complex problem solving performance was also investigated. Although both Principal Components (PC) analysis and Maximum Likelihood (ML) factor solutions were computed and compared for interpretability, only the PC solutions are reported here, given the close approximation between the respective factor solutions.
CHAPTER 7

RESULTS AND DISCUSSION: STUDY 2

7.1 Data Screening

The measures of cognitive abilities, complex problem solving, personality, and interests that are analysed here, are described in detail in Section 6.2. The psychometric properties of such tasks were reviewed and evaluated. Prior to main analyses, data screening procedures were executed on the variables. Specifically, variables were checked for accuracy of data entry. SPSS FREQUENCIES was used for evaluations of the assumptions of multivariate analyses. With the use of a $p < .001$ criterion for Mahalanobis distance, no outliers among cases were identified, with the exception of 10 outliers on Forestry system (FSYS 2.0) task performance. The Forestry system identified 10 cases whose high FSYS 2.0, SKAPCOR scores (described in Section 6.2.5.6) resulted from problematic decisions making, rather than superior complex problem solving skills. SKAPCOR scores $> 100$ are identified by the system as outliers, the 10 outlying cases in Study 2, all had SKAPCOR scores $> 136$, well above the cut off. These outliers were removed where Forestry system performance scores were included in data analyses. No cases had missing data and no suppressor variables were found, $N = 296$. Examination of the correlation matrix reveals no evidence of multicollinearity and singularity among the independent variables. Residual scatterplots revealed that assumptions of normality, linearity, homoscedasticity and independence of residuals were met. Overall, the distributions of variables fit the assumptions of multivariate analysis. Exploratory statistical techniques were employed throughout data analyses. The remainder of assumptions (e.g., assumptions and practical considerations
underlying the application of factor analysis) will be examined as they occur in the presentation of results.

7.2 Descriptive Statistics for Cognitive Abilities Variables

Data screening procedures included the usual checks for normality, linearity, and multicollinearity. No problems were detected. There were 15 cognitive abilities tests, including 9 computerised tasks (inclusive of a measure of speed of test taking, Gs) and 6 paper and pencil tasks, described in Section 6.2.1 and Section 6.2.2 respectively. The 15 cognitive abilities tests plus the speed of test taking measure provide 16 cognitive abilities variables in total. The means, standard deviations, and Cronbach alpha reliability estimates for each of the cognitive abilities tests are presented in Table 7.1.
Table 7.1

Descriptive Statistics for Cognitive Abilities Tests

<table>
<thead>
<tr>
<th>Cognitive abilities</th>
<th>( \alpha )</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swaps</td>
<td>.97 (speed .92)</td>
<td>81.25</td>
<td>15.78</td>
</tr>
<tr>
<td>Triplets</td>
<td>.88 (speed .97)</td>
<td>85.71</td>
<td>15.31</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.80 (speed .89)</td>
<td>75.33</td>
<td>14.45</td>
</tr>
<tr>
<td>Proverbs Matching</td>
<td>.70 (speed .84)</td>
<td>60.13</td>
<td>15.67</td>
</tr>
<tr>
<td>Numerical Operations</td>
<td>.95 (speed .83)</td>
<td>84.11</td>
<td>12.45</td>
</tr>
<tr>
<td>Financial Reasoning</td>
<td>.80 (speed .81)</td>
<td>56.68</td>
<td>19.31</td>
</tr>
<tr>
<td>Line Length</td>
<td>.68 (speed .93)</td>
<td>65.17</td>
<td>13.06</td>
</tr>
<tr>
<td>Letter Spotting</td>
<td>.74 (speed .92)</td>
<td>56.35</td>
<td>12.71</td>
</tr>
<tr>
<td>Speed of Test Taking (of the above 8 computerised tests)</td>
<td>As above in brackets</td>
<td>18.73</td>
<td>3.48</td>
</tr>
<tr>
<td>Paper Folding</td>
<td>.86</td>
<td>12.47</td>
<td>4.05</td>
</tr>
<tr>
<td>General Knowledge</td>
<td>.40</td>
<td>9.30</td>
<td>2.37</td>
</tr>
<tr>
<td>Letter Series</td>
<td>.57</td>
<td>10.30</td>
<td>1.49</td>
</tr>
<tr>
<td>Esoteric Analogies</td>
<td>.52</td>
<td>8.26</td>
<td>2.07</td>
</tr>
<tr>
<td>Raven’s Advanced Progressive Matrices</td>
<td>.60</td>
<td>5.41</td>
<td>1.65</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>.76</td>
<td>7.51</td>
<td>3.41</td>
</tr>
<tr>
<td>Critical Reasoning</td>
<td>.68</td>
<td>8.86</td>
<td>2.94</td>
</tr>
</tbody>
</table>

\( N = 296 \)

It can be seen from Table 7.1 that the internal consistency reliability estimates (alpha) for the cognitive abilities tests ranged from .40 for General Knowledge to .97 for...
Swaps. The original 24 item, General Knowledge test, had poor reliability, $\alpha = .24$.

Inspection of item-total statistics, cronbach’s alpha if item deleted, indicated that removing items 6, 13, 14, 20, 9, 5, 11, and 4, would lead to the greatest increase in internal consistency reliability, $\alpha = .40$. The 16 remaining items are: 1, 2, 3, 7, 8, 10, 12, 15, 16, 17, 18, 19, 21, 22, 23, and 24. Despite the increase in reliability, the revised 16-item, General Knowledge test does not meet the $\alpha > .50$ cut-off and is not sufficiently reliable for research purposes. Overall, the majority of the cognitive abilities tests in the battery, showed reliability estimates of .70 or higher, indicating moderate to high internal consistency for their respective ratings.

### 7.2.1 Correlations Among Cognitive Abilities Variables

Table 7.2 displays the correlations among the 16 cognitive abilities variables.

#### Table 7.2

**Correlation Matrix of Cognitive Abilities Tests**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Swaps</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Triplets</td>
<td>.24</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Vocabulary</td>
<td>.13</td>
<td>.15</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Proverbs Matching</td>
<td>.21</td>
<td>.12</td>
<td>.56</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Numerical Operations</td>
<td>.39</td>
<td>.23</td>
<td>.03</td>
<td>.10</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Financial Reasoning</td>
<td>.41</td>
<td>.16</td>
<td>.18</td>
<td>.25</td>
<td>.41</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Line Length</td>
<td>.27</td>
<td>.12</td>
<td>.12</td>
<td>.17</td>
<td>.22</td>
<td>.27</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>8. Letter Spotting</td>
<td>.26</td>
<td>.22</td>
<td>.20</td>
<td>.22</td>
<td>.30</td>
<td>.25</td>
<td>.30</td>
<td>*</td>
</tr>
<tr>
<td>9. Speed</td>
<td>-.02</td>
<td>.17</td>
<td>-.17</td>
<td>.04</td>
<td>.05</td>
<td>-.05</td>
<td>.14</td>
<td>.00</td>
</tr>
<tr>
<td>10. Paper Folding</td>
<td>.37</td>
<td>.17</td>
<td>.14</td>
<td>.24</td>
<td>.33</td>
<td>.46</td>
<td>.34</td>
<td>.26</td>
</tr>
<tr>
<td>11. General Knowledge</td>
<td>.12</td>
<td>.03</td>
<td>.37</td>
<td>.25</td>
<td>.01</td>
<td>.24</td>
<td>.09</td>
<td>.13</td>
</tr>
<tr>
<td>13. Analogies</td>
<td>.47</td>
<td>.19</td>
<td>.45</td>
<td>.37</td>
<td>.23</td>
<td>.46</td>
<td>.13</td>
<td>.28</td>
</tr>
<tr>
<td>14. Raven’s Matrices</td>
<td>.32</td>
<td>.17</td>
<td>.04</td>
<td>.08</td>
<td>.32</td>
<td>.37</td>
<td>.21</td>
<td>.18</td>
</tr>
<tr>
<td>16. Critical Reasoning</td>
<td>.35</td>
<td>.09</td>
<td>.57</td>
<td>.44</td>
<td>.21</td>
<td>.42</td>
<td>.19</td>
<td>.18</td>
</tr>
</tbody>
</table>
As shown in Table 7.2, Vocabulary, Proverbs Matching, Critical Reasoning, and Esoteric Analogies were significantly correlated with one another ($r > .4$, $p < .05$), suggesting the possibility of a common basis for these ratings. Similarly, Problem Solving, Numerical Reasoning, Raven’s Progressive Matrices, Paper Folding, Financial Reasoning, Swaps, Letter Series, and Esoteric Analogies variables were significantly positively correlated with one another ($r > .40$, $p < .05$). The Line Length and Letter Spotting variables were likewise significantly intercorrelated ($r > .30$, $p < .05$). Sizeable correlations among most of the variables in Table 7.2 suggest factorability of the correlation matrix.
7.2.2 The Structure of Cognitive Abilities Variables

Assumptions and practical considerations underlying the application of Factor Analysis were considered before applying Principle Component (PC) analysis on the cognitive abilities tests. The sample size, $N = 296$ was acceptable and cognitive abilities variables were normally distributed. There were no outliers among cases. The assumption of multicollinearity and singularity is not relevant for PC and was not considered. Bartlett’s test of sphericity was large and significant. In addition, the Kaiser-Meyer-Olkin measure of sampling adequacy was greater than .6. Thus, factorability of the correlation matrix presented in Table 7.2 was assumed.

However, the General Knowledge and Triplets tests did not have correlations in excess of .3 with the other variables. In addition, removal of items to improve reliability failed to achieve adequate reliability of the General Knowledge test for research purposes, $\alpha < .50$ (Tabachnick & Fidell, 2000). Communalities for the General Knowledge test were also inadequate, $h^2 < .20$.

Communality for the Triplets test was also inadequate for research purposes, $h^2 < .20$. The internal consistency reliability for the Triplets test reported in Table 7.1, $\alpha = .88$, was from the test manual (Stankov, 1999). Cronbach’s alpha of the Triplets test could not be determined specifically for this sample because the Triplets test is automatically computer scored and item level data was unavailable. The low reliability of the Triplets test, in Study 2 of this thesis, could be explained by a design fault. While conducting the experiment, many participants informed the test administrator that they had missed the rule provided at the beginning of the Triplets test. Consequently, participants could
not decide whether subsequent items were consistent or inconsistent with the rule. In addition, the General Knowledge and Triplets tests were outliers among variables, these variables had low squared multiple correlations with all other variables, and low correlations with all pockets of intercorrelations (factors). Thus, both the General Knowledge and Triplets variables were removed from factor analysis of the cognitive abilities tests.

![Scree plot indicating the extraction of three factors.](image)

Factors with eigenvalues > 1 should be retained; the remaining factors are ‘scree’ and should not be extracted (Norman, 1987). Consistent with the scree plot in Figure 7.1, Table 7.3 shows the analysis extracted three factors with eigenvalues greater than 1, which together accounted for 58.43% of the variance in the cognitive abilities performance scores. Table 7.3 presents the pattern matrix, proportion of variance
accounted for, and communalities ($h^2$) for the cognitive abilities tests. All factor loadings greater than .30 are underlined to facilitate interpretation. The communality values for the cognitive abilities tests ranged from .36 for Letter Series to .78 for Vocabulary.

Table 7.3

Factor Loadings, Communalities ($h^2$), Percents of Variance for Principal Component Analysis and Oblique Rotation of Cognitive Abilities Tests

<table>
<thead>
<tr>
<th>Cognitive Abilities</th>
<th>Fluid Reasoning (Gf)</th>
<th>Crystallised Knowledge (Gc)</th>
<th>Broad Visualisation (Gv)</th>
<th>$h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven’s Matrices</td>
<td>.88</td>
<td>-.09</td>
<td>-.23</td>
<td>.64</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>.85</td>
<td>-.07</td>
<td>-.01</td>
<td>.68</td>
</tr>
<tr>
<td>Paper Folding</td>
<td>.70</td>
<td>.03</td>
<td>.06</td>
<td>.55</td>
</tr>
<tr>
<td>Financial Reasoning</td>
<td>.62</td>
<td>.09</td>
<td>.13</td>
<td>.52</td>
</tr>
<tr>
<td>Numerical Operations</td>
<td>.54</td>
<td>-.22</td>
<td>.36</td>
<td>.50</td>
</tr>
<tr>
<td>Esoteric Analogies</td>
<td>.54</td>
<td>.48</td>
<td>-.10</td>
<td>.67</td>
</tr>
<tr>
<td>Swaps</td>
<td>.54</td>
<td>.03</td>
<td>.25</td>
<td>.47</td>
</tr>
<tr>
<td>Letter Series</td>
<td>.47</td>
<td>-.09</td>
<td>.28</td>
<td>.36</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>-.18</td>
<td>.92</td>
<td>.07</td>
<td>.78</td>
</tr>
<tr>
<td>Proverbs Matching</td>
<td>-.14</td>
<td>.81</td>
<td>.16</td>
<td>.65</td>
</tr>
<tr>
<td>Critical Reasoning</td>
<td>.39</td>
<td>.66</td>
<td>-.14</td>
<td>.72</td>
</tr>
<tr>
<td>Letter Spotting</td>
<td>-.06</td>
<td>.16</td>
<td>.76</td>
<td>.60</td>
</tr>
<tr>
<td>Line Length</td>
<td>.06</td>
<td>.03</td>
<td>.67</td>
<td>.49</td>
</tr>
<tr>
<td>Eigenvalue</td>
<td>4.81</td>
<td>1.70</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>Percent of Variance</td>
<td>37.01</td>
<td>13.09</td>
<td>8.32</td>
<td></td>
</tr>
</tbody>
</table>

Factor Correlation Matrix

<table>
<thead>
<tr>
<th>Cognitive Abilities</th>
<th>Fluid Reasoning (Gf)</th>
<th>Crystallised Knowledge (Gc)</th>
<th>Broad Visualisation (Gv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluid Reasoning (Gf)</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crystallised Knowledge (Gc)</td>
<td>.37</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Broad Visualisation (Gv)</td>
<td>.36</td>
<td>.16</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note. Loadings > .40 are underlined.

Factor loadings > .40 are adequate for the sample size, $N > 200$ (Stevens, 1992). Oblique rotation was employed to simplify the pattern of factor loadings, this method of
rotation assumes intercorrelations amongst all of the cognitive abilities variables, which was the case, as shown in Table 7.2. As shown in the pattern matrix, displayed in Table 7.3, the first factor was defined as Fluid Reasoning (Gf), with Raven’s Advanced Progressive Matrices, Problem Solving, Paper Folding, Financial Reasoning, Numerical Operations, Swaps, and Letter Series loading highly (> .40) on this factor. These Gf tasks require inductive, deductive, conjunctive, and disjunctive reasoning to arrive at understanding relations among stimuli, comprehend implications, and draw inferences (Horn & Noll, 1993). Factor 2 was labelled Crystallised Knowledge (Gc), defined by Vocabulary, Proverbs Matching, and Critical Reasoning loading highly (> .40) on this factor. These Gc tasks indicate breadth and depth of the knowledge of the dominant culture (Horn & Noll, 1993). Two variables, Critical Reasoning and Esoteric Analogies, have dual loadings on both the Gf and Gc factors. This is consistent with previous research, which has found that the analogies task commonly loads on both Gf and Gc factors (Stankov, 1997). Common features of Critical Reasoning and Esoteric Analogies tasks are that they both require inductive thinking. Firstly, the participant must understand the premises using knowledge and literacy to construct a representation of the problem. Reading comprehension is important for building representations in Critical Reasoning (Wilhelm, 2005). This information processing requires crystallised knowledge (Gc). Subgroups of participants may employ different strategies for initial construction of problem content (Ford, 1995; Stenning & Oberlander, 1995; Sternberg & Turner, 1981). Secondly, a tentative hypothesis regarding the relationship between the premises and the target is constructed. Thirdly, models are evaluated, maintained, modified, or rejected (Wilhelm, 2005). In addition to crystallised knowledge (Gc), these last two stages involve fluid reasoning (Gf).
The third factor was defined as Broad Visualisation (Gv), with Letter Spotting and Line Length loading highly (> .40) on this factor. These Gv tests assess spatial skills. The factor correlation matrix indicated significant positive correlations between the Gf factor and both the Gc and Gv factors, $r = .37$, $p < .05$ and $r = .36$, $p < .05$, respectively. In contrast, the correlation between Gc and Gv factors was not significant, $r = .16$, $p > .05$.

During the experimental design phase of Study 2, Problem Solving, Financial Reasoning, and Numerical Operations were employed as measures of Quantitative Knowledge (Gq) requiring understanding and application of the concepts and skills of mathematics (Horn & Noll, 1993). However, Principle Component (PC) analysis of cognitive abilities tests did not define Gf and Gq as separate factors. The fact that Gq tests loaded on the Gf factor is compatible with Carroll’s (1993, p. 590) suggestion that the Gq factor (Carroll’s RQ) consistently loads on the Gf factor (Carroll’s 2F). Also during study design, Paper Folding was included in the test battery as a Gv marker, however, Paper Folding has loaded on the Gf factor. A possible explanation for the loading of Paper Folding on the Gf factor is that spatial tests, such as Paper Folding, tend to load on the general ‘g’ factor of intelligence, which is usually assessed by Raven’s Progressive Matrices (Lohman, Pellegrino, Alderton, & Regian, 1987). Thus, in Study 2, two marker tests rather than three define the Gv factor. It is acknowledged that this is inconsistent with the suggestion that at least three measures should be employed to define a factor (Dunteman, 1989). However, the Gv factor, which explains 8.32% of the variance, will be considered in data analyses, after accounting for Gf and Gc, which explain 37.01% and 13.09% of the variance respectively.
Factor scores (Kim & Mueller, 1978) were calculated for each of the 3 intelligence factors: Fluid Reasoning (Gf), Crystallised Knowledge (Gc), and Broad Visualisation (Gv). These factor scores were employed in correlational and regression analyses, investigating the relationship between specific components of intelligence and complex problem solving performance on computer simulated tasks. A confirmatory factor analysis (CFA) was also conducted to test the construct validity of the test battery. The CFA solution (Appendix H) is consistent with the principle components analysis (PCA) described in this section.

### 7.3 Descriptive Statistics for Complex Problem Solving Variables

The descriptive statistics for complex problem solving performance on the 3 computer simulations (Furniture Factory, Tailorshop, and Forestry System) are presented in Table 7.4. Aggregate scores are the average scores of all trials within a computer simulation, e.g. average of Furniture Factory performance scores trial 1 to trial 18. Goal achievement scores are calculated by summing the total number of trials in which the aim of the computer simulated task was achieved, e.g. the number of trials on the Furniture Factory task that employee efficiency was improved from trial to trial, out of a total of 18 trials. Final trial scores are the scores in the last trial of the game, e.g. single score of Furniture Factory performance on the 18th trial.
Table 7.4

Descriptive Statistics for Complex Problem Solving Variables

<table>
<thead>
<tr>
<th>Complex Problem Solving Tasks</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Furniture Factory (aggregate)</td>
<td>101.84</td>
<td>19.06</td>
</tr>
<tr>
<td>Furniture Factory (goal achievement)</td>
<td>11.40</td>
<td>6.88</td>
</tr>
<tr>
<td>Furniture Factory (final trial)</td>
<td>104.39</td>
<td>26.42</td>
</tr>
<tr>
<td>Tailorshop (aggregate)</td>
<td>206571.91</td>
<td>83846.56</td>
</tr>
<tr>
<td>Tailorshop (goal achievement)</td>
<td>2.01</td>
<td>2.84</td>
</tr>
<tr>
<td>Tailorshop (final trial)</td>
<td>143762.19</td>
<td>160625.55</td>
</tr>
<tr>
<td>Forestry System (aggregate)</td>
<td>12939996.55</td>
<td>1022864.53</td>
</tr>
<tr>
<td>Forestry System (goal achievement)</td>
<td>22.44</td>
<td>8.31</td>
</tr>
<tr>
<td>Forestry System (final trial)</td>
<td>11966552</td>
<td>2117857.50</td>
</tr>
</tbody>
</table>

N = 296 (Furniture Factory and Tailorshop)
N = 286 (Forestry System)

Reliabilities for were not calculated for the 3 complex problem solving tasks for reasons outlined in Section 4.1.2. Goal achievement scores were devised for this thesis to be a consistent scoring measure to allow comparison across computer simulations. Section 6.3.1 describes the development of the new goal achievement scores. Examination of the standard deviations shown in Table 7.4, are consistent with the fact that scores on the Tailorshop task were widely dispersed and did not fit a normal distribution. This finding is consistent with previous findings where Tailorshop raw scores differed considerably from a normal distribution (Süß, 2001).

In contrast, performance on the Furniture Factory and Forestry System tasks were normally distributed. The means of the ‘goal achievement’ scores, for each of the 3 simulations, were employed to calculate the average percentage of trials on which participants achieved the main goal of each computer simulated task. On average, participants achieved the goal of the Furniture Factory task on 63.33% of trials, \([\frac{11.40}{18 \text{ trials}} \times 100]\). Participants achieved the goals of the Tailorshop task and the
Forestry System task on 16.75% [(2.01 / 12 trials) X 100] and 44.88% [(22.44 / 50 trials) X 100] of trials respectively.

In summary, Tailorshop task performance scores are not normally distributed and participants experienced great difficulty in achieving the goal of the Tailorshop computer simulation. The 1\textsuperscript{st} trial of the Tailorshop task is the most important. Mistakes in early trials of the computer simulation cannot be recovered from in subsequent trials; this explains why scores are not normally distributed and could be problematic for the assessment of complex problem solving competencies. Thus, aggregate scores, goal achievement scores, and scores on the final trial of the Tailorshop task must be interpreted with caution.
7.3.1 Correlations Among Complex Problem Solving Variables

Correlations among the three complex problem solving variables are presented in Table 7.5.

Table 7.5
Correlation Matrix of Complex Problem Solving Variables

<table>
<thead>
<tr>
<th>CPS Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Furniture Factory (aggregate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Furniture Factory (goal achievement)</td>
<td>.91</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Furniture Factory (final trial)</td>
<td>.95</td>
<td>.85</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Tailorshop (aggregate)</td>
<td>.30</td>
<td>.23</td>
<td>.29</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Tailorshop (goal achievement)</td>
<td>.13</td>
<td>.15</td>
<td>.12</td>
<td>.49</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Tailorshop (final trial)</td>
<td>.30</td>
<td>.23</td>
<td>.29</td>
<td>.98</td>
<td>.55</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Forestry System (aggregate)</td>
<td>.21</td>
<td>.20</td>
<td>.18</td>
<td>.10</td>
<td>.14</td>
<td>.12</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Forestry System (goal achievement)</td>
<td>.25</td>
<td>.26</td>
<td>.22</td>
<td>.21</td>
<td>.19</td>
<td>.22</td>
<td>.85</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>9. Forestry System (final trial)</td>
<td>.21</td>
<td>.21</td>
<td>.17</td>
<td>.09</td>
<td>.10</td>
<td>.10</td>
<td>.94</td>
<td>.86</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296 (Furniture Factory and Tailorshop)
N = 286 (Forestry System)

Note. Ten participants were identified by SKAPCOR scores as exploiting loopholes on the Forestry System task and were removed from the sample.

Table 7.5 shows significant positive correlations between Furniture Factory performance and Forestry System performance for all 3 scoring methods: Final trial (week 18 of the Furniture Factory and month 50 of the Forestry System), aggregate (mean of all trials), and goal achievement scores (number of ‘won’ trials), \( r = .17, p < .05; r = .21, p < .05; r = .26, p < .05 \), respectively. The correlation between performance
scores on the two computer simulations is increased by employing aggregate scores rather than single final trial scores. Employing goal achievement scores rather than aggregate scores further increases the correlation. Both the goal achievement method and aggregated scoring methods reveal more shared variance between the tasks than a single overall performance score on the final trial of the computer simulation. Furthermore, for comparison of complex problem solving performance across different computer simulations, employing the more consistent goal achievement scoring technique over the aggregated scoring method identifies a stronger relationship. Thus, the newly devised scoring technique, goal achievement, may offer an improved method for comparison of performance across different complex computer simulated tasks.

Correlations between Tailorshop performance and both the Furniture Factory and Forestry System performance scores must be interpreted with caution for reasons outlined earlier. Despite the problems associated with Tailorshop performance scores, there was a significant correlation between performance on the Tailorshop and the Furniture Factory, $r = .30, p < .05$, and between the Tailorshop and the Forestry System, $r = .20, p < .05$. Overall, the intercorrelations between the 3 complex problem solving tasks were positive and significant. Results suggest that the three complex problem solving tasks are measuring some common abilities. The relationships between complex problem solving tasks can be manipulated by scoring procedures. Thus, consistent scoring procedures must be employed when investigating the relationship between computer simulated tasks. For example, participants who have exploited ‘loopholes’ must be removed from the sample, as these scores may corrupt true correlations between the measures. In addition, the distribution of scores is also an important consideration. The three different scoring methods will be employed to
investigate the relationship between specific components of intelligence and complex problem solving competencies.

7.4 The Relation of Complex Problem Solving and Intelligence

Relationships between intelligence and performance on the Furniture Factory, Tailorshop, and Forestry System computer simulations will now be described in turn.

7.4.1 The Relation of Furniture Factory Performance and Intelligence

Table 7.6 presents the correlations between the 16 cognitive abilities variables and complex problem solving performance on the Furniture Factory simulation.
Table 7.6

Correlation Matrix of Cognitive Abilities and Furniture Factory Performance

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Swaps</td>
<td>.</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Triplets</td>
<td>.24</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Vocabulary</td>
<td>.13</td>
<td>.15</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Proverbs Matching</td>
<td>.21</td>
<td>.12</td>
<td>.56</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Numerical Operations</td>
<td>.39</td>
<td>.23</td>
<td>.03</td>
<td>.10</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Financial Reasoning</td>
<td>.41</td>
<td>.16</td>
<td>.18</td>
<td>.25</td>
<td>.41</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Line Length</td>
<td>.27</td>
<td>.12</td>
<td>.12</td>
<td>.17</td>
<td>.22</td>
<td>.27</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Letter Spotting</td>
<td>.26</td>
<td>.22</td>
<td>.20</td>
<td>.22</td>
<td>.30</td>
<td>.25</td>
<td>.30</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>9. Speed</td>
<td>-.02</td>
<td>.17</td>
<td>-.17</td>
<td>.04</td>
<td>.05</td>
<td>-.05</td>
<td>.14</td>
<td>.00</td>
<td>*</td>
</tr>
<tr>
<td>11. General Knowledge</td>
<td>.12</td>
<td>.03</td>
<td>.37</td>
<td>.25</td>
<td>.01</td>
<td>.24</td>
<td>.09</td>
<td>.13</td>
<td>.01</td>
</tr>
<tr>
<td>13. Analogies</td>
<td>.47</td>
<td>.19</td>
<td>.45</td>
<td>.37</td>
<td>.23</td>
<td>.46</td>
<td>.13</td>
<td>.28</td>
<td>.03</td>
</tr>
<tr>
<td>14. Raven’s Matrices</td>
<td>.32</td>
<td>.17</td>
<td>.04</td>
<td>.08</td>
<td>.32</td>
<td>.37</td>
<td>.21</td>
<td>.18</td>
<td>.06</td>
</tr>
<tr>
<td>15. Problem Solving</td>
<td>.48</td>
<td>.19</td>
<td>.14</td>
<td>.20</td>
<td>.51</td>
<td>.53</td>
<td>.22</td>
<td>.19</td>
<td>-.05</td>
</tr>
<tr>
<td>16. Critical Reasoning</td>
<td>.35</td>
<td>.09</td>
<td>.57</td>
<td>.44</td>
<td>.21</td>
<td>.42</td>
<td>.19</td>
<td>.18</td>
<td>.02</td>
</tr>
<tr>
<td>17. Furniture (aggregate)</td>
<td>.15</td>
<td>.24</td>
<td>.17</td>
<td>.13</td>
<td>.18</td>
<td>.24</td>
<td>.13</td>
<td>.17</td>
<td>.19</td>
</tr>
</tbody>
</table>

Table 7.6 continued below

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Paper Folding</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. General Knowledge</td>
<td>.21</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Letter Series</td>
<td>.30</td>
<td>.11</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Analogies</td>
<td>.46</td>
<td>.24</td>
<td>.37</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Raven’s Matrices</td>
<td>.56</td>
<td>.16</td>
<td>.24</td>
<td>.47</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Problem Solving</td>
<td>.52</td>
<td>.17</td>
<td>.41</td>
<td>.43</td>
<td>.55</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Critical Reasoning</td>
<td>.38</td>
<td>.25</td>
<td>.20</td>
<td>.59</td>
<td>.41</td>
<td>.40</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Furniture (aggregate)</td>
<td>.22</td>
<td>.10</td>
<td>.21</td>
<td>.34</td>
<td>.17</td>
<td>.12</td>
<td>.36</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Furniture (goal achievement)</td>
<td>.20</td>
<td>.13</td>
<td>.17</td>
<td>.37</td>
<td>.15</td>
<td>.10</td>
<td>.41</td>
<td>.91</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>19. Furniture (final trial)</td>
<td>.22</td>
<td>.08</td>
<td>.23</td>
<td>.33</td>
<td>.17</td>
<td>.13</td>
<td>.33</td>
<td>.95</td>
<td>.85</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296
Table 7.7

Correlation Matrix of Intelligence Factors and Furniture Factory Performance

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fluid Reasoning (Gf)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Crystallised Knowledge (Gc)</td>
<td>.45</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Broad Visualisation (Gv)</td>
<td>.34</td>
<td>.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Furniture Factory (aggregate)</td>
<td>.22</td>
<td>.40</td>
<td>.18</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Furniture Factory (goal achievement)</td>
<td>.21</td>
<td>.41</td>
<td>.16</td>
<td>.91</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>6. Furniture Factory (final trial)</td>
<td>.22</td>
<td>.37</td>
<td>.25</td>
<td>.95</td>
<td>.85</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296

7.4.2 The Relation of Tailorshop Performance and Intelligence

Table 7.8

Correlation Matrix of Cognitive Abilities and Tailorshop Performance

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Swaps</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Triplets</td>
<td>.24</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Vocabulary</td>
<td>.13</td>
<td>.15</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Proverbs Matching</td>
<td>.21</td>
<td>.12</td>
<td>.56</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Numerical Operations</td>
<td>.39</td>
<td>.23</td>
<td>.03</td>
<td>.10</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Financial Reasoning</td>
<td>.41</td>
<td>.16</td>
<td>.18</td>
<td>.25</td>
<td>.41</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Line Length</td>
<td>.27</td>
<td>.12</td>
<td>.12</td>
<td>.17</td>
<td>.22</td>
<td>.27</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Letter Spotting</td>
<td>.26</td>
<td>.22</td>
<td>.20</td>
<td>.22</td>
<td>.30</td>
<td>.25</td>
<td>.30</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>9. Speed</td>
<td>-.02</td>
<td>.17</td>
<td>-.17</td>
<td>.04</td>
<td>.05</td>
<td>-.05</td>
<td>.14</td>
<td>.00</td>
<td>*</td>
</tr>
<tr>
<td>11. General Knowledge</td>
<td>.12</td>
<td>.03</td>
<td>.37</td>
<td>.25</td>
<td>.01</td>
<td>.24</td>
<td>.09</td>
<td>.13</td>
<td>.01</td>
</tr>
<tr>
<td>13. Analogies</td>
<td>.47</td>
<td>.19</td>
<td>.45</td>
<td>.37</td>
<td>.23</td>
<td>.46</td>
<td>.13</td>
<td>.28</td>
<td>.03</td>
</tr>
<tr>
<td>14. Raven’s Matrices</td>
<td>.32</td>
<td>.17</td>
<td>.04</td>
<td>.08</td>
<td>.32</td>
<td>.37</td>
<td>.21</td>
<td>.18</td>
<td>.06</td>
</tr>
<tr>
<td>15. Problem Solving</td>
<td>.48</td>
<td>.19</td>
<td>.14</td>
<td>.20</td>
<td>.51</td>
<td>.53</td>
<td>.22</td>
<td>.19</td>
<td>-.05</td>
</tr>
<tr>
<td>16. Critical Reasoning</td>
<td>.35</td>
<td>.09</td>
<td>.57</td>
<td>.44</td>
<td>.21</td>
<td>.42</td>
<td>.19</td>
<td>.18</td>
<td>.02</td>
</tr>
<tr>
<td>17. Tailorshop (aggregate)</td>
<td>.14</td>
<td>.10</td>
<td>.08</td>
<td>.09</td>
<td>.18</td>
<td>.13</td>
<td>-.05</td>
<td>.03</td>
<td>.06</td>
</tr>
<tr>
<td>18. Tailorshop (goal achievmt)</td>
<td>.05</td>
<td>.03</td>
<td>.04</td>
<td>-.02</td>
<td>.25</td>
<td>.19</td>
<td>.05</td>
<td>.07</td>
<td>.01</td>
</tr>
<tr>
<td>19. Tailorshop (final trial)</td>
<td>.15</td>
<td>.10</td>
<td>.06</td>
<td>.08</td>
<td>.22</td>
<td>.15</td>
<td>-.05</td>
<td>.05</td>
<td>.08</td>
</tr>
</tbody>
</table>
Table 7.8 continued below

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Paper Folding</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. General Knowledge</td>
<td>.21</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Letter Series</td>
<td>.30</td>
<td>.11</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Analogies</td>
<td>.46</td>
<td>.24</td>
<td>.37</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Raven’s Matrices</td>
<td>.56</td>
<td>.16</td>
<td>.24</td>
<td>.47</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Problem Solving</td>
<td>.52</td>
<td>.17</td>
<td>.41</td>
<td>.43</td>
<td>.55</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Critical Reasoning</td>
<td>.38</td>
<td>.25</td>
<td>.20</td>
<td>.59</td>
<td>.41</td>
<td>.40</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Tailorshop (aggregate)</td>
<td>.14</td>
<td>-.01</td>
<td>.13</td>
<td>.22</td>
<td>.22</td>
<td>.11</td>
<td>.29</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Tailorshop (goal achievement)</td>
<td>.20</td>
<td>.07</td>
<td>.25</td>
<td>.25</td>
<td>.34</td>
<td>.25</td>
<td>.29</td>
<td>.49</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>19. Tailorshop (final trial)</td>
<td>.17</td>
<td>.01</td>
<td>.15</td>
<td>.25</td>
<td>.25</td>
<td>.15</td>
<td>.28</td>
<td>.98</td>
<td>.55</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296

Table 7.9

Correlation Matrix of Intelligence Factors and Tailorshop Performance

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fluid Reasoning (Gf)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Crystallised Knowledge (Gc)</td>
<td>.45</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Broad Visualisation (Gv)</td>
<td>.34</td>
<td>.20</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Tailorshop (aggregate)</td>
<td>.24</td>
<td>.39</td>
<td>.05</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Tailorshop (goal achievement)</td>
<td>.37</td>
<td>.29</td>
<td>.04</td>
<td>.49</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>6. Tailorshop (final trial)</td>
<td>.29</td>
<td>.37</td>
<td>.06</td>
<td>.98</td>
<td>.55</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296
### 7.4.3 The Relation of Forestry System Performance and Intelligence

Table 7.10

Correlation Matrix of Cognitive Abilities and Forestry System Performance

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Swaps</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Triplets</td>
<td>.24</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Vocabulary</td>
<td>.13</td>
<td>.15</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Proverbs Matching</td>
<td>.21</td>
<td>.12</td>
<td>.56</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Numerical Operations</td>
<td>.39</td>
<td>.23</td>
<td>.03</td>
<td>.10</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Financial Reasoning</td>
<td>.41</td>
<td>.16</td>
<td>.18</td>
<td>.25</td>
<td>.41</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Line Length</td>
<td>.27</td>
<td>.12</td>
<td>.12</td>
<td>.17</td>
<td>.22</td>
<td>.27</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Letter Spotting</td>
<td>.26</td>
<td>.22</td>
<td>.20</td>
<td>.22</td>
<td>.30</td>
<td>.25</td>
<td>.30</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>9. Speed</td>
<td>-02</td>
<td>.17</td>
<td>-17</td>
<td>.04</td>
<td>.05</td>
<td>.14</td>
<td>.00</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>11. General Knowledge</td>
<td>.12</td>
<td>.03</td>
<td>.37</td>
<td>.25</td>
<td>.01</td>
<td>.24</td>
<td>.09</td>
<td>.13</td>
<td>.01</td>
</tr>
<tr>
<td>13. Analogies</td>
<td>.47</td>
<td>.19</td>
<td>.45</td>
<td>.37</td>
<td>.23</td>
<td>.46</td>
<td>.13</td>
<td>.28</td>
<td>.03</td>
</tr>
<tr>
<td>14. Raven’s Matrices</td>
<td>.32</td>
<td>.17</td>
<td>.04</td>
<td>.08</td>
<td>.32</td>
<td>.37</td>
<td>.21</td>
<td>.18</td>
<td>.06</td>
</tr>
<tr>
<td>15. Problem Solving</td>
<td>.48</td>
<td>.19</td>
<td>.14</td>
<td>.20</td>
<td>.51</td>
<td>.53</td>
<td>.22</td>
<td>.19</td>
<td>-.05</td>
</tr>
<tr>
<td>16. Critical Reasoning</td>
<td>.35</td>
<td>.09</td>
<td>.57</td>
<td>.44</td>
<td>.21</td>
<td>.42</td>
<td>.19</td>
<td>.18</td>
<td>.02</td>
</tr>
<tr>
<td>17. Forestry (aggregate)</td>
<td>.18</td>
<td>.25</td>
<td>.19</td>
<td>.19</td>
<td>.11</td>
<td>.18</td>
<td>.06</td>
<td>.18</td>
<td>.16</td>
</tr>
<tr>
<td>18. Forestry (goal achievevt)</td>
<td>.26</td>
<td>.24</td>
<td>.22</td>
<td>.23</td>
<td>.09</td>
<td>.25</td>
<td>.03</td>
<td>.17</td>
<td>.16</td>
</tr>
<tr>
<td>19. Forestry (final trial)</td>
<td>.23</td>
<td>.26</td>
<td>.17</td>
<td>.19</td>
<td>.10</td>
<td>.19</td>
<td>.05</td>
<td>.19</td>
<td>.19</td>
</tr>
</tbody>
</table>

Table 7.10 continued below

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>10. Paper Folding</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. General Knowledge</td>
<td>.21</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Letter Series</td>
<td>.30</td>
<td>.11</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Analogies</td>
<td>.46</td>
<td>.24</td>
<td>.37</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Raven’s Matrices</td>
<td>.56</td>
<td>.16</td>
<td>.24</td>
<td>.47</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Problem Solving</td>
<td>.52</td>
<td>.17</td>
<td>.41</td>
<td>.43</td>
<td>.55</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Critical Reasoning</td>
<td>.38</td>
<td>.25</td>
<td>.20</td>
<td>.59</td>
<td>.41</td>
<td>.40</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Forestry (goal achievevt)</td>
<td>.35</td>
<td>.16</td>
<td>.24</td>
<td>.38</td>
<td>.27</td>
<td>.21</td>
<td>.30</td>
<td>.85</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>19. Forestry (final trial)</td>
<td>.27</td>
<td>.13</td>
<td>.16</td>
<td>.24</td>
<td>.22</td>
<td>.23</td>
<td>.23</td>
<td>.94</td>
<td>.86</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 286
Table 7.11

Correlation Matrix of Intelligence Factors and Forestry System Performance

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fluid Reasoning (Gf)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Crystallised Knowledge (Gc)</td>
<td>.45</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Broad Visualisation (Gv)</td>
<td>.34</td>
<td>.20</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Forestry System (aggregate)</td>
<td>.31</td>
<td>.29</td>
<td>.19</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Forestry System (goal achievement)</td>
<td>.38</td>
<td>.39</td>
<td>.18</td>
<td>.85</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>6. Forestry System (final trial)</td>
<td>.28</td>
<td>.28</td>
<td>.19</td>
<td>.94</td>
<td>.86</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 286

A common pattern emerges through analyses of correlations shown in Tables 7.6, 7.8, and 7.10. For all three computer simulations, the highest correlations ($r > .30$, $p < .05$) are found between computer simulation performance and Esoteric Analogies and Critical Reasoning. Recall from Table 7.3, that both of these cognitive abilities variables had dual loadings on Gf and Gc. These tests are ‘impure’ tests, they are not ‘pure’ measures of a single factor. Perhaps that is another reason why previous research has failed to establish a relationship between specific components of intelligence and complex problem solving. Individual differences researchers generally take a factor analytic approach, which strives to obtain ‘pure’ measures of cognitive abilities in order to predict performance on a criterion measure.

In summary, results of Study 2, suggest that there is a relationship between intelligence and complex problem solving. The strength of this relationship is often corrupted by inclusion of participants who exploit ‘loopholes’, by using final scores rather than aggregate or goal achievement scores, and by using ‘pure’ tests of cognitive abilities. Conversely, in order to establish a relationship between intelligence and complex
problem solving, participants who exploit ‘loopholes’ must be removed from the sample, aggregate or goal achievement scores of performance should be used, ‘impure’ tests of cognitive abilities that tap into multiple intelligences should also be employed.

7.4.4  Stepwise Regression of Complex Problem Solving on Intelligence Factors

Factor scores that were derived from the Principal Component analysis of cognitive variables, shown in Table 7.3, were entered into a stepwise regression with three steps, one for each factor. Factor 1 was entered in the first step and included Raven’s Progressive Matrices, Problem Solving, Paper Folding, Financial Reasoning, Numerical Operations, Swaps, and Letter Series variables, which were assumed to reflect a Fluid Reasoning (Gf) factor, based on the results of the factor analysis. The second step added factor 2, which included Vocabulary, Proverbs Matching, and Critical Reasoning variables, which were assumed to represent Crystallised Knowledge (Gc). The third step added factor 3, which included Letter Spotting and Line Length, which were assumed to load on a Broad Visualisation (Gv) factor.

Analysis was performed using SPSS REGRESSION and SPSS FREQUENCIES for evaluation of assumptions. With the use of a p < .001 criterion for Mahalanobis distance, no outliers among the cases were identified. No cases had missing data and no suppressor variables were found, N = 296.
Firstly, a stepwise regression with Furniture Factory (goal achievement) as the dependent variable was performed. Table 7.12 displays $R$, $R^2$, Adjusted $R^2$, and $R^2$ change after entry of the three independent variables.

Table 7.12

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>$R$</th>
<th>$R^2$ inc</th>
<th>Adjusted $R^2$</th>
<th>$R^2$ Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 Factor 1 (Fluid Reasoning)</td>
<td>.21</td>
<td>.04</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>Step 2 Factor 2 (Crystallised Knowledge)</td>
<td>.41</td>
<td>.17</td>
<td>.15</td>
<td>.13</td>
</tr>
</tbody>
</table>

As shown in Table 7.12, $R^2$ change was significantly different from zero at the end of steps 1 and 2. The third independent variable, Broad Visualisation (Gv), failed to meet the selection criteria. After step 1, with factor 1 (Fluid Reasoning) in the equation, $R^2$ inc = .04 (adjusted $R^2$ = .03), $F(1, 288) = 3.32, p < .01$. Therefore, when entered first in the equation, fluid reasoning (Gf) tasks pick up 4% of the variance in Furniture Factory performance. After step 2, with factor 2 (Crystallised Knowledge) added to the prediction of Furniture Factory performance, $R^2$ inc = .17 (adjusted $R^2$ = .15), $F(1, 287) = 7.50, p < .01$. Thus, the crystallised knowledge tasks, which load on factor 2, pick up 13% of the variance in Furniture Factory performance after fluid reasoning has been controlled for, $R^2$ change = .13.

Secondly, a stepwise regression with Tailorshop (goal achievement) as the dependent variable was performed. Table 7.13 displays $R$, $R^2$, Adjusted $R^2$, and $R^2$ change after entry of the three independent variables.
Table 7.13

Stepwise Regression of Tailorshop Performance on Intelligence Factors

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>R</th>
<th>$R^2$ inc</th>
<th>Adjusted $R^2$</th>
<th>$R^2$ Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 Factor 1 (Fluid Reasoning)</td>
<td>.37</td>
<td>.14</td>
<td>.13</td>
<td>.14</td>
</tr>
<tr>
<td>Step 2 Factor 2 (Crystallised Knowledge)</td>
<td>.39</td>
<td>.15</td>
<td>.13</td>
<td>.02</td>
</tr>
</tbody>
</table>

As shown in table 7.13, $R^2$ change was significantly different from zero at the end of steps 1 and 2. The third independent variable, Broad Visualisation (Gv), failed to meet the selection criteria. After step 1, with factor 1 (Fluid Reasoning) in the equation, $R^2$ inc = .14 (adjusted $R^2$ = .13), $F(1, 288) = 13.90$, $p < .01$. Therefore, when entered first in the equation, fluid reasoning (Gf) tasks pick up 14% of the variance in Tailorshop performance. After step 2, with factor 2 (Crystallised Knowledge) added to the prediction of Tailorshop performance, $R^2$ inc = .15 (adjusted $R^2$ = .13), $F(1, 287) = 7.83$, $p < .01$. Thus, the crystallised knowledge tasks, which load on factor 2, pick up 2% of the variance in Tailorshop performance after fluid reasoning has been controlled for, $R^2$ change = .02.

Thirdly, a stepwise regression with Forestry System (goal achievement) as the dependent variable was performed. Table 7.14 displays $R$, $R^2$, Adjusted $R^2$, and $R^2$ change after entry of the three independent variables.
Table 7.14

Stepwise Regression of Forestry System Performance on Intelligence Factors

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>R</th>
<th>$R^2_{\text{inc}}$</th>
<th>Adjusted $R^2$</th>
<th>$R^2_{\text{Change}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 Factor 1 (Fluid Reasoning)</td>
<td>.38</td>
<td>.15</td>
<td>.14</td>
<td>.15</td>
</tr>
<tr>
<td>Step 2 Factor 2 (Crystallised Knowledge)</td>
<td>.45</td>
<td>.21</td>
<td>.19</td>
<td>.06</td>
</tr>
</tbody>
</table>

Results in Table 7.14, show that $R^2$ change was significantly different from zero at the end of steps 1 and 2. The third independent variable, Broad Visualisation (Gv), failed to meet the selection criteria. After step 1, with factor 1 (Fluid Reasoning) in the equation, $R^2_{\text{inc}} = .15$ (adjusted $R^2 = .14$), $F(1, 278) = 18.06, p < .01$. Therefore, when entered first in the equation, fluid reasoning (Gf) tasks pick up 15% of the variance in Forestry System performance. After step 2, with factor 2 (Crystallised Knowledge) added to the prediction of Forestry System performance, $R^2_{\text{inc}} = .21$ (adjusted $R^2 = .19$), $F(1, 277) = 13.62, p < .01$. Thus, the crystallised knowledge tasks, which load on factor 2, pick up 6% of the variance in Furniture Factory performance after fluid reasoning has been controlled for, $R^2_{\text{change}} = .06$.

Overall, the main intelligence factor accounting for variance in Furniture Factory performance was crystallised knowledge (Gc), which accounted for 13% of the variance. Fluid Reasoning (Gf) was the main intelligence factor accounting for 14% of the variance in Tailorshop, and 15% of the variance in Forestry System performance. In the above regressions, cognitive abilities factors were entered as the independent variables and computer simulation scores were entered as the dependent variables. However, it is important to bear in mind that these results must be interpreted with
caution because the direction of causality is unknown (Wenke, Frensch, & Funke, 2005). It is unknown whether individual differences in intelligence cause individual differences in problem solving competencies, or whether the direction of causality is the other way around. This causes problems for traditional statistical analyses techniques such as stepwise regression, in which the independent and dependent variables must be specified. For these same reasons, confirmatory analyses may not be appropriate. The direction of the relationship, between intelligence and complex problem solving, described by the theoretical model (Figure 4.1) must also be interpreted with caution. Where possible, analyses of results in Studies 1 and 2 have employed an exploratory approach.

7.5  The Relation of Complex Problem Solving and Personality Factors

The relationship between complex problem solving performance and the Five Factor model of personality (Costa & McCrae, 1992) was explored for each of the three computer simulations. Table 7.15 shows the correlations between personality and complex problem solving performance on the Furniture Factory.
Correlation Matrix of Personality and Complex Problem Solving Performance on the Furniture Factory

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Openness</td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Conscientiousness</td>
<td>.22</td>
<td>.88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Extraversion</td>
<td>.08</td>
<td>.04</td>
<td>.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Agreeableness</td>
<td>.29</td>
<td>.28</td>
<td>.40</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Neuroticism</td>
<td>.11</td>
<td>.02</td>
<td>-.20</td>
<td>-.10</td>
<td>.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Furniture Factory (aggregate)</td>
<td>.01</td>
<td>-.04</td>
<td>.04</td>
<td>-.06</td>
<td>-.08</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Furniture Factory (goal achievement)</td>
<td>.01</td>
<td>-.05</td>
<td>.05</td>
<td>-.02</td>
<td>-.10</td>
<td>.91</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>8. Furniture Factory (final trial)</td>
<td>-.04</td>
<td>-.03</td>
<td>.06</td>
<td>-.05</td>
<td>-.17</td>
<td>.95</td>
<td>.85</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296

Note. Cronbach alpha reliability estimates for personality variables are shown on the diagonal.

Correlations displayed in Table 7.15 show that there is no significant correlation between complex problem solving performance on the Furniture Factory and the personality factors of openness, conscientiousness, extraversion, and agreeableness. Correlations range from the lowest (r = .01, p > .05) for openness and Furniture Factory performance over all trials, to the highest (r = -.06, p > .05) for agreeableness and Furniture Factory performance across all trials of the computer simulation. In contrast, a significant negative correlation, albeit a rather low correlation, was observed between neuroticism and complex problem solving performance on the Furniture Factory task, r = -.17, p < .05, on the final trial of the simulation.

Correlations between personality and complex problem solving performance on the Tailorship task are presented in Table 7.16.
Correlation Matrix of Personality and Complex Problem Solving Performance on the Tailorshop

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Openness</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Conscientiousness</td>
<td>.22</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Extraversion</td>
<td>.08</td>
<td>.04</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Agreeableness</td>
<td>.29</td>
<td>.28</td>
<td>.40</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Neuroticism</td>
<td>.11</td>
<td>.02</td>
<td>-.20</td>
<td>-.10</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Tailorshop (aggregate)</td>
<td>.05</td>
<td>.00</td>
<td>.06</td>
<td>.03</td>
<td>-.09</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Tailorshop (goal achvmnt)</td>
<td>.12</td>
<td>.03</td>
<td>-.06</td>
<td>-.03</td>
<td>-.21</td>
<td>.49</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>8. Tailorshop (final trial)</td>
<td>.07</td>
<td>-.01</td>
<td>.01</td>
<td>.00</td>
<td>-.09</td>
<td>.98</td>
<td>.55</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296

Correlations displayed in Table 7.16 show that there is no significant correlation between complex problem solving performance on the Tailorshop and the personality factors of openness, conscientiousness, extraversion, and agreeableness. Correlations range from the lowest (r = .00, p > .05) for openness and agreeableness with Tailorshop performance over all trials and on the final trial respectively, to the highest (r = .12, p > .05) for openness and Tailorshop performance across all trials of the computer simulation. In contrast, a significant negative correlation, albeit a rather low correlation, was observed between neuroticism and complex problem solving performance on the Tailorshop task, r = -.21, p < .05, for the newly devised goal achievement score.

Correlations between personality and complex problem solving performance on the Forestry System are presented in Table 7.17
Table 7.17

Correlation Matrix of Personality and Complex Problem Solving Performance on the Forestry System

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Openness</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Conscientiousness</td>
<td>.22</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Extraversion</td>
<td>.08</td>
<td>.04</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Agreeableness</td>
<td>.29</td>
<td>.28</td>
<td>.40</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Neuroticism</td>
<td>.11</td>
<td>.02</td>
<td>-.20</td>
<td>-.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Forestry System (aggregate)</td>
<td>-.01</td>
<td>.06</td>
<td>.00</td>
<td>.06</td>
<td>-.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Forestry System (goal achievement)</td>
<td>.00</td>
<td>-.02</td>
<td>-.08</td>
<td>.04</td>
<td>-.14</td>
<td>.85</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>8. Forestry System (final trial)</td>
<td>-.03</td>
<td>.02</td>
<td>-.07</td>
<td>.04</td>
<td>-.10</td>
<td>.94</td>
<td>.86</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 286

Correlations displayed in Table 7.17 show that there is no significant correlation between personality and complex problem solving performance on the Forestry System. Correlations range from the lowest (r = .00, p > .05) for openness and Forestry System performance over all trials, to the highest (r = -.14, p > .05) for neuroticism and Forestry System performance over all trials of the computer simulation.

Overall, results indicate that the relationship between complex problem solving (assessed by Furniture Factory and Tailorshop computer simulations) and neuroticism is negative and significant. Neutoticism reflects emotional instability, so the negative correlation between neuroticism and complex problem solving performance suggests that participants with greater emotional stability (or lower neuroticism) tend to achieve higher performance scores on complex problem solving tasks. However, there were no significant correlations observed between complex problem solving, assessed by the Forestry System computer simulation, and any of the 5 personality factors. The results for complex problem solving (assessed by Furniture Factory and Tailorshop tasks) are consistent with recent interest in the role of emotion in complex problem solving.
performance, where negative emotions have a detrimental effect on complex problem solving (Spering, Wagener, & Funke, 2005).

### 7.6 The Relation of Complex Problem Solving and Interest Variables

The relationship between complex problem solving performance and 16 domains of career interests was explored for each of the three computer simulations. Table 7.18 shows the correlations between careers interests and complex problem solving performance on the Furniture Factory.

#### Table 7.18

**Correlation Matrix of Interests and Furniture Factory Performance**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mechanical/Fixing</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Protective Services</td>
<td>.48</td>
<td>.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Nature/Outdoors</td>
<td>.49</td>
<td>.33</td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Mathematics</td>
<td>.38</td>
<td>.25</td>
<td>.17</td>
<td>.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Science</td>
<td>.40</td>
<td>.34</td>
<td>.32</td>
<td>.60</td>
<td>.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Medical</td>
<td>.15</td>
<td>.16</td>
<td>.21</td>
<td>.37</td>
<td>.57</td>
<td>.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Creative Arts</td>
<td>.29</td>
<td>.04</td>
<td>.42</td>
<td>.16</td>
<td>.23</td>
<td>.26</td>
<td>.82</td>
<td></td>
</tr>
<tr>
<td>8. Writing</td>
<td>.05</td>
<td>-.08</td>
<td>.16</td>
<td>-.02</td>
<td>.13</td>
<td>.22</td>
<td>.59</td>
<td>.90</td>
</tr>
<tr>
<td>9. Community Service</td>
<td>.15</td>
<td>.18</td>
<td>.38</td>
<td>.17</td>
<td>.12</td>
<td>.34</td>
<td>.38</td>
<td>.27</td>
</tr>
<tr>
<td>10. Educating</td>
<td>.17</td>
<td>.30</td>
<td>.30</td>
<td>.17</td>
<td>.06</td>
<td>.24</td>
<td>.22</td>
<td>.20</td>
</tr>
<tr>
<td>11. Child Care</td>
<td>.08</td>
<td>.05</td>
<td>.31</td>
<td>.08</td>
<td>-.08</td>
<td>.18</td>
<td>.36</td>
<td>.22</td>
</tr>
<tr>
<td>12. Public Speaking</td>
<td>.11</td>
<td>.10</td>
<td>.17</td>
<td>.11</td>
<td>.22</td>
<td>.24</td>
<td>.32</td>
<td>.45</td>
</tr>
<tr>
<td>14. Sales</td>
<td>.17</td>
<td>.30</td>
<td>.09</td>
<td>.22</td>
<td>.11</td>
<td>.11</td>
<td>.18</td>
<td>.02</td>
</tr>
<tr>
<td>15. Office Practices</td>
<td>.18</td>
<td>.23</td>
<td>.13</td>
<td>.32</td>
<td>.26</td>
<td>.12</td>
<td>.16</td>
<td>.09</td>
</tr>
<tr>
<td>16. Food Service</td>
<td>.30</td>
<td>.12</td>
<td>.44</td>
<td>.19</td>
<td>.23</td>
<td>.23</td>
<td>.59</td>
<td>.32</td>
</tr>
<tr>
<td>17. Furniture Factory (aggregate)</td>
<td>.03</td>
<td>.03</td>
<td>.11</td>
<td>.02</td>
<td>.03</td>
<td>-.01</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>18. Furniture Factory (goal achievement)</td>
<td>.01</td>
<td>.07</td>
<td>.11</td>
<td>-.03</td>
<td>.00</td>
<td>-.03</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>19. Furniture Factory (final trial)</td>
<td>-.02</td>
<td>.01</td>
<td>.08</td>
<td>.02</td>
<td>.01</td>
<td>-.02</td>
<td>-.03</td>
<td>-.04</td>
</tr>
</tbody>
</table>
Table 7.18 continued below

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.90</td>
<td>.37</td>
<td>.54</td>
<td>.27</td>
<td>.36</td>
<td>.76</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>12. Public Speaking</td>
<td></td>
<td>.82</td>
<td></td>
<td>.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Business</td>
<td></td>
<td></td>
<td></td>
<td>.27</td>
<td>.36</td>
<td>.76</td>
<td>.03</td>
<td>.05</td>
</tr>
<tr>
<td>14. Sales</td>
<td>.23</td>
<td>.31</td>
<td>.57</td>
<td>.65</td>
<td></td>
<td>.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Office Practices</td>
<td>.22</td>
<td></td>
<td></td>
<td>.02</td>
<td></td>
<td></td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>16. Food Service</td>
<td>.44</td>
<td>.35</td>
<td>.37</td>
<td>.37</td>
<td>.48</td>
<td>.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Furniture Factory (aggregate)</td>
<td>-.02</td>
<td>.00</td>
<td>.05</td>
<td>.03</td>
<td>.05</td>
<td>.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Furniture Factory (goal achievement)</td>
<td>-.05</td>
<td>-.02</td>
<td>.04</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>.91</td>
<td>*</td>
</tr>
<tr>
<td>19. Furniture Factory (final trial)</td>
<td>.00</td>
<td>.03</td>
<td>.06</td>
<td>.06</td>
<td>-.03</td>
<td>.04</td>
<td>.95</td>
<td>.85</td>
</tr>
</tbody>
</table>

**N = 296**

Note. Cronbach alpha reliability estimates for career interests variables are shown on the diagonal.

Correlations displayed in Table 7.18 show that there is no significant correlation between career interests and complex problem solving performance on the Furniture Factory. Correlations range from the lowest ($r = .00, p > .05$) to the highest ($r = .07, p > .05$).

Correlations between career interests and complex problem solving performance on the Tailorshop task are presented in Table 7.19.
Table 7.19

Correlation Matrix of Interests and Tailorshop Performance

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mechanical/Fixing</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Protective Services</td>
<td>.48</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Nature/Outdoors</td>
<td>.49</td>
<td>.33</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Mathematics</td>
<td>.38</td>
<td>.25</td>
<td>.17</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Science</td>
<td>.40</td>
<td>.34</td>
<td>.32</td>
<td>.60</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Medical</td>
<td>.15</td>
<td>.16</td>
<td>.21</td>
<td>.37</td>
<td>.57</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Creative Arts</td>
<td>.29</td>
<td>.04</td>
<td>.42</td>
<td>.16</td>
<td>.23</td>
<td>.26</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Writing</td>
<td>.05</td>
<td>- .08</td>
<td>.16</td>
<td>- .02</td>
<td>.13</td>
<td>.22</td>
<td>.59</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>9. Community Service</td>
<td>.15</td>
<td>.18</td>
<td>.38</td>
<td>.17</td>
<td>.12</td>
<td>.34</td>
<td>.38</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Educating</td>
<td>.17</td>
<td>.30</td>
<td>.30</td>
<td>.17</td>
<td>.06</td>
<td>.24</td>
<td>.22</td>
<td>.20</td>
<td>.59</td>
<td>*</td>
</tr>
<tr>
<td>11. Child Care</td>
<td>.08</td>
<td>.05</td>
<td>.31</td>
<td>.08</td>
<td>- .08</td>
<td>.18</td>
<td>.36</td>
<td>.22</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>12. Public Speaking</td>
<td>.11</td>
<td>.10</td>
<td>.17</td>
<td>.11</td>
<td>.22</td>
<td>.24</td>
<td>.32</td>
<td>.45</td>
<td>.54</td>
<td>.40</td>
</tr>
<tr>
<td>14. Sales</td>
<td>.17</td>
<td>.30</td>
<td>.09</td>
<td>.22</td>
<td>.11</td>
<td>.11</td>
<td>.18</td>
<td>.02</td>
<td>.39</td>
<td>.27</td>
</tr>
<tr>
<td>15. Office Practices</td>
<td>.18</td>
<td>.23</td>
<td>.13</td>
<td>.32</td>
<td>.26</td>
<td>.12</td>
<td>.16</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Food Service</td>
<td>.30</td>
<td>.12</td>
<td>.44</td>
<td>.19</td>
<td>.23</td>
<td>.23</td>
<td>.59</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Tailorshop (aggregate)</td>
<td>.09</td>
<td>- .10</td>
<td>.14</td>
<td>.21</td>
<td>.13</td>
<td>.16</td>
<td>- .07</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Tailorshop (goal achievement)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Tailorshop (final trial)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.19 continued below

<table>
<thead>
<tr>
<th></th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Child Care</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Public Speaking</td>
<td>.37</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Business</td>
<td>.27</td>
<td>.54</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Sales</td>
<td>.23</td>
<td>.36</td>
<td>.76</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Office Practices</td>
<td>.22</td>
<td>.31</td>
<td>.57</td>
<td>.65</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Food Service</td>
<td>.44</td>
<td>.35</td>
<td>.37</td>
<td>.37</td>
<td>.48</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Tailorshop (aggregate)</td>
<td>.10</td>
<td>.20</td>
<td>.07</td>
<td>.02</td>
<td>- .03</td>
<td>- .01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Tailorshop (goal achievement)</td>
<td>- .12</td>
<td>.20</td>
<td>.06</td>
<td>.00</td>
<td>.02</td>
<td>- .07</td>
<td>.49</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>19. Tailorshop (final trial)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N = 296
Correlations displayed in Table 7.19 show that there are significant positive correlations between Tailorshop performance and the career interest variables of public speaking, science, and mathematics, correlations are $r = .22, p < .05$, $r = .29, p < .05$, $r = .23, p < .05$, respectively. All other correlations between Tailorshop performance and the remaining 13 career interest variables were not significant.

Correlations between career interests and complex problem solving performance on the Forestry System are presented in Table 7.20
Table 7.20

Correlation Matrix of Interests and Forestry System Performance

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mechanic/Fixing</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Protective Services</td>
<td>.48</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Nature/Outdoors</td>
<td>.49</td>
<td>.33</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Mathematics</td>
<td>.38</td>
<td>.25</td>
<td>.17</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Science</td>
<td>.40</td>
<td>.34</td>
<td>.32</td>
<td>.60</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Medical</td>
<td>.15</td>
<td>.16</td>
<td>.21</td>
<td>.37</td>
<td>.57</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Creative Arts</td>
<td>.29</td>
<td>.04</td>
<td>.42</td>
<td>.16</td>
<td>.23</td>
<td>.26</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Writing</td>
<td>.05</td>
<td>-.08</td>
<td>.16</td>
<td>-.02</td>
<td>.13</td>
<td>.22</td>
<td>.59</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Community Service</td>
<td>.15</td>
<td>.18</td>
<td>.38</td>
<td>.17</td>
<td>.12</td>
<td>.34</td>
<td>.38</td>
<td>.27</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>10. Educating</td>
<td>.17</td>
<td>.30</td>
<td>.30</td>
<td>.17</td>
<td>.06</td>
<td>.24</td>
<td>.22</td>
<td>.20</td>
<td>.59</td>
<td>*</td>
</tr>
<tr>
<td>11. Child Care</td>
<td>.08</td>
<td>.05</td>
<td>.31</td>
<td>.08</td>
<td>-.08</td>
<td>.18</td>
<td>.36</td>
<td>.22</td>
<td>.62</td>
<td>.78</td>
</tr>
<tr>
<td>12. Public Speaking</td>
<td>.11</td>
<td>.10</td>
<td>.17</td>
<td>.11</td>
<td>.22</td>
<td>.24</td>
<td>.32</td>
<td>.45</td>
<td>.54</td>
<td>.40</td>
</tr>
<tr>
<td>14. Sales</td>
<td>.17</td>
<td>.30</td>
<td>.09</td>
<td>.22</td>
<td>.11</td>
<td>.11</td>
<td>.18</td>
<td>.02</td>
<td>.39</td>
<td>.27</td>
</tr>
<tr>
<td>15. Office Practices</td>
<td>.18</td>
<td>.23</td>
<td>.13</td>
<td>.32</td>
<td>.26</td>
<td>.12</td>
<td>.16</td>
<td>.09</td>
<td>.36</td>
<td>.21</td>
</tr>
<tr>
<td>16. Food Service</td>
<td>.30</td>
<td>.12</td>
<td>.44</td>
<td>.19</td>
<td>.23</td>
<td>.23</td>
<td>.59</td>
<td>.32</td>
<td>.47</td>
<td>.33</td>
</tr>
<tr>
<td>17. Forestry System (aggregate)</td>
<td>.12</td>
<td>.06</td>
<td>.07</td>
<td>.03</td>
<td>.16</td>
<td>.20</td>
<td>.07</td>
<td>.03</td>
<td>.07</td>
<td>.03</td>
</tr>
<tr>
<td>18. Forestry System (goal achvmnt)</td>
<td>.09</td>
<td>.03</td>
<td>.04</td>
<td>.06</td>
<td>.15</td>
<td>.13</td>
<td>.02</td>
<td>.04</td>
<td>.05</td>
<td>.03</td>
</tr>
<tr>
<td>19. Forestry System (final trial)</td>
<td>.09</td>
<td>.06</td>
<td>.04</td>
<td>.03</td>
<td>.16</td>
<td>.20</td>
<td>.10</td>
<td>.05</td>
<td>.08</td>
<td>.04</td>
</tr>
</tbody>
</table>

Table 7.20 continued below

<table>
<thead>
<tr>
<th></th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Child Care</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Public Speaking</td>
<td>.37</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Business</td>
<td>.27</td>
<td>.54</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Sales</td>
<td>.23</td>
<td>.36</td>
<td>.76</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Office Practices</td>
<td>.22</td>
<td>.31</td>
<td>.57</td>
<td>.65</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. Food Service</td>
<td>.44</td>
<td>.35</td>
<td>.37</td>
<td>.37</td>
<td>.48</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Forestry System (aggregate)</td>
<td>.02</td>
<td>.07</td>
<td>.02</td>
<td>-.02</td>
<td>.04</td>
<td>-.02</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Forestry System (goal achvmnt)</td>
<td>.00</td>
<td>.06</td>
<td>.00</td>
<td>-.05</td>
<td>.02</td>
<td>-.02</td>
<td>.85</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>19. Forestry System (final trial)</td>
<td>.01</td>
<td>.09</td>
<td>.06</td>
<td>.02</td>
<td>.07</td>
<td>.05</td>
<td>.94</td>
<td>.86</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296
Correlations displayed in Table 7.20 show that there is a significant positive correlation between Forestry System performance and the medical career interest variable, $r = .20, p < .05$. All other correlations between Forestry System performance and the remaining 15 career interest variables were not significant.

Overall, results indicate that the relationship between complex problem solving (assessed by Tailorshop, and Forestry System tasks) and career interests is significant for realistic and investigative interests (Holland, 1959, 1973). These results are consistent with the suggested link between realistic and investigative interests and intelligence (Ackerman & Heggestad, 1997). However, the correlations reported here are rather low and thus no further speculation will be entered into.

### 7.7 The Relation of Complex Problem Solving and Biodata

The correlations between complex problem solving and biodata, such as gender and University Admissions Index (UAI), are shown in Table 7.21.
Table 7.21

Correlation Matrix of Biodata with Intelligence and Complex Problem Solving

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gender</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. UAI</td>
<td>.00</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Gf</td>
<td>.19</td>
<td>.36</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Gc</td>
<td>.09</td>
<td>.21</td>
<td>.45</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Gv</td>
<td>.05</td>
<td>.08</td>
<td>.34</td>
<td>.20</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Furniture Factory (aggregate)</td>
<td>.10</td>
<td>.08</td>
<td>.22</td>
<td>.40</td>
<td>.18</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>7. Furniture Factory (goal achvmnt)</td>
<td>.13</td>
<td>.04</td>
<td>.21</td>
<td>.41</td>
<td>.16</td>
<td>.91</td>
<td>*</td>
</tr>
<tr>
<td>8. Furniture Factory (final trial)</td>
<td>.08</td>
<td>.07</td>
<td>.22</td>
<td>.37</td>
<td>.25</td>
<td>.95</td>
<td>.85</td>
</tr>
<tr>
<td>9. Tailorshop (aggregate)</td>
<td>-.01</td>
<td>.04</td>
<td>.24</td>
<td>.39</td>
<td>.05</td>
<td>.30</td>
<td>.23</td>
</tr>
<tr>
<td>10. Tailorshop (goal achvmnt)</td>
<td>.07</td>
<td>-.04</td>
<td>.37</td>
<td>.29</td>
<td>.04</td>
<td>.13</td>
<td>.15</td>
</tr>
<tr>
<td>11. Tailorshop (final trial)</td>
<td>.01</td>
<td>.04</td>
<td>.29</td>
<td>.37</td>
<td>.06</td>
<td>.30</td>
<td>.23</td>
</tr>
<tr>
<td>12. Forestry System (aggregate)</td>
<td>.00</td>
<td>.03</td>
<td>.31</td>
<td>.29</td>
<td>.19</td>
<td>.21</td>
<td>.20</td>
</tr>
<tr>
<td>13. Forestry System (goal achvmnt)</td>
<td>.02</td>
<td>.00</td>
<td>.38</td>
<td>.39</td>
<td>.18</td>
<td>.25</td>
<td>.26</td>
</tr>
<tr>
<td>14. Forestry System (final trial)</td>
<td>.00</td>
<td>.04</td>
<td>.28</td>
<td>.28</td>
<td>.19</td>
<td>.21</td>
<td>.21</td>
</tr>
</tbody>
</table>

Table 7.21 continued below

<table>
<thead>
<tr>
<th></th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Furniture Factory (final trial)</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Tailorshop (aggregate)</td>
<td>.29</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Tailorshop (goal achvmnt)</td>
<td>.12</td>
<td>.49</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Tailorshop (final trial)</td>
<td>.29</td>
<td>.98</td>
<td>.55</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Forestry System (aggregate)</td>
<td>.18</td>
<td>.10</td>
<td>.14</td>
<td>.12</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Forestry System (goal achvmnt)</td>
<td>.22</td>
<td>.21</td>
<td>.19</td>
<td>.22</td>
<td>.85</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>14. Forestry System (final trial)</td>
<td>.17</td>
<td>.09</td>
<td>.10</td>
<td>.10</td>
<td>.94</td>
<td>.86</td>
<td>*</td>
</tr>
</tbody>
</table>

N = 296

As shown in Table 7.21, there are no significant correlations between gender and complex problem solving performance. Gender issues in this area of research were investigated in Study 1, Chapter 4 and were not a main aim of Study 2.
Table 7.21 shows that the University Admissions Index (UAI) is significantly correlated with fluid reasoning (Gf), \( r = .36, p < .05 \). In contrast, the correlation between UAI and computer simulation performance, for all three simulations, is not significant, the highest correlation is between UAI and Furniture Factory (aggregate) performance, \( r = .08, p > .05 \). Correlations presented in Table 7.5 indicate that there is common variance among the three simulations. A static criterion, such as UAI, is adequate for a static task, such as Raven’s Progressive Matrices (the most common Gf measure). However, finding a dynamic criterion that is related to dynamic tasks, such as computer simulations, is a major problem in this area of research (J. Beckmann, personal communication, September 20, 2005). Assessment centre tasks or job performance ratings could be adequate dynamic criteria for complex problem solving performance in future research.

### 7.8 Final Analyses: The Factor Structure of Complex Problem Solving Variables

SPSS was employed to investigate the structure underlying the matrix of intercorrelations obtained with the complex problem solving variables using Principle Components (PC) analysis with oblique rotation. Assumptions and practical considerations underlying the application of PC (outlined earlier in the factor analysis of cognitive abilities) were met. Factor analyses of the 3 simulations, Furniture Factory, Tailorshop, and Forestry System, will now be described in turn.
7.8.1 The Relation of the Furniture Factory Component with Intelligence Factors

Two factors were extracted during preliminary factor analysis of Furniture Factory performance. Performance during simulated week 1 (trial 1) loaded on a separate factor from the remaining 17 simulated weeks (trials 2 to 18). In subsequent analyses, Furniture Factory week 1 was removed from the factor analyses due to low initial communalities. In addition, the factors, in the 2 factor solution, are highly correlated, \( r = .80, p < .05 \), indicating that there is only one component. The removal of Furniture Factory week 1 performance scores from the analysis is consistent with the fact that week 1 was not considered to be a complete trial of the computer simulation. As outlined in Section 4.2.6, actual job completion times (roughly equivalent to Furniture Factory performance) are based on 2 actions in the 1st week, ‘job assignment’ and ‘production targets’. In subsequent trials of the simulation, actual job completion times are calculated based on ‘job assignment’ and ‘production targets’ in the current trial, in addition to ‘feedback’ and ‘rewards’ from the previous trial. Thus, Furniture Factory week 1 performance is based on 2 participant-actions, whereas, Furniture Factory performance during weeks 2 to 18 is based on 4-participant actions. After the removal of performance scores for week 1, performance scores for weeks 2 to 18 were reanalysed using Principle Components analysis. Table 7.22 displays the component matrix.
Table 7.22

Component Matrix for Furniture Factory performance

<table>
<thead>
<tr>
<th>Furniture Factory Performance</th>
<th>Component 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 18</td>
<td>.95</td>
</tr>
<tr>
<td>Trial 13</td>
<td>.95</td>
</tr>
<tr>
<td>Trial 12</td>
<td>.94</td>
</tr>
<tr>
<td>Trial 15</td>
<td>.94</td>
</tr>
<tr>
<td>Trial 10</td>
<td>.93</td>
</tr>
<tr>
<td>Trial 11</td>
<td>.92</td>
</tr>
<tr>
<td>Trial 17</td>
<td>.91</td>
</tr>
<tr>
<td>Trial 14</td>
<td>.91</td>
</tr>
<tr>
<td>Trial 16</td>
<td>.89</td>
</tr>
<tr>
<td>Trial 9</td>
<td>.89</td>
</tr>
<tr>
<td>Trial 8</td>
<td>.88</td>
</tr>
<tr>
<td>Trial 6</td>
<td>.86</td>
</tr>
<tr>
<td>Trial 7</td>
<td>.85</td>
</tr>
<tr>
<td>Trial 4</td>
<td>.83</td>
</tr>
<tr>
<td>Trial 5</td>
<td>.81</td>
</tr>
<tr>
<td>Trial 2</td>
<td>.71</td>
</tr>
<tr>
<td>Trial 3</td>
<td>.68</td>
</tr>
</tbody>
</table>

As shown in Table 7.22, Furniture Factory performance on trials 2 to 18 loaded on a single component, after the removal of trial 1. The component score derived from this analysis, accounts for 76.84% of the variance in Furniture Factory performance. This measure could be employed as an aggregated score of Furniture Factory performance, which may be a more accurate reflection of true performance than any of the scoring techniques suggested so far, e.g. mean performance across all trials, goal achievement, and final trial scores. By applying factor analytic techniques to computer simulation performance, irrelevant trials, which should not be included in calculations of mean performance or goal achievement scores, can be removed. Table 7.23 presents the correlation matrix of cognitive abilities factor scores and the Furniture Factory component score.
Table 7.23

Correlation Matrix of Cognitive Abilities Factors and the Furniture Factory Component Score

<table>
<thead>
<tr>
<th></th>
<th>Gf</th>
<th>Gc</th>
<th>Gv</th>
<th>Furniture Factory Factor Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluid Reasoning (Gf)</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Crystallised Knowledge (Gc)</td>
<td>.45</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad Visualisation (Gv)</td>
<td>.34</td>
<td>.20</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Furniture Factory Factor Score</td>
<td>.21</td>
<td>.39</td>
<td>.18</td>
<td>*</td>
</tr>
</tbody>
</table>

As shown in Table 7.23, employing a more accurate measure of overall Furniture Factory performance, based on a factor score, significant positive correlations between Furniture Factory performance and specific components of intelligence have emerged. The correlation between Furniture Factory performance and fluid reasoning (Gf) is significant, $r = .21$, $p < .05$. Similarly, the correlation between Furniture Factory performance and crystallised intelligence (Gc), is significant, $r = .39$, $p < .01$. These correlations are compatible with those obtained using alternative scoring methods, presented in Table 7.7.

7.8.2 The Relation of the Tailorshop Component with Intelligence Factors

Factor Analysis of performance scores on the Tailorshop and Forestry System computer simulations will be described next.

Performance on the 12 individual trials of the Tailorshop computer simulation are represented by ‘xgekap 1’ to ‘xgekap 12’, these scores represent total assets in the
current month (described earlier in Section 6.2.5.2). Table 7.24 displays the component matrix for Tailorshop performance scores.

Table 7.24

Component Matrix for Tailorshop performance

<table>
<thead>
<tr>
<th>Tailorshop Performance</th>
<th>Component 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 7</td>
<td>.99</td>
</tr>
<tr>
<td>Trial 6</td>
<td>.99</td>
</tr>
<tr>
<td>Trial 8</td>
<td>.98</td>
</tr>
<tr>
<td>Trial 9</td>
<td>.98</td>
</tr>
<tr>
<td>Trial 5</td>
<td>.97</td>
</tr>
<tr>
<td>Trial 10</td>
<td>.97</td>
</tr>
<tr>
<td>Trial 11</td>
<td>.96</td>
</tr>
<tr>
<td>Trial 12</td>
<td>.95</td>
</tr>
<tr>
<td>Trial 4</td>
<td>.94</td>
</tr>
<tr>
<td>Trial 3</td>
<td>.90</td>
</tr>
<tr>
<td>Trial 2</td>
<td>.83</td>
</tr>
<tr>
<td>Trial 1</td>
<td>.75</td>
</tr>
</tbody>
</table>

As shown in Table 7.24, Tailorshop performance on trials 1 to 12 loaded on a single component, which accounted for 87.60% of the variance in performance on the computer simulated task. By applying factor analytic techniques to computer simulation performance scores, information regarding factor loadings is captured in the factor score, which is not captured by simply averaging performance scores across all trials, e.g. mean performance. Thus, the component score derived from this analysis could be used as an aggregated score of Tailorshop performance, which may be a more accurate reflection of true performance than any of the scoring techniques suggested so far, e.g. mean performance across all trials, goal achievement, and final trial scores. Table 7.25 presents the correlation matrix of cognitive abilities factor scores and the Tailorshop component score.
Table 7.25

Correlation Matrix of Cognitive Abilities Factors and the Tailorshop Component Score

<table>
<thead>
<tr>
<th></th>
<th>Gf</th>
<th>Gc</th>
<th>Gv</th>
<th>Tailorshop Factor Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluid Reasoning (Gf)</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Crystallised Knowledge (Gc)</td>
<td>.45</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad Visualisation (Gv)</td>
<td>.34</td>
<td>.20</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Tailorshop Factor Score</td>
<td>.20</td>
<td>.39</td>
<td>.03</td>
<td>*</td>
</tr>
</tbody>
</table>

As shown in Table 7.25, employing a more accurate measure of overall Tailorshop performance, based on a factor score, significant positive correlations between Tailorshop performance and specific components of intelligence have emerged. The correlation between Tailorshop performance and fluid reasoning (Gf) is significant, \( r = .20, p < .05 \). Similarly, the correlation between Tailorshop performance and crystallised intelligence (Gc), is significant, \( r = .39, p < .01 \). These correlations are compatible with those obtained using alternative scoring methods, presented in Table 7.9.

7.8.3 The Relation of Forestry System Factors with Intelligence Factors

Factor Analysis of performance scores on the Forestry System computer simulation will be described next. Recall from Section 6.2.5.5, which described the Forestry system task, that there are 50 simulated months. Total asset values were calculated at the end of the each of the 50 months (50 trials). Performance scores on all 50 trials were entered into a Principle Components (PC) analysis with oblique rotation. The scree plot, presented in Figure 7.2, indicated the extraction of five factors.
Factors with eigenvalues > 1 should be retained; the remaining factors are ‘scree’ and should not be extracted (Norman, 1987). Consistent with the scree plot in Figure 7.2, Table 7.26 shows the analysis extracted five factors with eigenvalues greater than 1, which together accounted for 94.81% of the variance in the Forestry System performance scores. Table 7.26 presents the pattern matrix, proportion of variance accounted for, and communalities (h²) for the Forestry System performance scores across all 50 trials. All factor loadings greater than .30 are underlined to facilitate interpretation.
Table 7.26
Factor Loadings, Communalities ($h^2$), Percents of Variance for Principal Component Analysis and Oblique Rotation on Forestry System Performance

<table>
<thead>
<tr>
<th>Forestry System Trials</th>
<th>Factor</th>
<th>h$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials 29 to 50</td>
<td>Trials 16 to 28</td>
<td>Trials 8 to 15</td>
</tr>
<tr>
<td>Trial 44</td>
<td>.99</td>
<td>-.10</td>
</tr>
<tr>
<td>Trial 43</td>
<td>.99</td>
<td>-.10</td>
</tr>
<tr>
<td>Trial 45</td>
<td>.99</td>
<td>-.10</td>
</tr>
<tr>
<td>Trial 46</td>
<td>.99</td>
<td>-.10</td>
</tr>
<tr>
<td>Trial 42</td>
<td>.99</td>
<td>-.08</td>
</tr>
<tr>
<td>Trial 41</td>
<td>.99</td>
<td>-.07</td>
</tr>
<tr>
<td>Trial 47</td>
<td>.99</td>
<td>-.08</td>
</tr>
<tr>
<td>Trial 40</td>
<td>.99</td>
<td>-.04</td>
</tr>
<tr>
<td>Trial 48</td>
<td>.99</td>
<td>-.07</td>
</tr>
<tr>
<td>Trial 39</td>
<td>.99</td>
<td>-.02</td>
</tr>
<tr>
<td>Trial 49</td>
<td>.99</td>
<td>-.06</td>
</tr>
<tr>
<td>Trial 50</td>
<td>.99</td>
<td>-.07</td>
</tr>
<tr>
<td>Trial 38</td>
<td>.99</td>
<td>.00</td>
</tr>
<tr>
<td>Trial 37</td>
<td>.97</td>
<td>.03</td>
</tr>
<tr>
<td>Trial 36</td>
<td>.95</td>
<td>.06</td>
</tr>
<tr>
<td>Trial 35</td>
<td>.91</td>
<td>.10</td>
</tr>
<tr>
<td>Trial 34</td>
<td>.89</td>
<td>.14</td>
</tr>
<tr>
<td>Trial 33</td>
<td>.85</td>
<td>.19</td>
</tr>
<tr>
<td>Trial 32</td>
<td>.80</td>
<td>.26</td>
</tr>
<tr>
<td>Trial 31</td>
<td>.74</td>
<td>.35</td>
</tr>
<tr>
<td>Trial 30</td>
<td>.68</td>
<td>.42</td>
</tr>
<tr>
<td>Trial 29</td>
<td>.58</td>
<td>.53</td>
</tr>
<tr>
<td>Trial 22</td>
<td>.07</td>
<td>.93</td>
</tr>
<tr>
<td>Trial 23</td>
<td>.14</td>
<td>.91</td>
</tr>
<tr>
<td>Trial 21</td>
<td>.03</td>
<td>.91</td>
</tr>
<tr>
<td>Trial 24</td>
<td>.19</td>
<td>.88</td>
</tr>
<tr>
<td>Trial 20</td>
<td>.00</td>
<td>.88</td>
</tr>
<tr>
<td>Trial 25</td>
<td>.25</td>
<td>.84</td>
</tr>
<tr>
<td>Trial 19</td>
<td>-.02</td>
<td>.84</td>
</tr>
<tr>
<td>Trial 18</td>
<td>-.07</td>
<td>.80</td>
</tr>
<tr>
<td>Trial 26</td>
<td>.34</td>
<td>.76</td>
</tr>
<tr>
<td>Trial 17</td>
<td>-.10</td>
<td>.74</td>
</tr>
<tr>
<td>Trial 16</td>
<td>-.11</td>
<td>.68</td>
</tr>
<tr>
<td>Trial 27</td>
<td>.44</td>
<td>.68</td>
</tr>
<tr>
<td>Trial 28</td>
<td>.51</td>
<td>.60</td>
</tr>
</tbody>
</table>
Table 7.26 continued below

Forestry System Trials

<table>
<thead>
<tr>
<th>Trials 29 to 50</th>
<th>Trials 16 to 28</th>
<th>Trials 8 to 15</th>
<th>Trials 3 to 7</th>
<th>Trials 1 to 2</th>
<th>$h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 10</td>
<td>.08</td>
<td>-.02</td>
<td>.96</td>
<td>-.02</td>
<td>-.02</td>
</tr>
<tr>
<td>Trial 9</td>
<td>.10</td>
<td>-.11</td>
<td>.95</td>
<td>.13</td>
<td>.03</td>
</tr>
<tr>
<td>Trial 11</td>
<td>.07</td>
<td>.06</td>
<td>.92</td>
<td>-.10</td>
<td>-.07</td>
</tr>
<tr>
<td>Trial 8</td>
<td>.10</td>
<td>-.14</td>
<td>.89</td>
<td>.35</td>
<td>.06</td>
</tr>
<tr>
<td>Trial 12</td>
<td>.02</td>
<td>.18</td>
<td>.86</td>
<td>-.12</td>
<td>-.07</td>
</tr>
<tr>
<td>Trial 13</td>
<td>-.01</td>
<td>.30</td>
<td>.79</td>
<td>-.12</td>
<td>-.07</td>
</tr>
<tr>
<td>Trial 14</td>
<td>-.06</td>
<td>.43</td>
<td>.70</td>
<td>-.11</td>
<td>-.07</td>
</tr>
<tr>
<td>Trial 15</td>
<td>-.10</td>
<td>.56</td>
<td>.60</td>
<td>-.09</td>
<td>-.07</td>
</tr>
<tr>
<td>Trial 5</td>
<td>-.04</td>
<td>.05</td>
<td>-.03</td>
<td>.98</td>
<td>-.01</td>
</tr>
<tr>
<td>Trial 6</td>
<td>-.07</td>
<td>.03</td>
<td>.24</td>
<td>.94</td>
<td>.18</td>
</tr>
<tr>
<td>Trial 4</td>
<td>.00</td>
<td>.09</td>
<td>-.23</td>
<td>.90</td>
<td>-.16</td>
</tr>
<tr>
<td>Trial 3</td>
<td>-.02</td>
<td>.18</td>
<td>-.28</td>
<td>.71</td>
<td>-.40</td>
</tr>
<tr>
<td>Trial 7</td>
<td>.01</td>
<td>-.10</td>
<td>.64</td>
<td>.69</td>
<td>.16</td>
</tr>
<tr>
<td>Trial 2</td>
<td>-.03</td>
<td>.05</td>
<td>-.10</td>
<td>-.02</td>
<td>.99</td>
</tr>
<tr>
<td>Trial 1</td>
<td>-.03</td>
<td>.04</td>
<td>-.09</td>
<td>-.04</td>
<td>.99</td>
</tr>
</tbody>
</table>

Eigenvalue 31.72  7.24  4.09  2.66  1.70

Percent of Variance 63.44  14.48  8.19  5.31  3.39

Factor Correlation Matrix

<table>
<thead>
<tr>
<th>Trials 29 to 50</th>
<th>*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trials 16 to 28</td>
<td>.66</td>
</tr>
<tr>
<td>Trials 8 to 15</td>
<td>.30</td>
</tr>
<tr>
<td>Trials 3 to 7</td>
<td>.07</td>
</tr>
<tr>
<td>Trials 1 to 2</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note. Loadings > .60 are underlined.

Factor loadings > .40 are adequate for the sample size, N > 200 (Stevens, 1992).

Oblique rotation was employed to simplify the pattern of factor loadings, this method of rotation assumes intercorrelations amongst all of the cognitive abilities variables, which was the case. As shown in the pattern matrix, displayed in Table 7.26, the first factor
was defined as Forestry System trials 29 to 50, with the final trials of the simulation loading on this factor. Factor 1 accounts for 63.44% of the variance in Forestry System performance. The second factor was defined as Forestry System trials 16 to 28, with the middle trials of the simulation loading on this factor. Factor 2 accounts for 14.48% of the variance in Forestry System performance. Factor 3 was defined as Forestry System trials 8 to 15, with early trials of the simulation loading on this factor. Factor 3 accounts for 8.19% of the variance in Forestry System performance. Factor 4 was defined as Forestry System trials 3 to 7, with the initial trials of the simulation loading on this factor. Factor 4 accounts for 5.31% of the variance in Forestry System performance. Factor 5 is was defined as trials 1 to 2 and accounts for 3.39% of the variance in Forestry System performance.

The Factor Correlation Matrix, displayed in Table 7.26 above, shows that Factors 1, 2, and 3 are significantly intercorrelated. The correlation between Factor 1 and Factor 2 is significant, \( r = .66, p < .05 \). The correlation between Factor 1 and Factor 3 is significant, \( r = .30, p < .05 \). The correlation between Factor 2 and Factor 3 is also significant, \( r = .52, p < .05 \). However, Factor 4 does not correlate with Factors 1, 2, 3, and 5. The correlations between Factor 4 and Factors 1, 2, 3, and 5 are not significant, \( r = .07, p > .05 \), \( r = -.13, p > .05 \), \( r = .00, p > .05 \), and \( r = -.11, p > .05 \), respectively. In addition, Factor 5 does not correlate with factors 1, 2, and 4. The correlations between Factor 5 and Factors 1, 2, and 4 are not significant, \( r = .05, p > .05 \), \( r = .02, p > .05 \), \( p > .05 \), and \( r = -.11, p > .05 \), respectively. Finally, Factor 3 and Factor 5 are correlated, \( r = .21 \).
The intercorrelations between Forestry System factors suggest that performance scores on trials 3 to 7 are different from performance scores on previous trials (1 to 2) and performance on later trials (8 to 50) of the simulation. These five factors will now be interpreted by examination of the characteristics of the Forestry System task (described in Section 6.2.5.5), by inspection of the correlations between Forestry System factors and cognitive abilities factors, and by inspection of the correlations between individual Forestry System trial scores and cognitive abilities factors.

Table 7.27

Correlation Matrix of Forestry System Factor Scores and Cognitive Abilities

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gf</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Gc</td>
<td>.45</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Gv</td>
<td>.34</td>
<td>.20</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Forestry System (trials 29 to 50) Factor 1</td>
<td>.29</td>
<td>.28</td>
<td>.20</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Forestry System (trials 16 to 28) Factor 2</td>
<td>.28</td>
<td>.26</td>
<td>.14</td>
<td>.66</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Forestry System (trials 8 to 15) Factor 3</td>
<td>.23</td>
<td>.18</td>
<td>.09</td>
<td>.30</td>
<td>.52</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Forestry System (trials 3 to 7) Factor 4</td>
<td>-.02</td>
<td>-.12</td>
<td>-.03</td>
<td>.07</td>
<td>-.13</td>
<td>.00</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>8. Forestry System (trial 1)</td>
<td>.41</td>
<td>.27</td>
<td>.04</td>
<td>.02</td>
<td>.00</td>
<td>.13</td>
<td>-.16</td>
<td>*</td>
</tr>
</tbody>
</table>

It was beyond practicality (and mostly redundant to the information presented through factor scores) to present correlations between intelligence factors and all 50 Forestry System performance trials. However, pockets of correlations observed between intelligence factors and the 50 trials, was consistent with the factor structure of the
Forestry System task for trials 3 to 50. The only information that was lost through the emergence of factor 5 (trials 1 to 2) was the fact that trial 1 is significantly correlated with intelligence while trial 2 is not significantly correlated with intelligence. The correlation between Factor 5 and intelligence does not reveal this information, which is why the trial 1 score was added to Table 7.27.

The correlations in Table 7.27 show that Forestry system performance in trial 1 is significantly correlated with Gf and Gc, $r = .41, p < .05$, $r = .27, p < .05$, respectively. This correlation is consistent with the task characteristics, described in Section 6.2.5.5, which indicates that the participant managers of the Forestry business commence the task with some trees already growing. During the first simulated month (trial 1), participant managers can immediately assign an ‘action’ to their workers. ‘Actions’ may include cutting down the pre existing trees, fertilizing empty plots in preparation for planting, etc. These ‘actions’ would automatically reduce the total assets of the company due to workers income costs, fertilizer costs, etc. However, this type of task management, within the first trial of the task, would not be logical because the participant manager has not explored the task, and has not accessed important information (e.g., the information text). A more logical approach in the initial simulated month would be to read the information text and learn about the system before implementing any ‘actions’. Thus, participants with greater Gf and Gc abilities did not take any ‘action’ during the first trial of the task, as evidenced by the fact that their total assets remained steady. In contrast, participants with lower Gf and Gc abilities assigned an ‘action’ to their workers during the initial trial of the task and consequently their total assets were reduced. Overall, results indicate that more intelligent participants tend to learn about the system before making their decisions, while their less intelligent
counterparts make decisions before exploring the task and learning the available system information.

In addition, the effects of actions, such as planting trees, also take several months to achieve profit, whilst the trees are growing. During months 3 to 7 (factor 4), the participant managers spent the majority of their time engaged in exploring the task. Therefore, factor 4 would reflect exploration strategies and learning rather than complex problem solving ability. Recall from Figure 4.1, that rule identification is assessed by exploration strategy, while rule application is assessed by simulation performance. Forestry System performance during trials 3 to 7 is assumed to reflect learning, it is hypothesised that a measure of exploration strategy during these trials, may correlate with the personality factor of Openness. Unfortunately, there is no trial-by-trial exploration strategy measure, built into the Forestry System scoring protocol, in order to test this hypothesis.

Correlations presented in Table 7.27 also show significant correlations between Forestry System Factors 1, 2, and 3, which represent trials 8 to 50, and intelligence factors. The correlation between Factor 1 (trials 29 to 50) and Gf and Gc is significant, $r = .29$, $p < .05$, $r = .28$, $p < .05$, respectively. The correlation between Factor 2 (trials 16 to 28) and Gf and Gc is significant, $r = .28$, $p < .05$, $r = .26$, $p < .05$, respectively. The correlation between Factor 3 (trial 8 to 15) and Gf is significant, $r = .23$, $p < .05$, however, the correlation between Factor 3 and Gc is not significant, $r = .18$, $p > .05$. In contrast, the correlation between Factor 4 (trials 3 to 7) and Gf and Gc is not significant, $r = -.02$ $p > .05$, $r = -.12$, $p > .05$, respectively. There are no significant correlations between Forestry System performance, during any block of trials, and broad visualisation (Gv).
Overall, the results in Table 7.27 show that Forestry System Factors 1, 2, and 3, which represent trials 8 to 50, are generally significantly positively correlated with intelligence factors (Gf and Gc). In contrast, Factor 4 (trials 3 to 7) is not significantly correlated with intelligence factors. These correlations can be explained by ‘rule application’ and ‘rule identification’ components of the theoretical model of complex problem solving performance (presented in Figure 4.1).

Factors 1, 2, 3 (trials 8 to 50) and Factor 5 (trials 1 to 2), of the Forestry System, reflect the rule application component of complex problem solving ability, identified in Figure 4.1. This is consistent with the observed positive significant correlations, shown in Table 7.27, between these Forestry System factors and cognitive abilities. In contrast, Forestry System Factor 4 (trials 3 to 7) is assumed to reflect ‘rule identification’ and does not correlate with cognitive abilities.

Factor 4, which represents trials 3 to 7 reflect ‘rule identification’, and can be explained by the fact that participants were engaged in exploration of the underlying rules of the system during this phase. The Forestry System factor scores were calculated based on total assets at the end of each of the 50 trials of the simulation. These 50 performance scores reflect ‘rule application’ rather than ‘rule identification’. Thus results are consistent with the fact that during trials 3 to 7, participants were engaged in ‘rule identification’ rather than ‘rule application’. Unfortunately, a measure of ‘rule identification’ for trials 3 to 7 was not employed in the design of the complex problem solving task. If a measure of ‘rule identification’ were available, it would reflect learning and may correlate with personality traits such as Openness.
In summary, the correlations in Table 7.27, show that trials 8 to 50 of the Forestry System correlate with cognitive abilities, while trials 3 to 7 of the Forestry System do not correlate with cognitive abilities. Thus, for highly complex computer simulations, such as the Forestry System, (in contrast with the moderately complex Furniture Factory and Tailorshop tasks), aggregated scoring techniques, which simply take the mean performance across all trials, are not appropriate. Factor analysis revealed that performance during trials 3 to 7 was not reflected in ‘rule application’ scores and consequently did not correlate with cognitive abilities. It was suggested that performance during trials 3 to 7 reflects ‘rule identification’ and could be assessed by individual differences in exploration strategy. In contrast, performance during trials 1 to 2, and 8 to 50 reflect ‘rule application’ and can be assessed by performance scores on the appropriate trials of the computer simulation. Overall, results indicate that aggregated scoring techniques, to assess complex problem solving performance, must be refined to take into account different phases throughout simulation performance with a distinction between rule identification (learning) and rule application (complex problem solving performance), as described by Figure 4.1.

A summary of the results presented in this chapter, interpretation of findings, and suggestions for future research will now be presented in Chapter 8.
8 Conclusions

The results of Study 2 have led to several key findings regarding the relationship of complex problem solving with intelligence, personality, interests, and biodata. Interpretation and conclusions of the results analysed in Chapter 7 will now be provided: Firstly, scoring issues in complex problem solving research will be discussed. Secondly, key findings regarding the relationship of complex problem solving and intelligence will be provided. Thirdly, conclusions will be drawn from the observed relationship between complex problem solving and non-cognitive variables (personality and vocational interests). Finally, future directions in the area of complex problem solving research will be proposed.

8.1 Scoring issues

A relationship between intelligence and complex problem solving was found using three scoring techniques (aggregated, final trial, and goal achievement) to measure computer simulation performance. In accordance with Brunswik symmetry, in Study 2 the expectation that correlations between intelligence and complex problem solving would be increased by the use of specific cognitive abilities tests and aggregated computer
simulation scores, rather than general or factor scores of intelligence and final trial computer performance simulation scores, was supported.

8.1.1 Goal Achievement Scores

Goal achievement scores were developed and calculated for each of the three simulations. Goal achievement scores indicate the average percentage of trials on which participants achieved the goal of the task. Goal achievement scores also offer a way to measure the gap (barrier) between the problem state of the task and the goal state the participant is trying to achieve. The hypothesis that goal achievement will be highest for computer simulations characterised by smaller gaps e.g. Furniture Factory and conversely, goal achievement will be lowest for simulations in which the gap is greater e.g. Tailorshop, was supported. It was found that participants achieved their set goal more often on the Furniture Factory task than the Forestry System task, and more often on the Forestry System task than the Tailorshop task. Study 2 represents the first time that a measure for cross-task comparison has been developed in computer simulation research. In addition, goal achievement scores have significant positive correlations with intelligence components, over and above traditional scoring methods such as aggregated or final trial scores.

8.1.2 Factor Analysis of Complex Problem Solving Performance

Five factors representing different ‘phases’ of complex problem solving performance emerged from factor of analysis of individual trial scores on the Forestry System. The expectation that trials in which participants engage in ‘rule application’ would correlate
with cognitive abilities was supported. In addition, the hypothesis that trials in which participants do not engage in ‘rule application’ would not correlate with cognitive abilities was supported. Trial scores on the Furniture Factory and the Tailorshop loaded on a single factor, as these computer simulations were characterised by fewer trials and participants engaged in ‘rule identification’ and ‘rule application’ behaviour in each and every trial throughout performance. In contrast to the Forestry System, participants could not proceed to the next trial without making an ‘action’ assessed by ‘rule application’ in each and every trial. Thus, results factor analyses of computer simulation performance scores are consistent with task properties and may offer a more adequate method of aggregated scoring than simply taking the mean score all trials.

Factor analysis of complex problem solving performance, across all trials of computer simulated tasks, revealed that more complex tasks, such as the Forestry System, are characterised by several factors or ‘phases of task performance’. For example, there is a positive and significant correlation between cognitive abilities and performance on the first trial of the Forestry System, as measured by total assets in the computer simulation. This correlation suggests that more intelligent participants do not make uninformed decisions during the initial trial of the task, they do not take any action during the first trial and consequently total assets remain stable. In contrast, less intelligence participants make decisions before they have explored the information text that is available in the computer task. In addition, all phases of Forestry System performance were significantly positively correlated with intelligence, with the exception of an early phase (trials 3 to 7, out of a total of 50 trials) of the task. This factor derived from these trials, did not correlate with intelligence.
The lack of correlation is consistent with the fact that participants were using different strategies during this period. Some were continuing to learn from the information text, while others were taking a small number of actions, and others were implementing many decisions. In other words, individual differences in strategy during this phase did not correspond with optimal complex problem solving performance in a one to one ratio. Different approaches may have had similar chances of task success. Thus, performance during this phase did not correlate with intelligence. Performance measures of simulations, such as ‘total assets’ take into account the rule application component of the theoretical model presented in Figure 4.1, but do not take into account rule identification, which can be assessed exploration strategy. Recall, from Study 1 that Furniture Factory strategy had an almost perfect correlation with Furniture Factory performance. However, this was not the case for the Forestry System task. Therefore, scoring techniques such as aggregating performance (rule application) across all trials of a simulation, may not be appropriate for simulations, such as the Forestry System, where rule application during a certain phase of the game (e.g., trials 3 to 7) may not reflect complex problem solving performance.

Overall, results from the factor analysis of computer simulations, suggest that the relationship between intelligence and complex problem solving performance varies across all trials of the task. There appears to be a development of performance over time. In addition, the theoretical explanation for the relationship will change throughout different phases of the task. These results highlight the need to analyse data at the trial level. Thus, factor analysis of computer simulation performance could be a more appropriate technique to investigate the relationship between intelligence and complex
problem solving, rather than final trial scores or aggregated scores that have been used in previous research.

8.1.3 **Shared Variance in Complex Problem Solving Performance across Three Computer Simulated Tasks**

Despite the problems associated with computer simulation scoring techniques, results of Study 2 suggest that the three complex problem solving tasks were measuring common abilities, as indicated by positive significant correlations between the 3 complex problem solving tasks. The hypothesis that all three computer simulations would share common variance was supported, indicating that common processes, such as complex problem solving ability, underlie computer simulation performance. However, each simulation has more in common with specific cognitive abilities than with each other.

8.1.4 **Loopholes**

The presence of ‘loopholes’ in computer simulated tasks was identified in the Furniture Factory task employed in Study 1. There is little discussion of ‘loopholes’ in the literature, possibly due to the fact that most research of computer simulations is conducted by the test developers, who may regard loopholes in their mathematical models (causal connections between the variables), as error variance. The only research uncovered in the literature review for this thesis, that acknowledged the presence of loopholes, was conducted with the Space Fortress task. The Space Fortress research was a group project conducted by several international researchers. The presence of loopholes, which may be exploited by participants who achieve the goal without
following the directions of the task, has also been touched on (though not explicitly) by the developers of the Forestry System task (Wagener & Conrad, 1996). They included a scoring procedure, SKAPCOR, which identifies participants who have exploited loopholes. SKAPCOR scores revealed that 10 participants (3.38% of the sample) exploited loopholes on the Forestry System and achieved high performance scores (total assets) by means other than which the task was designed. Due to the fact that the causal connections between variables underlying all computer simulated tasks are described by mathematical algorithms, it is suggested that all computer simulations would feature loopholes. However, no detail is known regarding the presence of loopholes in the Furniture Factory or Tailorshop tasks. It is important to distinguish those participants whose scores reflect complex problem solving behaviour from those whose do not. The inclusion of participants’ scores, who exploit loopholes by not engaging in complex problem solving in order to achieve high performance scores, may corrupt the true correlations between computer simulation performance and intelligence. This may offer an additional explanation for the lack of a relationship between intelligence and complex problem solving in prior research.

8.2 Relationship between Intelligence and Complex Problem Solving

8.2.1 Relationship between Intelligence and Complex Problem Solving across Three Computer Simulations

It is unclear whether individual differences in intelligence cause individual differences in complex problem solving or whether the relationship is the other way around
Where possible, analyses of the results of Study 2 take an exploratory approach due to the lack of an established direction of causality. However, an attempt was made to explore the results by employing regression techniques, in which complex problem solving was defined as the dependent variable and intelligence factors were employed as predictor variables. Overall, fluid reasoning (Gf) and crystallised knowledge (Gc) explained roughly 20% of the variance in all three computer simulated tasks. However, these results must be interpreted with caution due to the assumed direction of causality.

Very little research has explored the correlation between specific cognitive abilities tests, as opposed to the factors on which they load, or more often general scores of intelligence. Thus, the relationship between sixteen specific cognitive abilities tests and complex problem solving performance on three computer simulations was also explored. It was found that the correlation between specific cognitive abilities tests, such as Esoteric Analogies and Critical Reasoning was significant and was a stronger relationship that that observed using factor scores (Gf, Gc) alone. The strong correlations observed between computer simulation performance with Esoteric Analogies and Critical Reasoning tests, was observed for all 3 computer simulations. These results are compatible with previous findings (Hussy, 1989; Kröner, Plass, & Leutner, 2005; Wagener, 2001) of a relationship between computer simulation performance and the BIS-K factor (processing capacity, capturing the ability to recognise relations and rules and to form logical inferences in figure series, number series, verbal analogies).
Overall, it is suggested that computer simulation performance depends on reasoning skills. In summary, consistent pattern of results was found across all three computer simulated tasks. Complex problem solving performance was significantly positively correlated with Esoteric Analogies and Critical Reasoning. These results may offer another possible explanation for the failure of the majority of previous research to establish a relationship between complex problem solving and intelligence. Individual differences research is characterised by factor analytic approaches that strive to obtain ‘pure’ measures of tasks. Thus, tasks that assess multiple cognitive abilities concurrently, e.g. analogies, have rarely been employed in this area of research. A notable exception is research conducted by Süß and colleagues (Süß, Kersting, & Oberauer, 1993; Süß, Oberauer, & Wittmann, 2005; Wittmann & Süß, 1999), who also employed an analogies task, and consistently found correlations between analogies performance and complex problem solving performance. The advantages of using highly specific cognitive abilities tests to explore the relationship between intelligence and complex problem solving outweigh the disadvantages due to loss of reliability of the more specific measures (in contrast to more reliable broad intelligence factors). The high specificity obtained on both sides of the Brusnwik lens by employing scores on individual trials of computer simulations on one side and highly specific cognitive abilities measures on the other, allows for a more detailed understanding of the complex problem solving construct. For example, if only final trial scores in computer simulations were used, the changing relationship between intelligence and complex problem solving over time could not be identified. Similarly, if only IQ scores were used, the consistent relationship of complex problem solving with reasoning skills would be concealed.
8.2.2 Predictors of Complex Problem Solving Performance

The prediction that there would be a significant correlation between the static variables (UAI and Gf) was supported. The hypothesis that the correlation between the static UAI scores and dynamic complex problem solving performance would not be significant was also supported. It was concluded that a static criterion, such as UAI is adequate for a static task, such as Raven’s Progressive Matrices (the most common Gf measure). However, finding a dynamic criterion that is related to dynamic tasks, such as computer simulations, is a major problem in this area of research (J. Beckmann, personal communication, September 20, 2005).

8.3 Relationship of Complex Problem Solving and Non-Cognitive Variables

8.3.1 Personality

The relationship between the five factors of personality (Costa & McCrae, 1992) and complex problem solving performance in three computer simulations was explored. Significant negative correlations were observed between Neuroticism and performance on the Furniture Factory and Tailorshop tasks. These findings are consistent with the moderate negative correlations observed between Neuroticism and Forestry System performance (Wagener, 2001). Interestingly, no such correlation was observed for the Forestry System in Study 2 of this thesis. It was also surprising that there were no correlations between Openness and complex problem solving because Openness is the personality factor that most often correlates with intelligence. The personality variable,
neuroticism, is a measure of emotional instability. Thus, the negative correlation
was interpreted as suggesting that participants with increased emotional stability are
more likely to achieve high performance on complex problem solving tasks. These
results are consistent with the recent suggestion that emotion is an important mediator
of complex problem solving performance (Spering, Wagener, & Funke, 2005).

8.3.2 Interests

The hypothesis that interests that correlate with intelligence (realistic, investigative, and
artistic) would correlate with complex problem solving performance (Furniture Factory,
Tailorshop, and Forestry System performance) was partially supported. Significant
positive correlations were observed between career interests (Mathematics, Science, and
Medicine) and complex problem solving (assessed by Tailorshop and Forestry System
tasks). These correlations are consistent with the cover stories of the tasks and task
demands. For example, the strongest correlation between Tailorshop performance and
career interests was for the domain of Mathematics. This is consistent with previous
findings of a relationship between Tailorshop performance and economics (Süß,
Oberauer, & Wittmann, 2005). Overall, the relationship between complex problem
solving and interests tends to emerge for Holland’s (1973) realistic and investigative
vocational interests. Results are consistent with the suggestion that realistic and
investigative interests are related to intelligence (Ackerman & Heggestad, 1997).
Interestingly, the findings that computer simulated task performance correlates with
intelligence as process variables (realistic and investigative). Realistic and investigative
vocational interests correlate with intelligence as process factors (reasoning, math),
intelligence as process also correlates with neuroticism (Ackerman & Heggestad).
Thus, the findings that complex problem solving behaviour correlates with both process factors (realistic and investigative interests) and the personality factor of neuroticism are compatible with PPIK theory.

8.4 Future Directions

Findings in Study 1 and 2 of this thesis were planned to be extended from a sample predominantly comprising university students to a professional sample (i.e., the Police Service). It is predicted that a dynamic criterion such as assessment centre task performance would be related to performance on computer simulated tasks as both tasks require complex problem solving processes. In addition, the role of emotion in mediating complex problem solving performance was to be investigated. Ethical approval from the NSW Police Service was obtained. However, ethical delays at the research end, which have now been resolved, meant that the study had to be abandoned for the purposes of this thesis. Ethics approval has been obtained for the conduction of future research in this area.

Current empirical evidence suggests that the relationship between complex problem solving and intelligence may be mediated by interaction effects between the person, the task, and the environment (Wenke, Frensch, & Funke, 2005). Theory driven interaction effects, such as the interaction between fluid reasoning (Gf) and self-efficacy also deserve future research attention. Previous research suggested that ability-motivation interactions were important during complex skill acquisition (Kanfer & Ackerman, 1989). However, the relationship of complex problem solving with the interaction of Gf and self-efficacy was explored for the Furniture Factory task and no significant
relationship was observed. The Gf/self-efficacy interaction did not account for a significant proportion of the variance after accounting for Gf and self-efficacy (Wood & Ryan, unpublished). Notwithstanding, interaction effects are a promising area of future research into complex problem solving processes.

There is a need to bring together different literatures to account for the research findings. In addition to Brunswick symmetry, other relevant areas for future research include the relationship between intelligence and work performance. Also, relevant is research on skill acquisition and the changing pattern of ability requirements as one learns a complicated computer simulation (e.g., Ackerman’s TRACON task).

Limitations of present research include the use of fairly homogenous samples, which make effects, such as age difficult to investigate. In addition, there is a lack of longitudinal research designs that would assess the causal direction of the relationship between intelligence and complex problem solving.

Individual differences research in the assessment of complex problem solving by computer simulated tasks has led to inconsistent results that are difficult to interpret. Researchers must engage in a cumulative effort to understand the characteristics of computer simulations as research tools. This thesis has attempted to bring individual differences research in the area a step closer to obtaining stable results from which generalisations about complex problem solving tasks can be made.
REFERENCES


Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*(2), 219-245.


AltaVista. (2004). Babel Fish. from babelfish.altavista.com


Games, E. A. (2002-2003). The Sims Online™: EA GAMES, the EA GAMES logo, Maxis, the Maxis logo and Electronic Arts are trademarks or registered trademarks of Electronic Arts Inc. in the U.S. and/or other countries. All rights reserved. EA GAMES™ is an Electronic Arts™ brand.


Microsoft. (1986-2004). ToCA (TM) is a registered trademark of TOCA Limited used under license.: Microsoft, Xbox, Xbox Live, the Live logo, and the Xbox logos are registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.


Appendix A

A  Computer simulated scenarios employed in individual differences research

AIR TRAFFIC CONTROLLER (ATC: Kanfer & Ackerman, 1989) is a simulation of some of the activities performed by air traffic controllers. The participant’s goal is to land planes safely and efficiently. The task variables consist of four runways, 12 hold pattern positions, and a queue stack indicating planes requesting permission to enter the hold pattern. TRACON (Terminal Radar Approach Controller: Wessen, 1991) is a real time problem solving simulation of the tasks performed by air traffic controllers and is similar to the ATC task (Kanfer & Ackerman). TRACON is widely used in North America by the Federal Aviation Administration (FAA), the National Aeronautics and Space Administration, the Department of Defence, and many training schools for air traffic controllers (Ackerman & Cianciolo, 2000). Significant correlations were found between general intelligence and performance in the TRACON task. However, results suggest that specific ability components (e.g., spatial, perceptual speed, perceptual/psychomotor abilities, etc.) and specific task components are more beneficial in explaining the ability-skill relationship, than global IQ or complex task scores (Ackerman, 1992).

CABOT (Computer-Aided Bridge Operation Task: Sauer, Wastell, Hockey, Crawshaw, & Downing, 2003), is a simulation of an advanced ship’s bridge environment of a commercial vessel. The participant plays the role of the watch keeper. There are three variables, navigation, engine control, and cargo control. Navigation is defined as the
primary goal, while engine and cargo control are secondary goals. When level of difficulty was manipulated by increasing or decreasing the number of potential collisions, performance on the secondary tasks was reduced, while performance on the main goal of navigation was generally maintained. In addition, under highly difficult conditions, participants changed their exploration strategies in terms of use of decision support systems. This study highlights the fact that overall performance scores may not accurately describe changes in performance. Other aspects of performance such as strategy use must also be investigated. CABOT is polytelic, with multiple goals, and the method of scoring is important for analysis of results. For example, in the study just described, navigation (the main goal) was not affected by increased collisions, whereas engine and cargo control (secondary goals) were affected by increased collisions. Polytelic simulations allow multiple scoring techniques to calculate performance and have a major impact on experimental results and conclusions drawn. Scoring issues will be addressed in Chapter 5 of this thesis.

COLDSTORE (Reichert & Dörner, 1988) consists of six variables. The participant plays the role of the manager of a supermarket. The automatic steering of the cold storage depot breaks down and the only way to control the temperature to protect the goods is by means of a control wheel and a thermometer. The scales of the control wheel are different to the scales on the cold store temperature gauge and the participant must infer the exact relations. Correlations between COLDSTORE and Raven’s Advanced Progressive Matrices, a measure of fluid intelligence, were around .3 (Rigas, Carling, & Brehmer, 2002).
DORI (Hesse, 1982) consists of 12 variables and is the offspring of DAGU. The complex system simulates the living conditions of a nomad tribe in the Sahel region, whose prosperity depends on cattle raising. There is an analogous abstract condition in which the variables are replaced by Latin Letters, termed the semantic-free condition. In addition, a graphical display of the connections between the variables was presented in a transparent condition and was not provided in the intransparent (opaque) condition. Hesse (1982) obtained moderate correlations between problem-solving performance and Raven’s Advanced Progressive Matrices (APM: Raven, Raven, & Court, 1993) in the semantic-free condition where variables were replaced by Latin letters. In the semantic-free condition, when the underlying rules of the game were made transparent, the correlation between problem solving competency and intelligence was increased, for the transparent condition $r = .38$ and for the intransparent condition $r = .46$.

FSYS 2.0 (Wagener & Conrad, 1996) was translated from English to German for this thesis. The Forestry task (Wagener, 2001) consists of five forests that must be tended to by the participant manger. The main goal is to increase the total assets of the company. The participant manager must allocate one of four tasks (plant trees, cut down trees, pest control, fertilize) to their workers to carry out. This computer simulation has been used in German studies exploring the relationship between complex problem solving competencies and intelligence (Wagener, 2001). These findings will be described in detail in Chapter 4. FSYS 2.0 has also been employed recently in a study investigating the role of emotions in complex problem solving (Spering, Wagener, & Funke, 2005).
Results suggest that negative emotions are not detrimental to complex problem solving performance. However, participants in the negative emotions condition are more likely to use information seeking strategies than participants in the positive emotions condition.

FSYS 2.0 was employed to assess complex problem solving competencies in this thesis. The relationship between FSYS 2.0 performance and specific components of intelligence is discussed in detail in Chapter 4.

**FURNITURE FACTORY** (Wood & Bailey, 1985) consists of 25 variables. The FURNITURE FACTORY is designed for employment in cognitive experiments where the experimenter deliberately changes conditions (e.g., feedback specificity) to explore the dependent variable of interest (e.g., learning). The participant plays the role of managing a furniture company for 18 simulated weeks. The participant managers’ main goal is to increase the efficiency of their employees by motivating them to work as quickly as possible. Empirical studies have shown that participants who are provided with highly specific feedback outperform participants with less feedback specificity. However, those in the high feedback specificity condition are less likely to independently explore the dynamic system (Goodman, Wood, & Hendrickx, 2004).

Previous research into learning and transfer (Pillinger, 2004) has involved changing the “cover story” of the same computer simulation. For example, the test developer of the Furniture Factory (Wood & Bailey) designed an analogous complex system called the Cricket Game. In the Furniture Factory task, the participant plays the role of the manager of a company whose main goal is to improve the efficiency of the workers. In
the Cricket Game complex system, the participant plays the role of the coach of a cricket team whose main goal is to improve the batting performance of the players within a women’s cricket team. Both the Furniture Factory and the Cricket Game simulations are based on exactly the same underlying mathematical algorithms, only the “cover story” has changed.

The FURNITURE FACTORY was employed, for the first time, to assess complex problem solving competencies in this thesis. The relationship between FURNITURE FACTORY performance and specific components of intelligence is discussed in detail in Chapters 3 and 4.

HUNGER IN THE SAHEL (Leutner & Schrettenbrunner, 1989) is described by its author as comparable to computer simulation games such as SimCity. It is currently employed as an instructional computer simulation for geography classes in European schools. The participant plays the role of a farmer in North-Africa. The task is used to investigate discovery learning in complex domains (Leutner, 2002). The participant is instructed to acquire knowledge regarding how to survive as a farmer in this developing country and makes decisions regarding family planning, education, mechanical modernization, etc. In addition, the participant farmer attends to 10 lots of land that differ in geographical quality such as gradient. In a recent study (Leutner) employing HUNGER IN THE SAHEL, it was hypothesised that the relationship between intelligence and problem solving ability could be explained by an inverted u-shaped relationship between the score of the correlation coefficient (dependent variable) and the
extent of available domain-specific prior knowledge (independent variable). This explanation is based on the Elshout-Raaheim hypothesis (Elshout, 1987; Raaheim, 1988). Results supported the hypothesis, it was found that in conditions of low domain knowledge, the correlation was low. With increasing knowledge, the correlation also increased. With further increasing knowledge, the correlation decreases. When the problem has become a simple task, the problem is again low. These results were used in attempt to resolve the different cognitive demands (Dörner, 1986; Pulz-Osterloh, 1993; Rigas & Brehmer, 1999) versus low reliability (Buchner, 1995; Funke, 1992, 1995) debate (Leutner, 2002). However, results must be interpreted with caution because this simulation is subject to the difficulties encountered in analysing results of any highly complex task such as Lohhausen (Dörner, Kreuzig, Reither, & Stäudel, 1983). For example, absence of a well-defined scoring protocol and complexity beyond participant’s comprehension without extensive testing time. Total testing time in the study just described was 70 minutes (Leutner). This is significantly less time than would be allowed for a much less complex task such as FSYS 2.0 (Wagener, 2001). There has not been any subsequent research to support Leutner’s findings or its implications for the relationship between complex problem solving tasks and intelligence.

LOHHAUSEN (Dörner, Kreuzig, Reither, & Stäudel, 1983) consists of more than 2000 variables (e.g., number of inhabitants, earnings of the industry, etc.). Participants play the role of the mayor of a small German town named Lohhausen. They were instructed to take care of the future prosperity of the town over 10 simulated years. Testing time was eight two-hour sessions. The investigation is exploratory without a clearly defined
scoring protocol. Problem-solving competence on this task did not correlate with Raven’s Advanced Progressive Matrices (APM: Raven, Raven, & Court, 1993), or with scores on the Culture Fair Intelligence Test (CFT: Cattell & Weiss, 1980).

**LUNAR LANDER** the original version (Thalmaier, 1979) consisted of three variables, the revised version (Funke, 1998) consists of six variables and is mathematically well defined with a very specific goal. The participant’s goal is to control the landing manoeuvre of a spacecraft on the surface of the moon. Funke and Hussy (1984) presented LUNAR LANDER along with COOKING (an analogous task to LUNAR LANDER with a different cover story). It was expected that males would outperform females on LUNAR LANDER and females would outperform males on COOKING due to hypothesised gender differences in the semantic domain. However, no significant gender differences were found in complex problem solving competencies for either version of the task.

In another study, Hussy (1989) found that complex problem solving performance decreased as the number of variables (complexity) and their interrelations (connectivity) increased. They also found significant correlations between intelligence and LUNAR LANDER performance increased as the transparency of the task increased. Hussy found the BIS K-factor (processing capacity, capturing the ability to recognise relations and rules and to form logical inferences in figure series, number series, verbal analogies) to be the most predictive factor of complex problem solving competencies regardless of the transparency condition of the task and system knowledge. Hussy also found that the BIS-
G (indicating memory performance) was significantly correlated with LUNAR LANDER performance in the nontransparent condition. These findings are consistent with those found for the TAILORSHOP task (Hörmann & Thomas, 1989) although the TAILORSHOP is a more complex task than LUNAR LANDER so caution must be taken with direct comparison of results. These findings suggest that intransparent conditions create high memory demands as participants try to infer the underlying causal model of interconnections between variables (Buchner, Funke, & Berry, 1995).

**MORO** (Dörner & Kreuzig, 1983) consists of 49 variables. The participant plays the role of an advisor to a tribe in Africa whose goal is to improve the living conditions of nomads in the Sahel zone. In a study employing this task, it was found that test-retest reliability was greater for behavioural indices (e.g., the number of questions presented) than for the automatic state of the system (e.g., the number of starving people). Participants were debriefed and found the simulation to have high face validity and reported that their perceived complex problem solving performance during the task was equivalent to real world problem solving situations (Strohschneider, 1986). In another study, the specificity of the goals was manipulated. In the specific goal condition, participants were asked to reach specified values on critical variables (e.g., the number of starving people, the number of cattle, etc.). The low specific goal condition simply asked participants to aim for long-term improvements in the general living conditions of the MORO tribe. In the low specific goal condition, the correlation between problem solving competencies and general intelligence measured by the Berlin Intelligence Structure (BIS) test (Jäger, Süß, & Beauducel, 1997) was not significant. In contrast, the
correlation was significant ($r = .59$) between MORO performance and general intelligence in the specific goal condition (Strohschneider, 1991).

**MULTIFLUX** (Kröner, 2001) is the simulation of a fictitious machine. The participant’s goal is to identify the rules of how the four controls of the machine relate to its displays. MULTIFLUX is a relatively new computer simulation that was deliberately designed to possess high construct validity and to produce reliable scores. The developers of MULTIFLUX based their design on a review of previous naturalistic decision making (NDM) findings reporting low correlations between computer simulations and traditional intelligence tests (Sweeney & Sterman, 2003). Therefore, MULTIFLUX was deliberately designed, from the ground up, to have high correlations with traditional intelligence tests. The advantages and disadvantages of this deliberate design will be further discussed in Chapter 3. The correlation between Multiflux scores and BIS-K (processing capacity, capturing the ability to recognise relations and rules and to form logical inferences in figure series, number series, verbal analogies) was significant ($r=.65$) (Kröner, Plass, & Leutner, 2005). A theoretical model of complex problem solving performance was proposed with three latent factors: rule identification, rule knowledge, and rule application. Structural Equation Modelling (SEM) analyses favoured the three factor Multiflux variables over a single factor model (Kröner, Plass, & Leutner, 2005). An extended theoretical model of complex problem solving performance is provided in Chapter 5 of this thesis.
**POWERPLANT** (Wallach, 1997) is the simulation of a coal-fired power plant based on a real power plant near Saabrüken in Germany. The system’s structure is similar to the SUGAR FACTORY (Berry & Broadbent, 1987) or TRANSPORTATION (Broadbent, 1977). General intelligence was correlated .33 with POWERPLANT (Süß, Oberauer, & Wittmann, unpublished).

**SPACE FORTRESS** (Mané & Donchin, 1989) was designed by cognitive psychologists and used as a common platform for a group of international researchers, called the ‘Learning Strategies Project’ (Donchin, 1995). This group of psychologists (Arthur, Strong, Williamson, Jordan, & Regian, 1995; Shebilske, Goettl, Corrington, & Day, 1999) were able to collaborate their research findings and draw general conclusions regarding skill acquisition in a complex environment. Participants control the movement and weapons of a space ship with the main aim to destroy the SPACE FORTRESS while protecting their own ship from damage, from the shells fired by the space fortress and the mines that surround it. The game ends with destruction of the space fortress (participant wins) or destruction of the space ship (participant loses). The overall score is the damage to the hostile minus the damage to the ship. This information is presented to the participants on screen throughout the game (Mané & Donchin).

The SPACE FORTRESS features many of the characteristics of a computer game including sound, competition, motivation, enjoyment, win/lose ending, joystick and trigger control, and a score that is displayed on screen (Mané & Donchin, 1989). In this way, the task differs from traditional complex problem solving computer simulations. It
shares many features with those employed by researchers who are attempting to convert traditional cognitive abilities tests, such as processing speed (Gs) into game-like simulations (McPherson & Burns, 2005; Washburn, 2003). It has been suggested, perhaps in jest, that it would be difficult to justify funding to an external party if a complex problem solving task appeared too ‘game-like’ and not ‘serious and academic’ enough (Washburn).

SPACE FORTRESS is a much more complex research tool for studies of skill acquisition than those employed before the advent of computers into the laboratory (Adams & Reynolds, 1954). SPACE FORTRESS has a set of 50 parameters (e.g., speed of hostile elements), which can be adjusted by the experimenter for cognitive research purposes. The resulting data consists of 150 variables to describe the participant’s behaviour including response speed, joystick movements, number of missiles fired, etc. These results can be analysed to provide information regarding perceptual processes, cognitive abilities, motor skills, knowledge of rules and game strategy (Mané & Donchin, 1989).

Previous problems were detected with previous versions of SPACE FORTRESS and have now been rectified. Experimenters observed that some participants would achieve high levels of performance in the absence of skill acquisition. These participants disregarded the task instructions and found a loophole that they could exploit to win the game by means other than which the task was designed (Mané & Donchin, 1989). Recall that all computer simulations are based on underlying mathematical algorithms and are subject to loopholes. While exploiting loopholes for an easy path to success may be
acceptable in real life, it is not acceptable in experimental simulations. The resulting scores cannot be interpreted because the experimenter cannot distinguish high scores resulting from those who actually learnt how to perform the complex skill and high scores achieved from exploiting a loophole. An effort must be made to maintain the complexity of the game for all participants to allow investigation of complex problem solving performance. It is important that participants must attend to all the elements of the game in order to score well. These problems can affect any computer simulation and their presence will be examined in the computer simulations employed in this thesis.

TAILORSHOP designed by Dietrich Dörner and first used in a study by (Putz-Osterloh, 1981) consists of 24 variables. The participant plays the role of the manager of a small shirt manufacturing company. The main goal is to maximise company capital by purchasing raw materials and modifying the production capacity in terms of number of workers and number of sales outlets, etc. In a study by Putz-Osterloh and Lüer (1981) participants were administered a transparent condition of the task where they had access to a diagram depicting the relations between the system variables, another group of participants were administered the intransparent (opaque) condition in which no such diagram was provided. A statistically reliable relation ($\tau = .22$) was found between IQ and problem solving competence. It was concluded that the relationship between problem solving competence and global intelligence is mediated by the transparency of the problem solving task. Contradictory results were found by Funke (1983).
In subsequent studies (Süß, Kersting, & Oberauer, 1991, 1993) a correlation between TAILORSHOP performance and the BIS-K factor (processing capacity, capturing the ability to recognise relations and rules and to form logical inferences in figure series, number series, verbal analogies) was found ($r = .47$). It was concluded that reasoning ability, as assessed by the BIS-K was important for participants to infer the underlying causal model of the system, from which relationships between the variables could be established. These findings were consistent with those found by Hörmann and Thomas (1989). In the nontransparent condition of the TAILORSHOP when participants had high system knowledge (as assessed by a questionnaire), the BIS-K factor and the BIS-G factor (indicating memory performance) correlated with problem solving competencies, correlations are $r = .72$ and $r = .54$ respectively. In the transparent condition of the task, with participants who reported high system knowledge, the best predictor of problem solving competencies was the BIS-B (processing speed). When all participants were included in the sample, regardless of high or low system knowledge, no significant correlations were found in the nontransparent condition.

In studies of gender differences (Putz-Osterloh, 1993; Süß, Oberauer, & Wittmann, 2005), strong gender effects favouring males have been found on TAILORSHOP task performance. It was suggested that males’ superior general economics knowledge led to improved TAILORSHOP task performance (Süß et al., 2004). These findings are in contrast to the Funke and Hussy (1984) study. The TAILORSHOP was employed to assess complex problem solving competencies in this thesis. The relationship between
TAILORSHOP performance and specific components of intelligence is discussed in detail in Chapter 4.

**WATER PURIFICATION PLANT** (Gonzalez, Lerch, & Lebiere, 2003) this computer simulation was based on the real world task of mail sorting for the United States Postal Service USPS. However, the USPS interface has been simplified. The participant plays the role of the plant operator whose main goal is to distribute water to different locations before expiration of the deadline. The simulation defines the arrival time, the amount of water dispersed, and the destination tank. Correlations between performance on Water Purification Plant and two tests of cognitive ability Ravens Progressive Matrices (RPM, standard or advanced version: Raven, 1962, 1977), a measure of fluid intelligence, and the VSPAN, a measure of visual working memory (Shah & Miyake, 1996) were around $r = .3$. Correlations between performance on team Water Purification Plant and two cognitive abilities tests RPM and VSPAN were around $r = .6$ (Gonzalez, Thomas, & Vanyukov, 2005).
APPENDIX B

B Post-test questionnaire measuring rule knowledge of Furniture Factory simulation

Please indicate whether, in your opinion, each of the following statements about the Furniture Factory is true or false (enter a ‘t’ or ‘f’).

1. When an employee is performing badly, setting a difficult goal (i.e., 75% of estimated) lowers their performance in the following week.

2. Difficult goals (i.e. 75% of estimated) generally have a more positive effect on worker performance than ‘do your best’ goals.

3. It is better to set the same goal level for the whole work group than assign individuals different goals.

4. Level of reward for an individual worker should be based solely on their level of performance, regardless of group performance.

5. If more than 3 or 4 members of the group perform badly, all group members should receive no reward.

6. Low performers are more affected by level of reward than high performers.

7. If performance is below standard, the best form of feedback is to advise employees of their performance and discuss their work with them (advise and discuss).

8. If performance is above standard, the best form of feedback is to advise employees of their performance only (advise).

9. For a high performer, discussing their work with them over a number of weeks will cause them to lower their subsequent performance.

10. It is a good idea to move people regularly between jobs at the Furniture Factory.

11. The best job for Dave is in the warehouse.

12. Evelyn performs at her best in the sewing room but can also work well on fabric cutting.

13. Janice is multi-skilled in furniture making but performs particularly well in upholstery.

14. Dave is generally easier to motivate than Bert.
Appendix B continued

Post-test questionnaire measuring rule knowledge of Furniture Factory simulation

Please indicate whether, in your opinion, each of the following statements about the Furniture Factory is true or false (enter a ‘t’ or ‘f’).

15. Overall, Evelyn is more difficult to motivate than Hilary.

16. Sewing is the only skill that Hilary has.

17. Bert is best suited for work on assembling furniture.

18. Janice is very hard to motivate.

19. There are 3 levels of feedback that can be given to employees.

20. Advising on performance level in relation to estimated hours is the most specific form of feedback available.

21. The most difficult goal level that can be set for employees is to complete their task within 75% of estimated time.

22. Of the 5 options for setting employee targets, 1 is less specific than the other 4.

23. There are 3 levels of reward available.

24. Putting up a memo on the factory floor is the highest form of reward available.
APPENDIX C

C Participants handout for the Furniture Factory task

MANAGEMENT INFORMATION PROVIDED

To assist in these decisions, the following information is available:

**Descriptions of the jobs** that might be required for a special order. Different orders may require different combinations of jobs from the following set of five different kinds of jobs:

- **Assembly** – Finished timber is assembled.
- **Fabric Cutting** – Upholstery material is cut to pattern.
- **Sewing Room** – Cut material is sewn.
- **Upholstery** – Sewn material, padding and springs fixed to furniture.
- **Finished Goods Warehouse** – Storage and movement of finished goods.

**Details of employees** on the Special Order Roster, including skills and aptitudes. Different employees from the following list will be available for the Special Order Roster from week to week:

1. **Bert**: is recently out of school. He studied woodwork and metal work, but he is not highly skilled. He sometimes seems more interested in his car than in his job, and he tends to be rather slipshod in his approach.

2. **Dave**: has been with the company for a few years now. He began as a general carpenter, but he is now highly skilled in most forms of woodwork, and is starting to learn some upholstery work. He is highly motivated, and works quickly and carefully.

3. **Janice**: is a first-class upholsterer. She began in the trade as her father’s assistant, in his small furniture repair shop, and supplemented this practical apprenticeship with evening classes in upholstery and woodwork. She is meticulous in her approach. She has acquired a range of general woodworking skills and she can sew if necessary, but upholstery is her forte.

4. **Hilary**: has been with the company for only a few years. She is a seamstress and dressmaker by training, but was made redundant when her previous employer got into financial difficulties. She takes pride in her work, both in her furniture covers and the dressmaking and embroidery which occupy her evenings.

5. **Evelyn**: is a new employee who enjoys fabric cutting and is able to perform simple sewing tasks. She has few skills in other areas and is not motivated to learn new tasks.
Appendix D

D Tailorshop Simulation - Participant Reference Sheet

Important Information about the Tailorshop

The following information has been made available to you by the board of directors of the shirt factory to begin your position as manager. You can always refer to this sheet during your company leadership.

1. Investment costs and Company Expenditure:

<table>
<thead>
<tr>
<th></th>
<th>Buy</th>
<th>Sell</th>
</tr>
</thead>
<tbody>
<tr>
<td>50er-Machines</td>
<td>$7,500</td>
<td>$6,000</td>
</tr>
<tr>
<td>100-er Machines</td>
<td>$15,000</td>
<td>$12,000</td>
</tr>
<tr>
<td>Trade 50er for 100er</td>
<td>$9,000</td>
<td>--</td>
</tr>
<tr>
<td>Selling Position</td>
<td>$7,500</td>
<td>$6,000</td>
</tr>
</tbody>
</table>

- The Raw Material prices vary according to the market conditions. Last year the average price was 5 dollars per unit of Raw Material.

- If an employee is fired, you must pay them a months salary as termination compensation.

- The bank supplies you with a generous amount, but not unlimited, credit. If you can receive no more credit for a purchase, it will be displayed on the screen.

2. Company Organization

- For every machine – whether it is a 50er or a 100er – the following rule applies: One employee per machine is satisfactory for the machine to be operated properly.

- A sales representative does not count as an employee. They charge the company a monthly fee of $500.
3. Definition of the Important Variables

- **Workload Efficiency** refers to what percentage of the shirts the employees *can* produce, compared to what they *actually* produce.

- **Assets** is that which is all together in the bank account, from the sales value of the machines and sales outlets, and the value of the shirts in stock and the raw material in stock.

- **Sales Representatives** are agents independent of the factory and deal with wholesale and retail offers.

- **Machines**: With every **100er machine** you can manufacture about 100 shirts in a month, with every **50er Machine** about 50.

- **Machine Efficiency** refers to what percentage of the shirts the machines *can* produce, compared to what they *actually* produce.

- **Machine Breakdown** refers to what percentage of the machines are damaged. 100% machine damage means that all machines are completely broken.

- **Demand** is the quantity of shirts from your factory that the buyers want to purchase.

- **Raw material in stock** refers to the number of shirts that the raw material in stock is sufficient for.

- **Social costs** refers to the total expenditure for voluntary social services and benefits for the company, e.g., for social meeting rooms, a canteen etc.

- **Sales outlets** are the number of shirt shops that the factory owns.

- **Maintenance costs** is the expenditure for the maintenance and repair of machines. Maintenance and repair is provided by a service company, that is paid for according to number of working hours.
Appendix E

E  Tailorshop Simulation – Directions for The Second Practice Month

Quickly go through the given catalogue of interventions. What you undertook during your training months, is not suggested in the summary.

Carry out the following plans in 2 months:

- Set the number of sales representatives to 1 (the sales representatives do not count as employees)
- Raise the Maintenance costs to $1850
- Trade in three 50-Shirt-Machines for 100-Shirt Machines ??ILLOGICAL. (You will see it will immediately informs/retrains 3 employees)
- You buy five additional 100-Shirt Machines
- You recruit five employees for the 100-Shirt-Machines five employees
- You buy raw materials for 500 shirts
- Increase the wages to $1150
- You think about other options and would prefer to buy only 4 100-Shirt-Machines. Change your plans accordingly.

Beforehand you swapped three 50-Shirt-Machines for 100ers and have now bought an extra four, so now under Plans you will see there are 7 100-Shirt-Machines. If this is not the case, please report it.

When you have finished, you can enable your plans to be put into action and end the month by pressing the a key. You will then be informed of the end result of these 2 months. Please do this now. Of course, will see that your interventions have changed the conditions.

When you are finished, turn over this sheet of paper so that the experimenter can see that you are finished.

***

Information for the Trial Manager

Variable Definitions in SWS

In the following are definitions and explanations about the all variables (though some are hidden??) listed in the SWS. The variable definitions will be elucidated in the instructions and continually for any individual inquiries that may arise while the task is in progress. Detailed information about the variables can be taken from the handout.

Employees for 50er and for 100er Machines: The number of employees working on 50-shirt-machines and on 100-Shirt-Machines respectively.
**Workload Efficiency:** Percentage of the maximum possible shirt production of the employees, that is actually produced.

**Work Motivation:** Employee preparation, employee commitment. 100% employee motivation means that all employees are totally committed.

**Shirts in stock:** The number of finished shirts in stock

**50-er Machines and 100er Machines:** There are two different machine types. With the 50-Shirt-Machines you can make about 50 shirts at a time. With the 100-Shirt-Machines you can produce about 100 shirts in a month.

**Total Assets:** Total value of the shirt factory. The total assets consists of the money that is situated in the company bank account, along with the value of the machines, the raw materials, the shirts in stock and so on.

**Sales Representatives:** Sales representatives are independent employees from the factory, and handle the retail and wholesale shirt sales.

**Shirt price:** Selling price of a shirt.

**Bank balance:** Money from the bank account that is available at any time.

**Wages:** Monthly wages of the employees.

**Machine Efficiency:** Percentage of shirts actually produced in relation to the maximum possible number of shirts produced from the machines.

**Machine Damage:** The number, expressed as a percentage, of machines that have acquired damage in relation to the number of functioning machines.

**Demand:** Willingness of the shoppers to buy shirts.

**Production:** Quantity of shirts that will be produced in a month.

**Production Loss:** The quantity of possible production that is not produced or realized, expressed as a percentage.

**Raw Materials:** The number of shirts that can be produced from the raw material in stock.

**Raw Material Price:** The costs of raw materials per shirt.

**Social costs:** Expenditure on voluntary social services and benefits for the employees (e.g. expenditure on social meeting rooms, a canteen, etc..)
**Sales Outlets:** Sales outlets are shirt shops owned by the factory.

**Shirts sold:** Number of shirts sold.

**Maintenance:** Expenditure on the maintenance and repair of the machines.

**Advertising:** Advertising costs.
Appendix F

F Test manual for Forest microworld task FSYS 2.0

FSYS 2.0
Testmanual
Inhalt Contents

INHALT CONTENTS 293

SPEZIFIKATIONEN SPECIFICATIONS 294

TECHNISCHE DOKUMENTATION TECHINICAL DOCUMENTS 296

SYSTEMVORAUSSETZUNGEN FÜR DEN EINSATZ VON FSYS 2.0 SYSTEM REQUIREMENTS FOR THE USE OF FSYS 2.0 296
UNTERSTÜTZTE BETRIEBSSYSTEME SUPPORTED SYSTEM SOFTWARE 296
EMPFOHLENEN BETRIEBSSYSTEME REQUIRED SYSTEM SOFTWARE 296
HARDWARE-MINIMUM Minimum HARDWARE 296
EMPFOHLENEN MINDESTHARDWARE REQUIRED MINIMUM HARDWARE 296
GRAFIKAUFLÖSUNG DES SYSTEMS GRAPHICS REQUIREMENTS 296
BETRIEB UNTER WINDOWS NT Operation UNDER WINDOWS NT 297
DIE BESTANDTEILE VON FSYS 2.0 THE COMPONENTS OF FSYS 2.0 297
INSTALLATION VON FSYS 2.0 INSTALLATION OF FSYS 2.0 297

ANMERKUNG FÜR BENUTZER VON WINDOWS 3.1 EXPLANATORY NOTES FOR USERS OF
WINDOWS 3.1 298
DEINSTALLATION VON FSYS 2.0 DEINSTALLATION OF FSYS 2.0 298
EINSATZ EINES LIZENZSERVERS USE OF A LICENCE SERVER 298
Ablaufen einer Testsitzung mit FSYS 2.0 Running a Test Session With FSYS 2.0 300
Vorgehen bei Unterbrechungen der Testsitzung Overriding Interruptions of a Test
Session 301

AUSWERTUNG EINER PROTOKOLLDATEN ANALYSIS OF DATA FILES 302

AUSWERTUNG: ERSTE SCHritte ANALYSE: FIRST STEPS 302
AUSWERTUNG: EINSTELLUNGEN ANALYSE: SETTINGS 305
AUTOMATISCHE AUSWERTUNG AUTOMATIC ANALYSIS 306

SKALEN SCALES 307
S: Gesamtergebnis S: Total Score 307
SKAP 307
SKAPKOR 307
M: Maßnahmenqualität M: QUALITY OF ORDERS 307
MNOFEHL 307
MPRIORI 308
MEFFIZI 308
MVERSTA 309
I: INFORMATIONSMANAGEMENT I: INFORMATION MANAGEMENT 309
IORIENT 309
IVORHAND 309
IFEEDS 309
IKONTI 310
IKONTIS 310
IMODUS 310
E: SELBSTMANAGEMENT E: SELF MANAGEMENT 310
ESICHER 310
Spezifikationen

Specifications

Konstrukt: Fähigkeit zur Lösung komplexer Probleme (nach DÖRNER).
Construct: Ability to solve complex problems (after DÖRNER).
Medium: PC-gestützte Darbietung, direkte Interaktion des Teilnehmers mit dem PC.
Medium: PC-based presentation, direct participant interaction with the PC.
Einkleidung: Teilnehmer hat die Aufgabe, einen forstwirtschaftlichen Betrieb in einer fiktiven Zukunft zu leiten.
Presentation: Participants are presented with the task of managing an economic forestry business in a fictitious future.

Ablauf: ca. 20 Minuten Einführung des Teilnehmers (10 Minuten Lesen des Einführungstexts, dann 10 Minuten praktische Einführung), danach maximal 90 Minuten selbständige Bearbeitung des Szenarios durch den Teilnehmer (Anwesenheit eines Testleiters wünschenswert).
Procedure: Approximately 20 minutes Introduction for the participants (10 Minutes to read the Introductory text, then 10 minutes practical introduction), subsequently, a maximum of 90 minutes with participants independently working through the scenarios (the presence of a test leader is prefereable).

Teilnehmer: Grundsätzlich kommen alle Personen in Frage, die keine Angst vor dem Umgang mit einem PC haben. Falls keinerlei Vor erfahrung mit Microsoft Windows vorliegt, sollte eine individuelle Einführung in den Umgang mit Maus und Dialogboxen vorgeschaltet werden. Von seiner Auslegung her ist der Test eher zur Anwendung bei zumindest durchschnittlich intelligenten Personen gedacht. Die Teilnehmer sollten ausgeruht sein und sich geistig "fit" fühlen.
Participants: Basically, all people will be questioned to ensure they have no anxiety about working with a PC. In the case where an individual has no previous experience with Microsoft Windows at all, then experience should be gained by using the mouse and dialogue boxes. From the previous interpretation, the test is better implemented with a person thought to have at least average intelligence. The participant should be rested and feeling mentally “fit”.

Gruppen: Die Durchführung als Gruppentest ist möglich, sofern genügend geeignete PCs zur Verfügung stehen. Die Einführung dürfte dann etwas länger dauern.
Groups: The implementation of this as a group test is possible provided there are sufficient suitable PCs available. The introduction should then also take somewhat longer.

Auswertung: PC-gestützte, vollautomatische Auswertung (kriterienorientiert und normorientiert).

Analysis: PC-compatible, fully automatic analysis (criterion oriented and rule-oriented)

Objektivität: Da die Durchführung in direkter Interaktion des Teilnehmers mit dem Rechner stattfindet, und da keinerlei Zufallsvariablen den Systemzustand beeinflussen, ist die Objektivität grundsätzlich sehr hoch. Bei Personen mit sehr wenig Computererfahrung können allerdings naturgemäß Probleme entstehen, die durch zusätzliche individuelle Einführung ausgeglichen werden müssen.

Objectives: Given that the performance of the participants takes place via direct interaction with the computer, and assuming no random variables will influence the system status, the objectivity is, in principle, very high. For people with very little computer experience, problems can indeed emerge, and can be corrected by additional individual introduction.
Technische Dokumentation
Technical Documents

Systemvoraussetzungen für den Einsatz von FSYS 2.0
System Requirements for the use of FSYS 2.0

Unterstützte Betriebssysteme
- Windows bzw. WfW 3.1x mit Win32s Version 1.30
- Windows 95
- Windows NT 3.51
- Windows NT 4.0

Supported System Software
- Windows respectively Windows for Workgroups (WfW) 3.1X with Win32s Version 1.30
- Windows 95
- Windows NT 3.51
- Windows NT 4.0

Empfohlene Betriebssysteme
Required System Software
- Windows 95
- Windows NT 3.51
- Windows NT 4.0

Hardware-Minimum
Minimum Hardware
Intel 386, 8 MB RAM, Festplatte, VGA-Grafikkarte, 14" - Monitor, Maus
Intel 386, 8 MB RAM, Hard Disk, VGA Graphics card, 14“ Monitor, Mouse

Empfohlene Mindesthardware
Minimum Hardware Requirements (Recommended)
Intel 486, 33 MHz, 16 MB RAM, Beschleuniger-Grafikkarte mit passendem Treiber, 15"- Monitor
Intel 486, 33 MHz, 16 MB RAM, Accelerated Graphics Card with compatible driver, 15” monitor

Grafikauflösung des Systems
System Graphic Requirements
Grundsätzlich sind HiColor- (>32000 Farben) oder True-Color-Grafikmodi zu empfehlen.
Basically HighColour (>32000 Colours) or True-Colour-Graphics mode is required.
Unter Windows 95 und NT 4.0 sind auch 256 Farben ausreichend. Unter Win 3.11 / NT 3.51 stellen 256 Farben nur eine Notlösung dar (das Programm funktioniert prinzipiell), 16 Farben-Modi stellen grundsätzlich nur eine Notlösung dar.
Under Windows 95 and NT 4.0 256 colours are also sufficient. Under Win. 3.11 / NT 3.51 256 colours only represents a temporary solution (the program will still function), 16-Colour-mode basically only represents a temporary solution.
Für den Betrieb genügen 640x480 Punkte. Höhere Auflösungen werden nicht ausgenutzt, können aber trotzdem eingestellt werden. Je nach Qualität und Größe des Monitors sollte eine Auflösung gewählt werden, bei der die Schrift gut lesbar, andererseits aber nicht zu groß ist. Auf einem 15"-Monitor wären also 640x480 oder 800x600 Punkte geeignet, auf einem 17"-Monitor 800x600 - 1024x768, usw.

For basic operation, a resolution of 640x480 pixels will suffice. A higher resolution will not be utilised, but may be used nonetheless. Resolution should be selected according to the quality and size of the monitor, so the font is readable, but on the other hand not too big. So for a 15-inch monitor, a resolution of 640x480 or 800x600 is suitable, for a 17-inch monitor 800x600 – 1024x768 and so forth.

**Betrieb unter Windows NT**

**Operation under Windows NT**


Under Windows NT, the license disk does not function in the usual manner. One then requires a separate license server that is connected with the NT-computer via the network. The license disk will then be written off via this license server. Every network-compatible PC (under DOS, Wfw 3.11, Windows 95) is qualified as a license server.

**Die Bestandteile von FSYS 2.0**

**The Components of FSYS 2.0**

FSYS 2.0 besteht aus den folgenden Programmen:

FSYS 2.0 is composed of the following programs:

- FSYS-Lernprogramm
- FSYS-Testdurchführung
- FSYS-Lizenzverwaltung
- FSYS-Auswertungsprogramm
- FSYS-Training Program
- FSYS-Test Performance/Execution
- FSYS-License Management
- FSYS-Analysis Program


These programmes can all be installed from the FSYS Installation disk. In order to start an FSYS Test Session, an FSYS License disk will be required. These measures help to protect the copyright of the FSYS software.

**Installation von FSYS 2.0**

**Installation of FSYS 2.0**

Legen Sie die Installationsdiskette in das Laufwerk ein, und starten sie das Programm SETUP.EXE auf der Diskette. (falls Sie nicht wissen, wie man das macht, schlagen Sie bitte in
Ihrem Windows-Benutzerhandbuch nach.) Das FSYS-Installationsprogramm erlaubt dann zunächst die Angabe des Pfades für die Installation. Sie können den Pfad entweder direkt eingeben, oder in dem Auswahlbereich mit der Maus einen Pfad aussuchen. Sie können auch beide Verfahren kombinieren.

Insert the Installation disc in the hard drive, and start the program SETUP.EXE on the disc. (If you do not know how to do this, please consult your Windows User Handbook). Once the FSYS Installation Program is run, you will need to specify the location where you wish to install FSYS. You can give the path either directly or browse for a path in the selection area using the mouse. You can also combine both methods.


After you have made your selection, the necessary data will be installed on your hard drive. A status bar will display the relative progress (of the installation). Finally, the icons for the individual FSYS Components will be generated.

Damit kann FSYS angewendet werden.

Now FSYS can be used.

Anmerkung für Benutzer von Windows 3.1
Remarks for Users of Windows 3.1

Da FSYS 2.0 ein 32-Bit Programm ist, kann es nicht ohne weiteres unter Windows 3.1 ausgeführt werden. Sie benötigen die Win32s-Erweiterung von Microsoft, die kostenlos erhältlich ist. Falls Sie keine andere Quelle haben, können Sie den Autor von FSYS 2.0 kontaktieren, um Win32s zu erhalten. Bevor Sie FSYS 2.0 installieren können, müssen Sie dann zuerst Win32s installiert haben.

FSYS 2.0 is a 32-Bit Program and it cannot be run under Windows 3.1. You can obtain the Win32 upgrade from Microsoft free of charge. In case you have no other sources, you can contact the author of FSYS 2.0 in order to obtain the Win32 upgrade. Before you can install FSYS 2.0, you must have first installed Win32s.

Unter Win32s kann FSYS nicht in einen bereits existierenden Pfad installiert werden. Sie erhalten dann die Meldung, daß das Verzeichnis nicht erzeugt werden kann. Ursache ist ein von Microsoft offiziell bestätigter Bug im Win32s.

FSYS cannot be installed in an already existing path under Win32s. You will then obtain the notification that the directory cannot be created. The reason is an officially confirmed bug from Microsoft in Win32s.

Deinstallation von FSYS 2.0
Deinstallation of FSYS 2.0

Die Entfernung von FSYS ist sehr einfach. Es genügt, wenn Sie das FSYS-Programmverzeichnis löschen. Schlagen Sie bitte in Ihrem Windows-Benutzerhandbuch nach, falls Sie nicht wissen, wie man die Programmicons von Ihrer Oberfläche löscht.

The removal of FSYS is very easy. It is enough if you delete the FSYS Program Directory. Please consult your Windows User Handbook if you do not know how to delete program icons from your desktop.

Einsatz eines Lizenzservers
Employment of a License Server

Bei Untersuchungen großer Gruppen ist die Benutzung einer FSYS-Lizenzdiskette für jeden einzelnen Teststart unhandlich. Durch Beschränkungen in der BIOS-Emulation von Windows...
NT und bei bestimmten, fehlerhaften BIOS-Versionen einzelner Rechner kann das FSYS-Lizenzverwaltungssystem unter Windows NT oder bei Fehlern im BIOS nicht eingesetzt werden. Wenn die Rechner untereinander vernetzt sind, kann ein beliebiger Rechner als Lizenzserver eingesetzt werden. Alle Abbuchungen von der Lizenzdiskette laufen dann über diesen Rechner.

The use of a FSYS license disk is inefficient when testing large groups. Due to restrictions in the BIOS-Emulation of Windows NT and certain erroneous BIOS-Versions, individual computers may not be able to use the FSYS License Management System. When the computers are networked among one another, any computer can be used as a License server. All direct debiting from the license disk then runs via this computer.


The use of these Licensing procedures assumes competence in DOS and Windows. The operation will be outlined briefly in the following.

Der PC, auf dem das Lizenzverwaltungsprogramm ausgeführt werden soll, sollte Schreib- und Leserechte in einem beliebigen Netzwerkpfad haben, ebenso die Lizenz-Clients (auf denen FSYS ausgeführt wird).

The PC, from which the licence management program will be run, should have sufficient disk space in a Network path, in the same way as the License clients (that will be implemented from FSYS).

In diesen Netzwerkpfad wird die Datei coopnet.exe von der Lizenzdiskette kopiert (mit EXPAND expandieren!). Diese Datei wird dann auf dem Lizenzserver ausgeführt, wobei sich natürlich eine Lizenzdisk im Laufwerk des Lizenzservers befinden sollte. Während der Starts der einzelnen Testsitzungen auf den Lizenz-Clients muß COOPNET aktiv bleiben.

In the selected network path the file coopnet.exe will be copied from the license disk (expanded with EXPAND). This file will then be run from the licence server. A license disk should be in the floppy drive of the license server. During the start of an individual test session from a License Client, COOPNET must remain active on the server.


FSYS should be run from the license-clients with the parameter –l(licensepath), so for example: “FSYS20.EXE –l:\temp\lizsrv\”. Please notice the backslash at the end of the path! Normally it would be sensible to create an icon or shortcut to call FSYS accordingly.


COOPNET has several service options that are not necessary under normal circumstances. That needs “Remount”. This is used to register a license disk that has been inserted after the start of COOPNET. It is unusual for the option “Purge” to be necessary. You should transfer the network dictionary/index to a clearly specified directory, one that is inaccessible to the client. This could be helpful for solving connection problems or other technical difficulties.

Nach Abschluß der Testsitzungen kann auch COOPNET mit "Abbruch" wieder beendet werden. After the completion of the test session you can once again end COOPNET with “Cancel".
Ablauf einer Testsitzung mit FSYS 2.0

Running a Test Session with FSYS 2.0

Stellen Sie bitte rechtzeitig vor Beginn der Sitzung sicher, daß FSYS auf den Computern korrekt installiert ist, daß genügend Lizenzen auf Lizenzdisketten vorhanden sind, und daß genügend einwandfreie Exemplare der FSYS-Probandeninstruktion zur Verfügung stehen.

Before beginning the session, please make sure that FSYS is correctly installed on the computer; ensure that sufficient licenses are available on the license disks, and that sufficient fault-free copies of the FSYS-Subject instructions are available.

Nach Begrüßung der Teilnehmer werden die Probandeninstruktionen ausgehändigt. Den Probanden wird mitgeteilt, daß sie die ersten zwei Seiten der Instruktion auf jeden Fall gründlich lesen sollen, und die restlichen vier Seiten so gründlich, wie sie dies für nötig erachten. Die Probanden dürfen die Instruktion während der eigentlichen Testsitzung behalten. Für Notizen erhalten sie einige Bögen DIN-A4-Papier und einen Stift.

After greeting the participants the subject instructions should be handed out. The subjects should then be informed that the first 2 pages of instructions must be read thoroughly in all cases, and the remaining four pages as thoroughly as considered necessary. The participants are allowed to retain the instructions during the actual test session. Have several sheets of A4 paper and a pen to make notes.


When the participants are finished with the introductory text, the training program should then be started. You will find the instructions for the training program in the appendix. After completing the training program you can begin the test. FSYS will prompt you for the location to which the test data should be written. Please select the appropriate directory and enter in the data names. For this purpose the participant’s ID is normally suitable. Please remember to insert the license disk into the drive near the beginning of the test program. The license administration system displays at the start the number of the licenses remaining on the license disk, and you can press “S” in order to use a license. After FSYS has started successfully, the license disk can once again be taken out from the drive.

Für die 50 simulierten Takte hat der Proband nun 90 Minuten Zeit. Sorgen Sie gegebenenfalls mit sanftem Druck dafür, daß diese Zeit annähernd eingehalten wird.

The participants have only 90 minutes for the 50 simulated cycles. You should make sure that this time is approximately adhered to with gentle pressure as required.

Anschließend muß die FSYS-Protokolldatei gesichert werden. Sie befindet sich an dem beim Start von FSYS angegebenen Ort. Sichern Sie sie auf eine Diskette oder besser auf ein zuverlässigeres Medium. Es lohnt sich, diese Dateien mit einem gängigen Packprogramm (zum Beispiel ZIP) einzupacken, da sie dadurch deutlich kleiner werden.

Afterwards the FSYS data must be saved. The data will be located in the directory you specified earlier. Save the data on a disk, or better still on a more reliable medium. It is worthwhile to pack up this data with a popular archiving program. For example, winzip (www.winzip.com), so that it will be decidedly smaller.
Vorgehen bei Unterbrechungen der Testsitzung Test Session Procedure With Interruptions

Durch technisches Versagen (z.B: Stromausfall) oder Fehlbedienung des Probanden (unbeabsichtigtes Betätigen von "Testleitung" - "Beenden") kann eine Testsitzung unterbrochen werden. Im Gegensatz zu den meisten anderen Szenarios erlaubt FSYS in diesem Fall die Wiederaufnahme der Bearbeitung an dem Punkt, an dem der Test unterbrochen wurde. Nur die Eingaben, die der Teilnehmer im aktuellen Simulationszyklus durchgeführt hatte, gehen verloren und müssen daher wiederholt werden.

Technical failure (e.g., power failure) or faulty operation by the participants (unintentional action for “Test management” – “Cancel/Close/End”) can interrupt a test session. For most other scenarios FSYS allows for the resumption of the test at the point where the test was interrupted. Only the input that the participants had carried out in the actual simulation cycle will be lost, and must therefore be redone.

Zur Weiterführung einer Testsitzung wird FSYS zunächst wie gewohnt gestartet. Es wird allerdings kein neuer Dateiname als Protokolldatei angegeben, sondern der Name der angefangenen Protokolldatei. Bitte beachten Sie, daß auch für Testfortführungen die Lizenzdiskette eingelegt werden muß, da eine neue Lizenz abgebucht wird.

For the continuation of a test session FSYS will start as normal. No new data name as protocol data will be displayed, instead the name of the initial protocol data will be shown. Please note that for test continuation the license disk must be inserted.
Auswertung einer Protokolldatei Analysis of a Data File
Zur Auswertung der FSYS-Protokolldateien (Dateierweiterung "dhi") steht ein spezielles Auswertungsprogramm zur Verfügung. Es erlaubt die Berechnung der FSYS-Skalen samt Normierung an einer wählbaren Vergleichsgruppe. Die Ergebnisse werden auf dem Bildschirm dargestellt und können ausgedruckt werden. Zugleich können die Ergebnisse in einer Datendatei abgespeichert werden, die sich in gängige Programme wie Microsoft Excel, SPSS oder Systat importieren läßt.

For analysis of the FSYS-protocol data (data suffix “.dhi”) there is a special analysis program available. It allows for the calculation of the FSYS-Scales together with the standardisation of a normal comparison/control group. The results will be presented on the screen and can then be printed out. At the same time you can save the results in a data file that you can then import into programs like Microsoft Excel, SPSS or Systat.

Auswertung: Erste Schritte Analysis: First Steps
Starten Sie bitte das FSYS-Auswertungsprogramm. Sie sehen ein Fenster wie im folgenden Bild:

Please start the FSYS Analysis Program. You will see a window like the following picture:

Für die meisten Auswertungen sollte es genügen, die ersten drei Knöpfe in der Button-Leiste zu bedienen.

For most analyses, it should be sufficient to use the first 3 buttons in the toolbar.

Der erste Knopf öffnet die FSYS-Protokolldatei. Betätigen Sie ihn bitte, und wählen Sie in der Dateiauswahlbox die Protokolldatei aus, die ausgewertet werden soll. (Ziehen Sie bitte Ihr Windows-Benutzermanual zu Rate, falls Sie Probleme mit der Dateiauswahl haben sollten.) Bestätigen Sie Ihre Auswahl mit "Öffnen".

The first button opens the protocol datafile. Please use this and choose the data that should be analysed. (Please consult your windows user manual if you have problems with the data selection). Confirm your selection with “Open”.

Zur Bestätigung erscheint im Fenster des Auswertungsprogramms der komplette Pfad der gewählten Protokolldatei und die Meldung "Diese Protokolldatei kann jetzt ausgewertet werden":

Für die meisten Auswertungen sollte es genügen, die ersten drei Knöpfe in der Button-Leiste zu bedienen.

For most analyses, it should be sufficient to use the first 3 buttons in the toolbar.

Der erste Knopf öffnet die FSYS-Protokolldatei. Betätigen Sie ihn bitte, und wählen Sie in der Dateiauswahlbox die Protokolldatei aus, die ausgewertet werden soll. (Ziehen Sie bitte Ihr Windows-Benutzermanual zu Rate, falls Sie Probleme mit der Dateiauswahl haben sollten.) Bestätigen Sie Ihre Auswahl mit "Öffnen".

The first button opens the protocol datafile. Please use this and choose the data that should be analysed. (Please consult your windows user manual if you have problems with the data selection). Confirm your selection with “Open”.

Zur Bestätigung erscheint im Fenster des Auswertungsprogramms der komplette Pfad der gewählten Protokolldatei und die Meldung "Diese Protokolldatei kann jetzt ausgewertet werden":
For confirmation, the complete path of the selected protocol data and the message “This data file can now be analysed” will appear in the window of the analysis program.

Um die Berechnungen zu starten, drücken Sie nun bitte den zweiten Knopf. Es erscheint die folgende Dialogbox:

In order to start the calculations, please now press the second button. It displays the following dialogue box:

Unter "Vergleichsgruppe" können Sie auswählen, mit welchen Teilnehmern der aktuelle Proband verglichen werden soll. Davon hängt die Berechnung der Prozentrangwerte ab, die dem aktuellen Teilnehmer zugewiesen werden.

Under “control group” you can select which participants are comparable to the actual participants. There it depends on the calculation of the percentage ranges (percentile bands or percentile rank order) that the actual participants will be assigned to.

Wenn Sie die Option "Ergebnisse zusätzlich in Dateien ... schreiben" aktivieren, werden die Ergebnisse des aktuellen Teilnehmers nach der Auswertung automatisch in die angegebenen Dateien geschrieben. "roh.txt" enthält dabei die Rohpunktwerte, und "proz.txt" die Prozentrangwerte. Falls die Dateien schon existieren, wird der neue Fall hinten angefügt. Dadurch können Sie am Ende einer Auswertungssitzung alle Ergebnisse auf einen Schlag in das Datenanalyseprogramm Ihrer Wahl importieren (Format: ASCII-Datei mit Tabulatortrennung).

When the option “Additional results in data .... write” is selected, the results of the actual participants will automatically be written in the specified data file after the analysis. “raw.txt” contains the raw item cores, and “proz.txt” the percentile bands. In the event that the data already exists, the new case will be appended to the file. Thus, you can import all results from a folder to the data analysis program of your choice after the analysis session. (Format: ASCII-Data with tab separations)
Wählen Sie eine Vergleichsgruppe aus, entscheiden Sie sich für oder gegen das Schreiben der Datendateien, und betätigen Sie dann "OK". Ein Meldungsfenster erscheint:

Select a comparison group, decide for or against the writing of the data file, and then execute and "OK". A message window displays:

Nach Abschluß der Berechnungen (Dauer je nach Hardware ca. 5 - 30 Sekunden) erscheinen die Ergebnisse. Die folgende Abbildung zeigt nur einen Ausschnitt:

After completion of the calculations (duration depends on hardware approx. 5 – 30 seconds) the results will be displayed. The following illustration shows an excerpt:


Please use the scroll bar in order to inspect all the scale values. The meaning of the scales is specified in the appropriate chapter of this manual. The bar always refers to the percentile bands. The longer the grey bar is, the higher is the percentage rank of the participants, and the better these participants were in comparison to the norm group. The percentile rank indicates the number of the participant of the control group that was exactly as good or bad as the evaluated participant.

Mit dem dritten Knopf in der Buttonleiste können Sie sich das Leistungsprofil auf jedem handelsüblichen Drucker ausgeben lassen (Papierformat A4). Anschließend können Sie mit dem ersten Knopf die nächste Protokolldatei zur Auswertung anwählen.

With the third button in the toolbar, you can print out the performance profiles (A4 paper format). Afterwards, you can choose the next data file for analysis with the first button.
Auswertung: Einstellungen Analysis: Settings

Das Verhalten des Auswertungsprogramms läßt sich in einigen Details an die Bedürfnisse des Anwenders anpassen. Wählen Sie dazu in der Menüleiste "Auswerten" - "Einstellungen". Es erscheint die folgende Dialogbox:

The user can configure the behaviour of the analysis program. To this end, select “Analysis” – “Settings” in the menu list. It displays the following dialogue box:

![Dialogbox](image)


Under path for norm data set, adjust the path so the program searches in the file for the control group. Normally, the default path is already indicated here, when FSYS is installed. Only in rare cases will you have to choose this path.

Die Einstellung für den "Standardpfad für Protokolldateien" wird noch nicht benutzt. Ändern Sie bitte das Arbeitsverzeichnis des Icons, mit dem Sie das Auswertungsprogramm starten, wenn Sie den Standardpfad für die Protokolldatei-Dialogbox ändern möchten.

The settings for the “standard path for a data file” will not be used just yet. Please choose the working directory of the icons that you used to start the analysis program when you would like to change the standard path for the data file-dialogue box.

Die Einstellung "Ausgabepfad für Ergebnisdateien" legt fest, in welches Verzeichnis die Dateien "roh.txt" und "proz.txt" geschrieben werden.

The settings “Output path for Results data” determines in which directory the files “raw.txt” and “proz.txt” will be written.


The “decimal seperator” is in the German language area traditionally a comma. Therefore expect that a number in the position after the decimal place will be seperated by a comma in several data analysis programs. Other programs in the anglo-saxon language area expect the usual full stop.

Für die Darstellung der FSYS-Leistungsprofile verwendet das Auswertungsprogramm unabhängig von dieser Einstellung immer den Punkt.
For the display of the FSYS Performance profile use the analysis program independently from the settings menu.

Alle Einstellungen des Auswertungsprogramms (auch die zuletzt gewählte Norm und die Option zum Schreiben von Ergebnisdateien) werden in der Datei "f20aus.ini" im Windows-Verzeichnis gespeichert.

All settings in the analysis program (also the last selected norms and the option for writing the results file) will be saved in the file “f20aus.ini” in the Windows-directory.

**Automatische Auswertung Automatic Analysis**


For automatic analysis, the Analysis program will select many data files. A simple batch file (with the DOS command FOR) can be used to analyse all subjects in a specific list, without manual interaction with the PC being necessary. In so doing, all the settings that were last used in interactive mode will still hold. In automatic mode, no performance profile will be written, only both results files “roh.txt” and “proz.txt”.

Die folgende Batchdatei sollte unter den meisten Windows-Varianten funktionieren:

The following batch file should function under most Windows versions:

```
for %%f in (c:\fsys20\*.dhi) do f20aus %%f
```

Unter Windows™ NT läßt sich die Priorität des Auswertungsprogramms niedrig einstellen. Dadurch kann man nebenher noch gut mit dem Rechner arbeiten, während die Auswertungen laufen:

Under Windows NT the priorities of the analysis program can be set to low. In this way, one can work alongside the computer while the analysis is running:

```
for %%f in (c:\fsys20\*.dhi) do start /LOW /WAIT f20aus %%f
```

Falls Sie Probleme mit der Anpassung dieser Batchdateien haben, wenden Sie sich bitte an uns.

In case you encounter problems with the scripting of these Batch files, please feel free to contact us.
Skalen Scales

Das FSYS-Auswertungsprogramm liefert die Rohpunktwerte und Prozentränge der FSYS-Skalen. Diese Skalen sind gemäß des theoretischen Konzepts hinter FSYS in vier Gruppen untergliedert (S, M, I und E). Die Rohpunktwerte aller Skalen liegen zwischen 0 und 100. Sie dürfen aber nicht als Prozentränge interpretiert werden. Die Prozentränge beziehen sich immer auf eine wählbare Vergleichsgruppe.

The FSYS-Analysis program provides the raw item values and percentage range for the FSYS scales. These scales are in accordance with the theoretical concepts behind FSYS and are partitioned into 4 groups (S, M, I, and E). The raw item scores for all scales lie between 0 and 100. You may interpret these, but not as percentage bands. The percentage ranges always apply to a control/comparison group.

S: Gesamtergebnis   S: Total Score
Note: You'll need to come up with new appropriate acronyms for these scales (my suggestions in blue)
Oh, you're right!

Die S-Skalen beziehen sich auf die Steuerleistung des Probanden insgesamt. The S-Scale refers to the participant’s control? performance in total.

SKAP SKAP (SCAP) OK

Final status of Economic Capital. This scale is directly equivalent to the conceptual task (maximisation of total assets). A higher total value implies higher problem-solving ability. A value of about 50 raw points means that the assets of the company at the end of the session are just as high as they were at the beginning. Thus, participants with a score over 50 have made an economic gain.

SKAPKOR SKAPOR SCAPCOR

A version of SKAP will be calculated from the effects of individual coarse? control mistakes. If this value is clearly higher than SKAP, then the subject has ruined the business through not enough interventions of consequence. Their actual problem solving performance over the total time will therefore underestimate SKAP?.

M: Maßnahmenqualität M: Measure Quality
Die M-Skalen bewerten die Güte der einzelnen Maßnahmen, die der Proband angeordnet hat. Maßnahmen sind alle Arbeitsaufträge für die simulierten Arbeiter. Sie können unter verschiedenen Aspekten beurteilt werden.

The M-Scales estimate the quality or value of the individual variables that the subject has arranged. Variables are all the business orders for the simulated worker. You can assess the subject on various aspects.

MNOFEHL MNOFEHL MNOMIST OK
Ausmaß, in dem prinzipiell vermeidbare direkte Fehlsteuerungen tatsächlich vermieden werden. Vermeidbar sind Fehlsteuerungen, wenn entsprechende Informationen abrufbar

A measure of directly avoidable control mistakes (management errors?) that are actually avoided. Management errors are avoidable when the appropriate information was available. Example: Strong over- or underdosage of pesticides (sensible dosages can be learnt from the Information text).

Ein hoher Punktwert bedeutet, daß die Maßnahmen des Probanden immer gemäß Infotexten richtig waren. Das bedeutet nicht, daß die Maßnahmen auch in der jeweiligen Situation sinnvoll waren, oder daß sie den gewünschten Effekt erzielt haben.

A higher score means that the subjects have used the infotext correctly throughout. It does not mean that the variable was also sensible in the respective situation, or that the desired effect was achieved.

**MPRIORI MPRIORI MPRIORI OK**

Richtige Einordnung der Prioritäten von Teilzielen. Wichtigen Maßnahmen sollte der Vorzug vor unwichtigeren gegeben werden.

Correct prioritisation of target components. Important variables should be given precedence over unimportant ones.

Ein hoher Punktwert bedeutet, daß der Proband zu jedem Zeitpunkt die drängendsten Probleme seines Betriebs angegangen hat. Das läßt keine Rückschlüsse zu, ob die Probleme auch erfolgreich bewältigt wurden.

A higher score means that, at every point in time, the subject tackled the most urgent problems of their company. This does not allow for conclusions/inferences about whether the problems are solved successfully.

**MEFFIZI MEFFIZI MEFFICI MEFFECT**

Beurteilung der Maßnahmen hinsichtlich ihrer Effizienz zur Erreichung des jeweils vom Probanden angestrebten Teilziels.

Assessment of the participant’s efficiency effectiveness for completing orders with respect to the target. [A good time to clarify the meaning of this scale by improving its name. Sorry for the confusion, but effectiveness matches better, also in German where I already started using effectiveness instead of efficiency!]

Ein hoher Punktwert bedeutet, daß die Maßnahmen des Probanden immer die beabsichtigte Wirkung entfalteten. Das erlaubt keine Aussage über eventuelle, unerwünschte Nebenwirkungen, die in Kauf genommen wurden. Außerdem können auch unangemessene Maßnahmen an sich effizient sein – und umgekehrt.

A higher score means that the participant’s orders display the intended effects. This does not allow for evidence of potential undesirable side effects, which would be received from purchase. Furthermore, inadequate orders can also be efficient – and vice versa.

Beispielsweise ist eine Schädlingsbekämpfung effizient, wenn hinterher alle Schädlinge ausgerottet sind. Daß vielleicht auch die zu schützenden Bäume vernichtet wurden, wird nicht bewertet. Ebensowenig wird für MEFFIZI beachtet, ob es überhaupt nötig war, die Schädlinge zu bekämpfen.

For example, a pest control is efficient when afterwards all pests have been exterminated. If the protected trees were also damaged, this would not be valued. Likewise few for MEFFIZI will be observed, if it was not necessary at all, to combat the pests.
MVERSTA MVERSTA MCOMPRE OK

Gibt an, wie gut der Proband das Gesamtsystem bereits am Anfang der Simulation überblickt und seine Handlungsmöglichkeiten ausschöpft.

This is an indicator of how good the subject is at overviewing the total system initially at the beginning of the simulation and exhausting their management possibilities.

Hohe Punktwerte lassen auf ein gutes Verständnis der Instruktion und frühzeitigen Überblick über das Gesamtsystem schließen.

Higher scores indicate a good understanding of the instructions and an overview of the total system at an early time.

I: Informationsmanagement I: Information Management


The I-Scales make statements about the connection of the participant with the source of information. During the simulation the subject can use an Information database with descriptions and “Instructions of Use” for all objects in the scenarios. In addition, they can and should provide information regularly about the status of their department within the company. For this, it will stand next to the concrete numbers also graphics and tables to existing developments interesting variables at ones disposal.

IORIENT IORIENT


A high value means that the subject has already provided an overview of the system in the first simulated cycle. They have accessed the Information text and checked over the status of the forests. It is easy to see that far-reaching decisions should not be made without prior attainment of the necessary knowledge.

IVORHAND IVORHAND IFIRST OK

Hohe Werte bedeuten, daß der Proband sich vor der erstmaligen Anordnung einer bestimmten Maßnahme über die Wirkungen und Nebenwirkungen dieser Maßnahme sowie potentielle Alternativen informiert. Dazu ruft er die entsprechenden Informationstexte ab.

High values mean that the participant has provided information of a certain order, before the first time period. Information is given about the effects and side-effects of these orders as well as potential alternatives. To do this they access the appropriate information text.

IFEEDS IFEEDS

Hohe Werte bedeuten, daß der Proband nach ausgeführten Maßnahmen deren konkrete Wirkung mit Hilfe der Statistiken analysiert. Dadurch kann er das auf Basis der Informationstexte erworbene Wissen über die Wirkung der jeweiligen Maßnahme verifizieren und differenzieren.

High values mean that the subject analysed the concrete effects of the performed orders with help from the statistics. Thus, they can verify and differentiate acquired knowledge on the basis of the information text.
IKONTI IKONTI ICONTII OK
Hohe Werte bedeuten, daß der Proband kontinuierlich alle Abteilungen des Betriebs inspiziert hat. Dadurch hatte er die Möglichkeit, sich einen guten Überblick über den Zustand seines Betriebs zu verschaffen.

High values mean that the subject continually inspected all sections of the company. Thus, he has the opportunity to gain a good overview about the status of his company.

IKONTIS IKONTIS ICONTIS OK

Continuous control of developmental trends important system variables about statistics. The utilisation of trend information can help to identify critical developments in sufficient time.

IMODUS IMODUS IMODE OK

The preference of graphical versus numerical information when demanding statistics. High scores mean that the subject has mainly chosen the graphic presentation of statistics. Low scores mean that the tabular representation was preferred.

Wenn ein Proband nie die Statistiken benutzt hat, kann dieser Wert nicht berechnet werden und fehlt daher in der Ergebnisdatei.

When a subject has never used statistics, these values cannot be calculated and therefore this information is absent in the results data.

E: Selbstmanagement E: Self-management
Die E-Skalen liefern Anhaltspunkte über Aspekte des Selbstmanagements des Probanden. Darunter werden Indikatoren verstanden, die allgemeinere Facetten des Umgangs mit der Aufgabenstellung betreffen.

The E-Scales provide clues about aspects of the subject’s self-management. Below are the indicators, relating to the general facets and their connection to the task.

ESICHER ESICHER ECERTAIN ECERTN ?

High values mean that the subject does not go back on decisions already made within the same cycle. Subjects with low values are either indecisive or unsure about using the system.

EGLEICH EDIRECT EBALANC ?
Hohe Werte bedeuten, daß der Proband die verschiedenen Abteilungen des Betriebs ausgewogen häufig aufgesucht hat. Die Aussage dieser Skala entspricht teilweise IKONTI.

High values mean that the subject has frequently called upon the various sections of the company in good balance. The interpretation of this scale corresponds in part to ICONTI.

EWERT EVALUE OK
Hohe Werte bedeuten, daß sich der Proband häufig über die Vermögensentwicklung seines Betriebs informiert hat. Er hat sich also direkt am Hauptkriterium für seinen Erfolg orientiert.
High values mean that the subject was informed about the growth of the company’s assets. He also oriented himself directly with the main criteria for success.

**Hinweise zur Interpretation Background for Interpretation**

Speziell in Auswahlsituationen sollten nicht alle Skalen in gleicher Weise berücksichtigt werden. Einige FSYS-Skalen dienen eher Forschungszwecken und sollten nicht überinterpretiert werden. Teilweise können die Benennungen dieser Skalen in die Irre führen.

Not all scales should be considered in the same manner, particularly in selection contexts. Several FSYS scales primarily serve research purposes and should not be overinterpreted. In particular, the naming of these scales can be misleading.

Nur mit Vorbehalt interpretiert werden sollten MVERSTA, IMODUS und ESICHER. Außerdem ist zu beachten, daß EGLEICH und IKONTI von der Bedeutung her relativ eng zusammenhängen, ebenso IVORHAND und IORIENT.

The scales MCOMPRE, IMODE and ECERTAIN should be interpreted with caution. Furthermore, it must be pointed out that EDIRECT and ICONTI from their meaning have a relatively small correlation, similarly IFIRST and IORIENT.

Um die Gesamtleistung eines Probanden einzuschätzen, ist SKAPKOR nach den vorliegenden Befunden das "fairere" Maß als SKAP. Es hängt aber immer von der Untersuchungsintention ab, ob nicht doch SKAP zu bevorzugen ist.

In order to assess the total performance of a subject, given the findings presented, SKAPKOR is a “fairer” measure than SKAP. Of course, it always depends on the purpose of the investigation, whether or not SKAP is preferred.

**Anhang: Appendix:**
- Skript zum Ablauf der praktischen Einführung des Probanden
- Einführungstext zur Aushändigung für den Probanden
- Script to run the practical introduction for participants
- Introductory text to hand out to participants
Skript zum Ablauf der praktischen Einführung des Probanden

Script to Run the Practical Introduction for Participants


[Preceding this are establishing rapport, the handout, reading through the Introductory texts, and handing out of notepaper. Now read the following text slowly and clearly. It should be respected that all participants will be following your speech]

Bevor Sie mit dem eigentlichen Test beginnen, erhalten Sie die Gelegenheit, die Programmbedienung ohne Konsequenzen für Ihren unternehmerischen Erfolg kennenzulernen. Dazu starten wir nun das FSYS-Lernprogramm und führen die wichtigsten Bedienungsschritte gemeinsam durch. Halten Sie sich bei der Übung bitte an meine Anweisungen und handeln Sie noch nicht eigenmächtig.

Before you begin the actual test, you will be provided with the opportunity to familiarise yourself with the program without consequences for your entrepreneurial success.

[Das FSYS-Lernprogramm starten.] [Start the FSYS-Training Program]


You can ignore the text menu above. For the operation of the program the buttons are sufficient. The first five are for the selection of the forest; the sixth selects the total overview. This will be pressed in a moment. Information will follow afterwards for the Information database.


Please press the button for the 3rd forest. Now the status of this forest will be displayed. Trees are already standing there that you now want to lumber. Please press the button for Forest 1. Here we want to afforest.


Please press the yellow button "Action". It appears in the Orders Dialogue Box, and you should now press “Afforest”. We select “Evergreen”. You will now see the costs for these orders. Click “OK”. You will see underneath the message: “Workers afforest WS 1” does this mean South-West? [No, Forest Plot !]

Solange wir nicht auf den Knopf "Weiter" drücken, passiert aber noch gar nichts. Um in den nächsten Monat zu gelangen, drücken Sie jetzt bitte "Weiter". Beantworten Sie die Sicherheitsabfrage mit "OK". Unten rechts steht jetzt "2. Monat". Und die Arbeiter sind wieder unbeschäftigt. Die Bäume wurden gepflanzt (siehe Anzeige), und sie fingen an, zu wachsen.

As long as we do not press the button “continue”, nothing further will actually happen. In order to arrive at the next month, please now press “Continue”. Answer the Security question with “OK”. The bottom right now states “2. Month”. And the workers are again waiting for an order. The trees will be planted (see display), and will begin to grow.

Wenn Sie sich lieber erst informieren wollen, ehe Sie etwas tun, können Sie die Info-Abteilung nutzen. Klicken Sie bitte auf "Info", und dann auf "Aktion". Sie können jetzt einen Infotext wählen. Wir wählen "Kontostand" und klicken auf OK. Im Test würde Ihnen jetzt die entsprechende Information angezeigt.
If you would prefer to be informed first about what you can do, you can use the Info-department. Please click on “Info” and then “Action”. You can now select an infotext. We will choose “Bank Account” and click on OK. In the text you would now be shown the appropriate information.


This month our workers will lumber the trees in Forest 3. Please click on the button for Forest 3. There the trees can bring in revenue through sale, when they are thick enough to lumber. We click on “Action”, “Lumber” and “OK”.

Die Arbeiter können immer nur eine Sache machen. Sie müssen entscheiden, was Ihnen am wichtigsten ist. Für diesmal heißt das: Wir können jetzt nicht gleichzeitig ein weiteres Waldstück aufforsten. Probieren wir's: Klicken sie auf "Waldstück 4" und "Aktion". Alle Maßnahmenknöpfe sind deaktiviert. Rechts steht der aktuelle Arbeitsauftrag für die Arbeiter.

The workers can only ever make one business action. You must decide what is most important for you. For this time that means: Click on “Forest 4” and “Action”. All Order buttons are deactivated. The actual manufacturing orders for the workers are on the right.

Wir könnten hier das Abholzen von WS 3 zurücknehmen und stattdessen etwas neues anordnen, aber das machen wir jetzt nicht. Klicken Sie deshalb auf "Abbruch".

Here we could have taken back the lumbering of Forest 3 and instead order something new, but we won’t do that just yet. Therefore, click on “Cancel”.


We will once again move ahead one month. Please click on “Continue” and “OK”. This time the workers are still engaged with an order. That is to say, not all orders can be finished off within a month. Notice the percentage-range/scale for the work progress.


In order to review the development of your company, you can use statistics. Please click on “Forest 3” and “Statistics”. From the bottom left of the list we will choose “Average Wood Thickness”. Now it appears as a chart, to reproduce the development of the wood thickness in this forest. Every small line/stroke on the level (horizontal) axis stands for one month since the beginning of the simulation. The average wood thickness thus increases, i.e. the trees are growing. You can scroll through the selection list by using the small arrow on the right border of the list. Now use this button in order to let the “Reported Grass-Spider Infestation” be displayed. When you click on this entry a second time, the corresponding curve will again fade out/go downwards.

Wenn Sie exaktere Angaben wollen, können Sie von der Grafik auf Tabellendarstellung umschalten. Klicken Sie bitte auf "als Zahlen". Nun sehen Sie die Zahlenwerte. Wir verlassen die Statistik mit "Fertig". Da die Arbeiter noch beschäftigt sind, gehen wir auf "Weiter" und auf "OK".
When you want the exact specifications, you can change from the graphic to the table presentation. Please click on “as numbers”. Now you will see the numerical values. We leave the statistics with “finish”. The workers are still engaged; we use “Continue” and “OK”.

Damit ist die Einführung beendet. Wenn Sie noch Fragen zur Bedienung haben, können Sie diese jetzt oder später beim Test gerne stellen. Versuchen Sie bitte, die 50 Zyklen in der vorgesehen Zeit von 90 Minuten zu absolvieren. Unten rechts wird die Restzeit in Stunden und Minuten angezeigt.

With this the introduction is finished. If you have questions about the program operation, you can ask them now or later. Please try to complete the 50 cycles in the allocated time of 90 minutes. The time remaining will be displayed on the bottom right in hours and minutes.

Ich wünsche Ihnen viel Erfolg bei der Bearbeitung.

I wish you every success with the task.

[Test starten]
[Start Test]
Dear Participant,

Thank-you for participating in our "Forest" simulation. This short text should help to familiarise you with the simulation so that you can be a successful company manager.

What is being investigated?
The "Forest" simulation task allows for a prediction about your handling of complex situations. In complex situations you must always weigh up which method to adopt in order to find a solution. There is no one definitively correct solution. Rather, only more or less satisfactory possibilities for the circumstances. In this respect, the simulation is equivalent to many situations that people face in "everyday life".

While you work through the problem (the management of a forestry company in a fictitious future), your actions on the computer will automatically be recorded in full and your results later evaluated. We are not only interested in your ability to manage the company successfully (economically), but we will also analyse how you proceed in the individual phases of the business management, and which decisions you make.

How does this program function?
The simulation involves a forestry business with five forests that you have just taken over. You do not need to possess any knowledge at all about the forestry industry or business management, because the corresponding circumstances will be reproduced in the program in the simplest form possible. Moreover, trees that appear in the program do not exist in the real world and grow significantly faster. Also, there are other ecological details that do not correspond to reality.

Your Business Establishment

Important elements of your business are:

- **Forests:** In total, trees in five forests will be planted, cultivated and lumbered (timber is cut down). This process generates profit. The lumbered trees will automatically be sold, and this is the only source of income in the simulation. The five forests are all the same size, but your soil conditions can be different and are affected by the increase or loss of trees. Basically, the forests are always a monoculture, so in every forest there is invariably only one type of tree planted. The growth of the trees begins with afforestation (planting of seedlings). Later, more trees can be planted, but not before the previous generation of stock in the current forest is completely cut down.

- **Workers/Employees:** The work team is responsible for the planting, fertilisation, pest control and cutting down of the trees in the five forests.
• **Bank Account:** All costs (wages and material costs) will automatically be withdrawn from your bank account. Similarly, all income from wood sales will automatically be deposited into your account.

• **Information Data Bank:** Here you can learn about almost everything of significance for the management of the company.

**The Management of the Business**

You are the decision maker in your business. You will make your respective decisions monthly. When you are happy with your decisions, the business for the month will continue to be calculated. After a total of 50 months, the simulation will end. Please note that you will be given a maximum period of 90 minutes to work through the task. Please take care that you complete all 50 months in this time.

The long-term objective of your decisions should be an increase of the total assets of your company. The total assets calculate themselves from the value of the 5 forests (depending on the tree stock and on the quality of the soil) and the status of the bank account.

**What comes next?**

Your task is to successfully manage the forestry company and to increase the total assets of your business. An important criterion for the quality of your performance is the status of the total assets after the 50 simulated months.

The following tips should facilitate you in the introduction of your business:

- Always keep all five forests in mind. Use the Statistics in order to better assess the Trends. The graphic representation often provides a very quick overview.

- All important details can be researched in the menu "Info". Use this opportunity before you make decisions, in order to avoid potentially fatal consequences!

If you still have problems with the operation of the program after the Introduction, please consult the test leader.

Good luck with the task!
Overview of the Operation of the Program

*(A precise read-through here is not necessary – the text remains with you!)*

The following picture shows a typical screen from FSYS.

The window content displays either the status of a forest, the status of the total company, or an Information text at a time.

On the bottom right border you will see the actual simulation month (here Month 11), and the remaining test time (here one hour and 28 minutes). The big buttons at the top border are divided into 2 groups. The green buttons switch between the sections of the company (Forests, Total Business, and Information), and the yellow button initiates the actions. The button "Action" serves as the order of actions for your workers and for the selection of Infotext. The button "Statistics" allows for the analysis of previously simulated months. You must also always activate "Advance" when you have made the decisions for a month. The simulated time will not continue on its own.

Meaning of the Terms:

**Tree Type:** You must select from five types of trees for planting trees.

**Age of the tree populations:** Elapsed time since afforestation of the forests.

**Average wood thickness:** The sales revenue depends on the trunk width of the trees. If the soil is suited to the trunk type, then the trunk thickness will increase.

**Tree status:** After planting trees this stands at 100% as a general rule. Due to pests and mineral deficits in the soil, the trees die off, and their status decreases accordingly.

**Sales Revenue per %:** The amount that you would currently obtain per percent tree
status for the logging. For example, you can obtain

\[ 2663.74 \times 97.1\% = 258649. \]

**Pest Infestation:** The severity of the infestation is determined by the three kinds of pests. The severity is classified as weak, medium and strong. For example, if Soldier-Beetles mainly infest the trees, they will quickly reduce the tree population.

**Phosphate-Magnesium**

The soil content is made up of 5 minerals. For optimal growth, the trees require every type of mineral in a precise amount. Each tree variety has suggested mineral requirements.

If you selected the total company instead of a single forest, other statements are displayed:

**Value of the Forests**

For each specific forest plot, the direct sum of the value of the soil and the tree population.

**Fixed costs:**

Salary costs and taxes cannot be affected and cannot be altered during the simulation.

**Material costs:**

In the last month the accrued costs for seedlings, fertilizer, and pesticides.

**Revenue from wood sales:**

Profit through wood sale in the last month.

**Bank account:**

Actual bank account. Can be overdrawn by any arbitrary value.

**Total Assets:**

The sum of the value of the five forests and the bank account.

---

**Reading the Information Text**

Your Information department supplies you with all the facts that you need for the management of the business. This can be accessed using the button labelled "Info" (one click with the left mouse key). Then use "Action" and select the topic of interest. Afterwards the Infotext will be displayed as shown in the example picture.
If you want to return to the Information section later, the last selected text will automatically be displayed. By activating the "Action" button in the Information department you can select a new Infotext.

**Allocation of Job Orders**

In order to issue a job order to your workers, first select the forest in which they should work. Then activate the button "Action". It displays the following dialogue box.

On the left there is a button for every possible job order. Simply activate one to request these orders. When the Text appears as shaded (in the picture with the button "Regenerate"), then these actions can not be directly ordered in the relevant forest. Planting trees is, for example, not possible when trees are already situated in the forest.

On the right, you will be shown what your workers can do directly. A new work order can only be given out if they are unengaged (as they are here). Otherwise, you must either wait until the workers are finished or the actual work is suspended (for this purpose there is the button on the right in the middle).

Please remember that you cannot divide your workers. You can invariably only finish off one work assignment per month. Most work that is especially extensive can also engage workers for several months. This is primarily the case with logging.
Statistics

You can accurately analyse the development of the individual forests and the company at any time. According to the rules, you should first select the section of interest. Then activate the button "Statistics". It displays a dialogue box as in the following picture.

![Statistics Dialogue Box]

From the list on the bottom left you can choose any desired data (simply click with the left mouse key). The corresponding trend will then be displayed in the top section. You can then choose between graphic presentation and presentation of the exact numbers in table form.

Too many indexes at the same time will become confusing. By a repeated click of the selected data this will be deactivated. Please also note the scroll bar on the right side of the list with the data. By clicking on both arrows (upwards and downwards) you can choose the indicated domain in the list – namely there is more data, than is adjusted in this area. This technology will also be applied with the selection of the Info text, so that not all topics can be displayed at the same time. Please note the scroll bar and use it if required.
### Appendix H

**H Confirmatory Factor Analysis of Cognitive Abilities Test Battery**

Factor Loadings, Communalities ($h^2$), Percents of Variance for Confirmatory Factor Analysis and Oblique Rotation of Cognitive Abilities Tests

<table>
<thead>
<tr>
<th>Cognitive Abilities</th>
<th>Fluid Reasoning (Gf)</th>
<th>Crystallised Knowledge (Gc)</th>
<th>Broad Visualisation (Gv)</th>
<th>$h^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raven’s Matrices</td>
<td>.85</td>
<td>-.09</td>
<td>-.18</td>
<td>.61</td>
</tr>
<tr>
<td>Problem Solving</td>
<td>.88</td>
<td>.03</td>
<td>-.12</td>
<td>.72</td>
</tr>
<tr>
<td>Paper Folding</td>
<td>.71</td>
<td>.06</td>
<td>.06</td>
<td>.57</td>
</tr>
<tr>
<td>Financial Reasoning</td>
<td>.65</td>
<td>.12</td>
<td>.08</td>
<td>.53</td>
</tr>
<tr>
<td>Numerical Operations</td>
<td>.58</td>
<td>-.18</td>
<td>.28</td>
<td>.49</td>
</tr>
<tr>
<td>Esoteric Analogies</td>
<td>.55</td>
<td>.47</td>
<td>-.11</td>
<td>.48</td>
</tr>
<tr>
<td>Swaps</td>
<td>.54</td>
<td>.04</td>
<td>.22</td>
<td>.45</td>
</tr>
<tr>
<td>Letter Series</td>
<td>.51</td>
<td>-.07</td>
<td>.26</td>
<td>.44</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>-.14</td>
<td>.88</td>
<td>.06</td>
<td>.75</td>
</tr>
<tr>
<td>Proverbs Matching</td>
<td>-.04</td>
<td>.81</td>
<td>.11</td>
<td>.66</td>
</tr>
<tr>
<td>Critical Reasoning</td>
<td>.37</td>
<td>.68</td>
<td>-.12</td>
<td>.76</td>
</tr>
<tr>
<td>Letter Spotting</td>
<td>-.06</td>
<td>.07</td>
<td>.82</td>
<td>.66</td>
</tr>
<tr>
<td>Line Length</td>
<td>.07</td>
<td>-.03</td>
<td>.71</td>
<td>.54</td>
</tr>
</tbody>
</table>

| Eigenvalue | 3.96 | 1.78 | 1.02 |
| Percent of Variance | 35.97 | 16.15 | 9.29 |

Factor Correlation Matrix

<table>
<thead>
<tr>
<th>Fluid Reasoning (Gf)</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystallised Knowledge (Gc)</td>
<td>.28</td>
</tr>
<tr>
<td>Broad Visualisation (Gv)</td>
<td>.39</td>
</tr>
</tbody>
</table>

*Note.* Loadings > .40 are underlined.