Use of Performance Predictors in Visual Analytics

by

Petri François Vitiello

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of

Doctor of Philosophy of Loughborough University

(2013)

[Intentionally left blank]

Abstract

Visual analytics (VA) is a multi-disciplinary field with interactive visualisation at its core. This field has emerged out of the need to analyse an ever-increasing amount of heterogeneous, large and often unstructured information data-sets. The aims of visual analytics are to promote and assist analytic reasoning and generate insights from the information presented. Furthermore, the human is an essential element in this system, and understanding the perceptual and cognitive factors is key to progress in this field. This research focuses on understanding the benefits of interaction in terms of insight generation and the effects on mental effort required to generate them.

This investigation also explores the compounding effects individual differences have with interaction when analysing data to generate insights by investigating the individual differences in two sets: a psychometric measure set composed of Locus of Control, Self-Efficacy and Self-Acceptance taken from the International Personality Item Pool (IPIP); and a multimodal learning style (VARK) to investigate the sensorial preferences relationship to insight generation. The psychometric measures were selected on the basis of their use as predictors of performance. The learning style model selection was driven by the ability to separate learning preferences into different sensorial modes.

This thesis analyses interaction from an information visualisation perspective, where human interaction can occur in three main parts (Data Transformation, Visual Mapping and View Transformation) according to the reference model defined by Card *et al.* [1]. The research explores the latter two parts of the model, Visual Mapping and View Transformation interaction, by isolating interaction as an independent variable. The analysis of the benefits of interaction uses the aptitude-by-treatment interaction (ATI) methodology. The ATI approach adopted enabled the assessment of the performance gains in terms of insight generation by using a pre-defined set level of individual differences measures.

The experimental design consisted of two experiments that isolated interaction as an independent variable. Experiments examined the visual mapping, where participants interacted with visual structures via a data analysis task. The second experiment studied the view transformation, allowing participants to interact by changing the views in 'serious game'-based VA simulation problem solving task. Additionally, this second experiment used two different visual representations of the problem – 2D and 3D, in order to gauge the effects of visual representation on view transformation interaction.

The experiment involves 42 participants, divided into two groups: one interacting with information visualisation; the other performing an equivalent non-interactive task. The participants were assessed prior to the experiment on their individual differences, using an online questionnaire consisting of the three psychometric measures associated with performance together with a multimodal learning preference style assessment. Then using the ATI methodology, a group comparison was performed, comparing individual differences conjoint effect with the visual structure and view interaction.

This thesis establishes the benefits of interaction as generating more insights and increasing accuracy. Further, the results show significant conjoint effects between interaction and individual differences. Additionally, this research revealed a performance difference between 2D and 3D visual representation in the 'serious game' problem solving context.

Overall, this thesis provides tangible proof that both visual structure and view interaction are beneficial to VA in generating insights. It also strengthens the view that interaction with the problem-set improves understanding, and the number of insights gleaned into the problem. The results showed that view transformation using a 2D representation in a game environment outperformed a 3D representation in the context of gathering insights into the problem set. Finally, these findings show compounding effects between interactions and psychometric measures and learning styles for both the visual structure and view interaction. Thus, adding more dimensions in the use of individual differences as a performance predictor in VA and providing additional credibility and confidence in the possibility of better profiling and selecting high performing visual analysts.

Acknowledgments

This is the opportunity to thank and be grateful to all the people that have made this PhD journey possible and a very rewarding experience. First and foremost I am indebted to my wife Delia and my son Zanekin and daughter Raija for the many sacrifices they've made, to allow me the time to concentrate and complete this thesis.

Special thanks to Dr John O'Brien, with whom the many conversations on life, science and philosophy have help me keep my 'sanity' in the challenging moments of this PhD.

Finally, I like to thank my supervisor Professor Roy S. Kalawsky, for his support and patience during this PhD journey.

I would like to thank Neil Fleming for his support and permission to the use the VARK tool, and acknowledge as well as, Tableau Software and Valve Corporation for their assistance in providing the software required for my experiments.

The Engineering and Physical Sciences Research Council (EPSRC) funded this research.

[Intentionally left blank]

Table of Contents

ABSTRACT I

ACKNOWLEDGMENTS III

TABLE OF CONTENTS V

LIST OF TABLES IX

LIST OF FIGURES XIII

LIST OF ABBREVIATIONS XV

CHAPTER 1. INTRODUCTION 1

1.1 Background and Motivation 2

1.2 Key Concepts and Scope 4

1.2.1 Insight 4

1.2.2 Mental Effort 4

1.2.3 Interaction 5

1.2.4 Individual Differences 6

1.3 Research Objectives 8

1.4 Thesis Outline 11

CHAPTER 2. LITERATURE REVIEW 13

2.1 Insight 14

2.1.1 Definition of insight 14

2.1.2 Characteristics of Insight 14

2.1.3 Evaluating and Measuring Insight 18

2.1.4 Insight Provenance 22

2.1.5 Insight: a summative overview 24

2.2 Interaction 25

2.2.1 Interaction and insight 25

2.2.2 Interaction and Games 30

2.2.3 Interaction: a summative overview 31
2.3 Individual Differences 32
2.3.1 Individual Differences in Information Visualisation 32
2.3.2 Individual differences : a summative overview 35
2.4 Thesis Objectives and Research Questions 37
2.4.1 Objective One 37
2.4.2 Objective Two 38
2.4.3 Objective Three 38
2.4.4 Objective Four 39
CHAPTER 3. METHOD 41
3.1 Introduction 42
3.1.1 Individual differences 42
3.1.1 Aptitude-by-Treatment Interaction 45
3.1.2 Workload Assessment 46
3.2 Experimental Design Overview 49
3.2.1 Participants 49
3.2.2 Design 49
3.2.1 Pre-study 51
3.3 VMI Experiment 53
3.3.1 Dependent Variables 54
3.3.2 Experiment Setting 58
3.3.1 Experiment Procedure 62
3.4 VTI Experiment 66
3.4.1 Dependent Variables 67
3.4.1 Experiment Setting 68
3.4.2 Experiment Procedure 70
3.5 Statistical Analysis 74
3.5.1 Power Analysis 78
CHAPTER 4. POPULATION SAMPLE CHARACTERISTICS AND PREPARATION 81
4.1 Participants 82
4.1.1 Description 82
4.1.2 2D and 3D Previous Experience 82
4.1.3 Individual Differences 83

4.1.4 ATI Methodology Profiles 87
4.2 VMI Experiment Data Preparation 90
4.2.1 Insights Categorisation 90
4.2.2 Mental Effort Based Insights Grouping 91
4.3 VTI Experiment Data Preparations 93
4.3.1 Post-Experiment Questionnaire 93
CHAPTER 5. VISUAL MAPPING INTERACTION EXPERIMENT 95
5.1 Statistical Analysis 96
5.2 NASA-TLX Analysis 97
5.3 Interaction Effects Analysis 98
5.4 Individual Differences and Interaction Compounding Effects Analysis 102
5.4.1 Overall Metrics Analysis 102
5.4.1 Yield Analysis 107
5.5 Summary 116
CHAPTER 6. VIEW TRANSFORMATION INTERACTION EXPERIMENT 119
6.1 Statistical Analysis 120
6.2 NASA-TLX Analysis 121
6.3 Interaction and Individual Differences Compounding Effects Analysis 124
6.3.1 Overall Percentage Analysis 124
6.3.2 Overall Score 126
6.4 Representation Analysis 127
6.4.1 Normalised Percentage Analysis 127
6.4.1 Score Analysis 131
6.5 Summary 138
CHAPTER 7. DISCUSSION 141
7.1 Research Summary 142
7.2 Findings 143
7.2.1 Objective One 143
7.2.2 Objective Two 145
7.2.1 Objective Three 149
7.2.1 Objective Four 152
7.3 Summary 158

CHAPTER 8. CONCLUSIONS 159
8.1 Conclusion 160
8.2 Contributions 162
8.3 Future Research 166
REFERENCES 169
APPENDICES 179
Appendix 1 Consent Form 181
Appendix 2 Participant Information Sheet 182
Appendix 3 Count Down Sheet 183
Appendix 4 Tutorial Insight Grid 184
Appendix 5 Empty Insight Grid 185
Appendix 6 Pre-Study Questionnaire 186
Appendix 7 Locus of Control Section of Pre-Study Questionnaire 187
Appendix 8 VARK Section of Pre-Study Questionnaire 188
Appendix 9 Self-Efficacy and Self-Acceptance Section of Pre-Study Questionnaire192
Appendix 10 VARK Scoring 193
Appendix 11 NASA-TLX 195
Appendix 12 VTI – 2D Representation Post Experiment Questionnaire 197
Appendix 13 VTI – 3D Representation Post Experiment Questionnaire 203
Appendix 14 VMI Statistical Description and Results Tables 208
Appendix 15 VTI Statistical Description and Results Tables 230

List of Tables

TABLE 1.1 – MENTAL EFFORT CATEGORIES DETAIL 5

TABLE 2.1. – INSIGHT FEATURES AND PERCENTAGE INSIGHTS PER FEATURE – ADAPTED FROM [35] 18

TABLE 2.2. – COMPARISON OF EMPIRICAL RESULTS, (ADAPTED FROM [48]) 21

TABLE 2.3. – SUMMARY OF BENCHMARK TASK AND INSIGHT METHODS [48] 22

TABLE 2.4. – SUMMARY OF GREEN AND FISHER RESULTS [9–11] 33

TABLE 3.1 – NASA-TLX RATING SCALE DEFINITIONS ADAPTED FROM [99] 48

TABLE 3.2 – EXPERIMENTAL PROCEDURE 52

TABLE 3.3 – INSIGHT CATEGORIES 55

TABLE 3.4 – MENTAL EFFORT CATEGORY GROUPING 56

TABLE 3.5 – ORDER OF VIEW TRANSFORMATION EXPERIMENTS 2D AND 3D REPRESENTATION 70

TABLE 3.6 – RESEARCH OBJECTIVES AND QUESTION WITH ASSOCIATED STATISTICAL ANALYSIS 77

TABLE 4.1 – INDIVIDUAL DIFFERENCES DESCRIPTION STATISTICS 83

TABLE 4.2 – VARK PROFILING EXAMPLE 84

TABLE 4.3 – VARK PROFILES 85

TABLE 4.4 – VARK PREFERENCE FREQUENCIES 85

TABLE 4.5 – SIMPLIFIED VARK PREFERENCE FREQUENCIES 86

TABLE 4.6 – INDIVIDUAL DIFFERENCES ATI CATEGORIES PARTICIPANTS NUMBERS (N) 88

TABLE 4.7 – INDIVIDUAL DIFFERENCES – LOW, MID AND HIGH GROUPING 89

TABLE 4.8 – NUMBER OF PARTICIPANTS BY LOW, MID AND HIGH GROUPING BY INDIVIDUAL DIFFERENCE 89

TABLE 4.9 – INDIVIDUAL DIFFERENCES PEARSON'S CORRELATIONS 89

TABLE 4.10 – INSIGHT CATEGORIES 90

TABLE 4.11 – MENTAL EFFORT CATEGORY GROUPING 91

TABLE 4.12 – MENTAL EFFORT CATEGORY GROUPING DETAIL 92

TABLE 4.13 – POST-EXPERIMENT 2D-3D ANSWERS CODING 93

TABLE 5.1 – TOTAL ACCURATE, TOTAL PROCEDURAL AND OVERALL NUMBER OF INSIGHTS PEARSON'S CORRELATIONS RESULTS 101

Table 6.1 – TLX Measures Statistic Description 122
Table 6.2 – TLX Measures Univariate ANOVA Results 123
Table 6.3 – Box's Test of Equality of Covariance Matrices Results for the Accurate, Inaccurate and Unidentified Percentages...128
Table 6.4 – 2D and 3D Score Statistical Description by Interactivity Treatment.....................138
Table 7.1 – Oneway ANOVA Results for VMI 144
Table 7.2 – Visual Analytics NASA-TLX Scores 148
Table 7.3 – 2D and 3D Scores Statistical Description by Interaction Treatment.....................155
Table VMI.1 – Number of Insights Statistical Description 208
Table VMI.2 – Omnibus Oneway ANOVA Results – Number of Insights, Accuracy and Mental Effort 209
Table VMI.3 – Overall Score Statistical Description by ATI-C and Interactivity Group 210
Table VMI.4 – Factorial ANOVA Results for the Overall score..211
Table VMI.5 – Overall Mental Effort Yield Statistical Description by ATI-C and Interactivity Group..212
Table VMI.6 – Overall Accuracy Yield Statistical Description by ATI-C and Interactivity Group 213
Table VMI.7 – Factorial ANOVA Results for the Overall Accuracy and Mental Effort Yield214
Table VMI.8 – Transformed Accuracy Yield Statistical Description by Mental Effort, Interactivity Group and Psychometric Measures ATI-C 215
Table VMI.9 – Transformed Accuracy Yield Statistical Description by Mental Effort, Interactivity Group and Learning Profiles ATI-C 216
Table VMI.10 – Univariate Factorial ANOVA Results for the Accuracy Yield 217
Table VMI.11 – Kruskal-Wallis Test Results for the Mental Effort Yield by Interactivity Treatment and by Accuracy and Interactivity Group as the Between Group Factor 218
Table VMI.12 – Mental Effort Yield Statistical Description by Accuracy, Interactivity Group and Psychometric Measures Scales 219
Table VMI.13 – Mental Effort Yield Statistical Description by Accuracy, Interactivity Group and Learning Profile Scales 220
Table VMI.14 (Part 1) – Kruskal-Wallis Test Results for the Mental Effort Yield by Accuracy Factor, Learning Preference, and Interactivity Treatment as the Between Group Factor 222

TABLE VMI.15 (PART 1) – KRUSKAL-WALLIS TEST RESULTS FOR THE MENTAL EFFORT YIELD BY ACCURACY FACTOR, LEARNING PREFERENCE, AND INTERACTIVITY TREATMENT AS THE BETWEEN GROUP FACTOR 224

TABLE VMI.16 – KRUSKAL-WALLIS TEST RESULTS FOR THE MENTAL EFFORT YIELD BY PSYCHOMETRIC MEASURE AND INTERACTIVITY GROUP AS THE BETWEEN GROUP FACTOR........ 226

TABLE VMI.17 – KRUSKAL-WALLIS TEST RESULTS FOR THE MENTAL EFFORT YIELD BY LEARNING PREFERENCE AND INTERACTIVITY GROUP AS THE BETWEEN GROUP FACTOR 227

TABLE VMI.18 – KRUSKAL-WALLIS TEST RESULTS FOR THE MENTAL EFFORT YIELD BY ACCURACY FACTOR,PSYCHOMETRIC MEASURE, AND INTERACTIVITY TREATMENT AS THE BETWEEN GROUP FACTOR 228

TABLE VMI.19 – KRUSKAL-WALLIS TEST RESULTS FOR THE MENTAL EFFORT YIELD BY INTERACTIVITY TREATMENT, PSYCHOMETRIC MEASURE, AND ACCURACY AS THE BETWEEN GROUP FACTOR........ 229

TABLE VTI.20 – OVERALL PERCENTAGE OF (ACCURATE, INACCURATE AND UNIDENTIFIED) INSIGHTS STATISTICAL DESCRIPTION BY INTERACTIVITY TREATMENT 230

TABLE VTI.21 – TRANSFORMED ACCURACY STATISTICAL DESCRIPTION BY INTERACTIVITY GROUP AND PSYCHOMETRIC MEASURES ATI-C........ 231

TABLE VTI.22 – TRANSFORMED ACCURACY STATISTICAL DESCRIPTION BY INTERACTIVITY GROUP AND LEARNING PROFILES ATI-C 232

TABLE VTI.23 – UNIVARIATE ANOVA RESULTS FOR THE OVERALL PERCENTAGE OF (ACCURATE, INACCURATE AND UNIDENTIFIED) INSIGHTS 233

TABLE VTI.24 – UNIVARIATE ANOVA RESULTS FOR THE OVERALL SCORE BY INTERACTION TREATMENT AND ATI-C........ 233

TABLE VTI.25 – OVERALL SCORE STATISTICAL DESCRIPTION BY INTERACTIVITY TREATMENT AND INDIVIDUAL DIFFERENCES ATI-C 234

TABLE VTI.26 – PERCENTAGE OF ACCURATE, INACCURATE AND UNIDENTIFIED INSIGHTS STATISTICAL DESCRIPTION BY REPRESENTATION AND INTERACTIVITY TREATMENT........ 234

TABLE VTI.27 – REPEATED MEASURES FACTORIAL ANOVA RESULTS FOR THE ACCURATE, INACCURATE AND UNIDENTIFIED PERCENTAGES WITH REPRESENTATION AS THE WITHIN SUBJECT FACTOR AND INTERACTIVITY GROUP AND INDIVIDUAL DIFFERENCES ATI-C AS BETWEEN SUBJECT FACTOR 235

TABLE VTI.28 – 2D AND 3D SCORE STATISTICAL DESCRIPTION BY INTERACTIVITY TREATMENT 236

TABLE VTI.29 – KRUSKAL-WALLIS TEST AND FRIEDMAN'S ANOVA RESULTS FOR POST-EXPERIMENT QUESTIONNAIRE SCORES BY REPRESENTATION AND INTERACTIVITY TREATMENT 236

TABLE VTI.30 – 2D AND 3D SCORES STATISTICAL DESCRIPTION BY INTERACTIVITY TREATMENT AND PSYCHOMETRIC MEASURES ATI-C....237

TABLE VTI.31 – MANN-WHITNEY RESULTS FOR THE 2D AND 3D SCORES BY PSYCHOMETRIC MEASURES ATI-C AND INTERACTION TREATMENT AS BETWEEN GROUP FACTOR....238

TABLE VTI.32 – 2D AND 3D SCORES STATISTICAL DESCRIPTION BY INTERACTIVITY TREATMENT AND LEARNING PREFERENCE ATI-C....239

TABLE VTI.33 – MANN-WHITNEY RESULTS FOR THE 2D AND 3D SCORES BY LEARNING PREFERENCES ATI-C (V) AND ATI-C (A) AND INTERACTION TREATMENT AS BETWEEN GROUP FACTOR....240

TABLE VTI.34 – MANN-WHITNEY RESULTS FOR THE 2D AND 3D SCORES BY LEARNING PREFERENCES ATI-C (R) AND ATI-C (K) AND INTERACTION TREATMENT AS BETWEEN GROUP FACTOR....241

List of Figures

FIGURE 1.1 – INFORMATION VISUALISATION REFERENCE MODEL (ADAPTED FROM [1]) 5
FIGURE 2.1 – FRAMEWORK OF INTERACTION COSTS ADAPTED FROM [57] 28
FIGURE 3.1 – ATI TYPES 46
FIGURE 3.2 – EXPERIMENTAL DESIGN OVERVIEW 50
FIGURE 3.3 – VMI ELEMENTS OF THE INFORMATION VISUALISATION REFERENCE MODEL 53
FIGURE 3.4 – SCREENSHOT OF TABLEAU READER – MERCER ISLAND DATA-SET 60
FIGURE 3.5 – NON-INTERACTIVE PDF SAMPLES – MERCER ISLAND DATA-SET 62
FIGURE 3.6 – VTI ELEMENTS OF THE INFORMATION VISUALISATION REFERENCE MODEL 66
FIGURE 3.7 – PORTAL ILLUSTRATION 69
FIGURE 3.8 – SCREENSHOT OF 2D PORTAL GAME 70
FIGURE 3.9 – SCREENSHOT 3D PORTAL GAME 71
FIGURE 3.10 – SAMPLE 2D AND 3D POST-EXPERIMENT QUESTIONNAIRE 72
FIGURE 3.11 – STATISTICAL ANALYSIS DECISION TREE FOR ONE INDEPENDENT VARIABLE ADAPTED FROM [105] 74
FIGURE 3.12 – STATISTICAL ANALYSIS DECISION TREE FOR MORE THAN ONE INDEPENDENT VARIABLE ADAPTED FROM [105] 75
FIGURE 4.1 – VARK PREFERENCES DISTRIBUTIONS 86
FIGURE 4.2 – SIMPLIFIED VARK PREFERENCES DISTRIBUTIONS 87
FIGURE 5.1 – TOTAL NUMBER OF INSIGHTS BY INTERACTION GROUPING 100
FIGURE 5.2 – TOTAL NUMBER OF ACCURATE INSIGHTS BY INTERACTION GROUPING 100
FIGURE 5.3 – TOTAL NUMBER OF ACCURATE PROCEDURAL INSIGHTS BY INTERACTION GROUPING 100
FIGURE 5.4 – OVERALL SCORE BY INTERACTION GROUPING * ATI-C (R) 103
FIGURE 5.5 – OVERALL SCORE BY INTERACTION GROUPING * R SCALE 104
FIGURE 5.6 – OVERALL MEAN MENTAL EFFORT SCORE BY INTERACTION GROUPING * ATI-C (SE) 105
FIGURE 5.7 – TOTAL PROCEDURAL AND INFERENTIAL INSIGHTS BY INTERACTION GROUPING * ATI-C (SE) 105

FIGURE 5.8 – OVERALL MEAN ACCURACY YIELD BY INTERACTION GROUPING * ATI-C (V) 106
FIGURE 5.9 – TOTAL ACCURATE AND INACCURATE INSIGHTS BY INTERACTION GROUPING * ATI-C (V)....106
FIGURE 5.10 – OVERALL ACCURACY YIELD BY INTERACTION GROUPING * V SCALE..............................106
FIGURE 5.11 – MEAN TRANSFORMED ACCURACY YIELD BY INTERACTION GROUPING * ATI-C (V) 108
FIGURE 5.12 – MENTAL EFFORT YIELD BY ACCURACY FACTOR, SELF-EFFICACY AND INTERACTIVITY TREATMENT 110
FIGURE 5.13 – ACCURATE AND INACCURATE INSIGHTS BY MENTAL EFFORT, SELF-EFFICACY AND INTERACTIVITY TREATMENT 111
FIGURE 5.14 – MENTAL EFFORT YIELD BY, INTERACTIVITY TREATMENT, READ-WRITE PROFILE AND ACCURACY FACTOR....................112
FIGURE 5.15 – ACCURATE AND INACCURATE BY MENTAL EFFORT, READ-WRITE-SCALE AND INTERACTION TREATMENT 113
FIGURE 5.16 – MENTAL EFFORT YIELD BY, INTERACTIVITY TREATMENT, KINAESTHETIC PROFILE AND ACCURACY FACTOR....................114
FIGURE 5.17 – ACCURATE AND INACCURATE BY MENTAL EFFORT, KINAESTHETIC-SCALE AND INTERACTION TREATMENT 115
FIGURE 6.1 – TLX MEASURES: 2D / 3D REPRESENTATION BY INTERACTION GROUPING.......................122
FIGURE 6.2 – ACCURATE AND INACCURATE OVERALL PERCENTAGE BY INTERACTION GROUPING 125
FIGURE 6.3 – OVERALL SCORE BY INTERACTIVITY TREATMENT 126
FIGURE 6.4 – ACCURATE AND INACCURATE PERCENTAGE BY REPRESENTATION 128
FIGURE 6.5 – ACCURATE PERCENTAGE BY REPRESENTATION AND INTERACTIVITY TREATMENT...............129
FIGURE 6.6 – UNIDENTIFIED PERCENTAGE BY REPRESENTATION AND INTERACTIVITY TREATMENT 130
FIGURE 6.7 – 2D AND 3D SCORES BY INTERACTIVITY TREATMENT..132
FIGURE 6.8 – 3D SCORES BY LOCUS OF CONTROL ATI-C AND INTERACTIVITY TREATMENT.....................132
FIGURE 6.9 – 2D AND 3D SCORES BY SELF-EFFICACY ATI-C AND INTERACTIVITY TREATMENT................133
FIGURE 6.10 – 2D AND 3D SCORES BY SELF-ACCEPTANCE ATI-C AND INTERACTIVITY TREATMENT 134
FIGURE 6.11 – 2D AND 3D SCORES BY VISUAL ATI-C AND INTERACTIVITY TREATMENT.......................135
FIGURE 6.12 – 2D AND 3D SCORES BY AURAL ATI-C AND INTERACTIVITY TREATMENT 136
FIGURE 6.13 – 2D AND 3D SCORES BY READ-WRITE ATI-C AND INTERACTIVITY TREATMENT 137
FIGURE 6.14 – 2D AND 3D SCORES BY KINAESTHETIC ATI-C AND INTERACTIVITY TREATMENT 137

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
A	Aural Preference
ANOVA	Analysis of Variance
AT	Activity Theory
ATI	Aptitude-by-Interaction
ATI-C	Aptitude-by-Interaction Categorisation
CDA	Collaborative Data Analysis
CPI	California Psychological Inventory
DCog	Distributed Cognition
DD-LSI	Dunn & Dunn Learning Style Index
DV	Dependent Variable
EEG	Electroencephalogram
F	F-Value
fMRI	Functional Magnetic Resonance Imaging
FS-ILS	Felder-Silverman Index of Learning Styles
GSD	Gregorc Style Delineator

HCI	Human-Computer-Interaction
IBM	Industrial Business Machines
IG	Interactive Group
IPIP	International Personality Item Pool
J	Jonckheere's test value
K	Kinaesthetic Preference
K-LSI	Kolb Learning Style Indicator
K-S	Kolmogorov-Smirnov
LCD	Liquid Crystal Display
LoC	Locus of Control
M	Mean
MANOVA	Multivariate Analysis of Variance
Mdn	Median
MR	Mean Rank
NASA	North American Space Agency
NIG	Non-Interactive Group
p	p-value
PDF	Portable Document File
R	Read-Write Preference
r	Pearson Correlation Coefficient
R&D	Research and Development

SA	Self-Acceptance
SD	Standard Deviation
SE	Self-Efficacy
Std. E	Standard Error
SPSS	Statistical Package for the Social Sciences
SR	Sum of Ranks
STM	Short Term Memory
SWAT	Subjective Workload Assessment Technique
TLX	Task Load Index
U	Mann-Whitney U-value
USA	United States of America
V	Visual Preference
VA	Visual Analytics
VARK	Visual, Aural, Read-Write, and Kinaesthetic
VDAR	Visual Data Analysis and Reasoning
VMI	Visual Mapping Interaction
VR	Virtual Reality
VTI	View Transformation Interaction
z	z-value
α	Cronbach's alpha
χ^2	Chi square

[Intentionally left blank]

Chapter 1. Introduction

This chapter introduces the research objectives of this thesis and the background to the subject matter. It defines the key concepts underlying this investigation and the thesis structure.

1.1 Background and Motivation 2

1.2 Key Concepts and Scope 4

1.2.1 Insight 4

1.2.2 Mental Effort 4

1.2.3 Interaction 5

1.2.4 Individual Differences 6

1.3 Research Objectives 8

1.4 Thesis Outline 11

1.1 Background and Motivation

In today's world, the amount of data available for analysis is increasing exponentially [2]. Data in their raw format have limited value; what is more valuable is the insight that can be extracted from these. Visual analytics (VA) stems from the visualisation and data analysis fields. Visualisation tools and techniques provide clarity through visual representations, and data analysis methods and algorithms help categorise and cluster data, to highlight otherwise hidden patterns. The first uses of interactive visualisation came from the data analysis field led by Tukey [3] with a shift from confirmatory analysis using static graphics to *Exploratory Data Analysis*, which is focused on interaction with the data. Then Card *et al.* [1] continued this process of interdisciplinary integration from the information visualisation field by depicting the use of vision to think from an information visualisation perspective. Interest in this new combined field has been on the increase ever since the term 'Visual Analytics' was first coined in 2004 [4]. The aims VA are to make this deluge of data an opportunity to synergise the strengths of computers and humans to make sense of 'big data'. Initial interest was generated in homeland security and emergency services, where VA has been used to investigate complex heterogeneous data-sets in the context of threats and rapid response options analysis. Within this context, Thomas and Cook's [5; p.4], define VA as *" … the science of analytical reasoning facilitated by interactive visual interfaces"*.

The guiding process in VA is a synergy between interactive visualisation and automated analysis of the data. In this context Keim *et al.* [6] define the exploration process as 'Analyse first, show the important, zoom/filter, analyse further, details on demand'. This practice is focused on the analysis of the data before their visual representation, and through interaction generating insights.

Scientific visualisation and information visualisation have been used for some time now to gain insights into complex data-sets. Upson *et al.* [7] have defined a framework describing the computational environment for scientific visualisations in which the analysis cycle is a key step. In this framework, scientists create and adapt data models based on their research and analysis cycles and then the outcome is used in the visualisation cycles. This human in the loop process is performed outside the visualisation tool, whereas in VA the human interactive analysis using visualisation is an integral part of the tool set.

The VA research and development agenda [5] calls for a 'science of interaction'. Pike at al [8] in their latest review regarding the progress of this 'science', identified many areas of development. One of the areas of research requiring attention relates to the evaluation of interaction, which is the key characteristic that this thesis investigates. Additionally, recent research [9–12] into the characteristics of the user population affecting the use of VA, identified the need to further explore interaction as an experimental variable. Investigations into these different human characteristics are referred to as 'individual differences studies'. Moreover, when considering interaction as *"… the process of active discourse of user with the data"* [13; p.18], the research on individual differences helps uncover the relationship between personality factors and learning styles with interaction performance. Further, as the main purpose of visualisation is insight [5, 14], this is the key dependent variable in these investigations.

This thesis, builds and expands upon the information visualisation research done by Green and Fisher [9–11] and Ziemkiewicz *et al.* [12], which focuses on the relationships between individual differences and visualisation structures and layout. Both these studies are insight-based investigations, and were performed in a controlled experimental setting. This thesis aims to expand the existing body of knowledge regarding the effects of individual differences in VA, by investigating not previously researched performance-related psychometric measures, such as self-efficacy and self-acceptance, as well as the effects of sensorial multimodal learning styles. In addition to examining interaction as the key independent variable of study, thus providing insight-based results with regards to the benefits of interaction in VA.

1.2 Key Concepts and Scope

1.2.1 Insight

The literature review chapter details the current definitions of insight, its measurement, characteristics and provenance in the context of VA. Starting with a broader view, the Oxford English Dictionary [15] defines insight as "*Internal sight, mental vision or perception, discernment; in early use sometimes, Understanding, intelligence, wisdom*". In this thesis, the insight definition used will concentrate on the latter part of the definition, centred on understanding.

As insight is the main purpose of visualisation, insight-based evaluations have become a standard approach in VA [16]. Insight-based evaluations are distinct from other quantitative evaluation and present measurement and characterisation challenges, for which other than absolute counting and timing no consensual agreement has been reached thus far. With regards to the classification of insights, a multitude of taxonomies exist, these are derived from both the insight characteristics and their provenance. Consequently, this research characterises insights in its simplest form and at the simplest level of interpretation, using counting as the data collection and provenance as categorisation mechanisms. Where provenance is defined as the process and rational by which insight is derived.

1.2.2 Mental Effort

Research studies [9–12] have investigated iterative procedural versus inferential learning using different visualisation interfaces. In these kinds of studies procedural tasks are considered from a bottom up perspective as a task requiring little conscious mental effort, due to their automatic and repetitive nature. Inferential tasks are about drawing conclusions from the data that require more conscious mental effort, and are used in reasoning activities such as induction, deduction, and comparison [17]. Using these definitions, insights were grouped into two categories based on the conscious mental effort required to obtain them. They are defined as procedural for the low mental effort insights and inferential for the high mental effort insights. Table 1.1 illustrates the kind of tasks involved in the different mental effort categories. The method section covers the mental effort categorisation in more detail.

ID	Procedural	ID	Inferential
1	Retrieve value	5	Sort
2	Filter	7	Characterise Distribution
3	Compute Derived Value	8	Find Anomalies
4	Find Extremum	9	Cluster
6	Determine Range	10	Correlate

Table 1.1 – Mental Effort Categories Detail

1.2.3 Interaction

Card *et al.* [1] define a reference model for information visualisation, depicted in Figure 1.1. In this model the raw data are transformed into data tables which are a description of the raw data though the use of meta-data. The next step in the process is the transformation of the data tables into visual structures through visual mappings that transform the data into graphical representations. Lastly, the visual structures have different views, which are created though view transformations that provide graphical parameters such as scaling and position. Human interaction occurs at three points in this model: in the Data Transformation, Visual Mapping, and View Transformation.

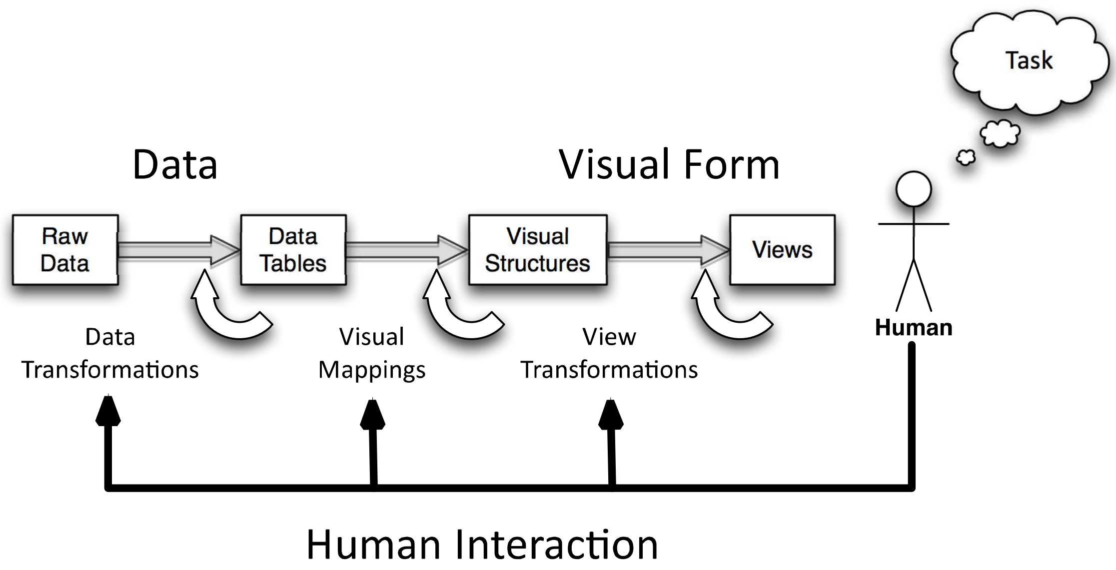


Figure 1.1 – Information Visualisation Reference Model (Adapted from [1]**)**

In this research, interactivity was studied from the visual mapping and the view transformation perspective as these are the key elements of interactivity in VA. In the visual mapping interaction (VMI), users interact with the visual structures in order to perform their exploration in a data analysis task. With regards to the view transformation interaction (VTI), users change the views within a game environment, used as a proxy to a spatio-temporal VA simulation problem-solving task as defined by Keim *et al.* [18].

Games are highly visual and interactive. Salem and Zimmerman [19] write comprehensively about the use of games as a system proxy, in particular about games as emergent systems. In this investigation the definition of the concept of emergence is taken from the Oxford English Dictionary as "*An effect produced by a combination of several causes, but not capable of being regarded as the sum of their individual effects. Opposed to resultant.*" [20]. The particular facets of interest in this thesis are the interaction aspect of games, and their use in exploring emergent system behaviours to understand a problem-solving data-set and extract insights. Hence in this study, the aspects of view transformations were studied using games in both 2D and 3D representations to understand the possible differences these two have when presented to the population sample.

1.2.4 Individual Differences

HCI has a long history of accounting for individual differences in user analysis research [21]. The individual differences investigated in this research are twofold. Firstly, performance-related psychometric measures, of which three were selected based on prior visualisation research. Secondly, a sensory based learning style preference model based on the differentiation of the verbal and non-verbal sensory modes as defined by Paivio [22].

The performance-related psychometric measures studied in this thesis are: Rotter's Locus of Control (LoC) [23], which is a measure that suggests how much a person attributes outcomes of actions to their own behaviour. (LoC is defined as a continuum going from internal to external LoC, depending whether the control is perceived internally or externally to the individual [24]); Bandura's self-efficacy (SE) measure [25], which measures the attitude towards goals and challenges (a high level is an indicator of high self-belief of good performance); and Self-acceptance (SA) [26] which is a measure concerned with confidence in personal decision-making and asserting one's own viewpoint, also associated with more resiliency to stress and higher effectiveness at work.

The sensory based learning style used is the VARK model, which profiles participants learning predispositions in terms of visual, aural, read-write and kinaesthetic sensory modes [27]. The different sensory modes are defined as:

- Visual preference, which refers to information presented in a symbolic manner without words such as graphs and charts (analogous to non-verbal processes in Paivio dual coding theory [22]);
- Aural preference, which relates to information presented by auditory means;
- Read-write preference, used to categorise people who absorb information in written format other than graphs and charts (analogous to verbal process in Paivio dual coding theory [22]); and
- Kinaesthetic preference, which relates to a practical learning preference.

1.3 Research Objectives

The VA agenda [5] calls for a science of interaction outlining in particular the need to look at the nature of interaction. Pike *et al.* [8] in their state of the art analysis, call for further research into the capacity of interaction as the inquiry process to generate knowledge. Research into the 'personal equation of interaction' [9–11] investigate the relationship between individual differences and interactive visualisation structures and call for more in-depth research into the value of interaction and the relationship with learning styles.

As a result, the research in this thesis will study interaction as an independent variable in the context of VA, in order to evaluate its benefits in terms of insight generation. Also this research will investigate the compounding effects individual differences have with interaction within this context.

Objective One: Investigate the effects of Visual Mapping Interaction in the context of performing an analytical task using information visualisation.

> *Research Question 1*: Does Visual Mapping Interaction affect the number of insights generated and their accuracy, when compared to an equivalent non-interactive task?
>
> *Research Question 2:* When insights are categorised based on mental effort (inferential for high and procedural for low mental effort), does Visual Mapping Interaction have an effect on the number and accuracy of insights generated in each mental effort category?

Objective Two: Investigate the compounding effects of Visual Mapping Interaction with performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles in the context of performing an analytical task using information visualisations.

> *Research Question 3:* Do Locus of Control, Self-efficacy, and Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences have compound effects with Visual Mapping Interaction, whereby according to the level of the

different measures, there will be a significant effect on the generation of insights and their accuracy?

Research Question 4: When categorising insights based on mental effort, do individual differences (LoC, SE, SA, V, A, R, and K) have a compounding effect with VMI with regards to the generation and accuracy of insights?

Objective Three: Investigate the effects of View Transformation Interaction in the context of a problem-solving task, where View Transformation Interaction is used to explore the problem data-set using a game-based simulation using a 2D and 3D visual representation.

Research Question 5: Does View Transformation Interaction affect the number and accuracy of insights identified in a problem data-set represented in a game-based simulation, when comparing to an equivalent non-interactive task?

Research Question 6: Does the view representation – 2D and/or 3D, have an effect on the number and accuracy of insights into a problem data-set represented in a game-based simulation?

Research Question 7: Does the representation – 2D and/or 3D and View Transformation Interaction have an interaction effect with regards to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?

Objective Four: Investigate the compounding effects of View Transformation Interaction with the performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles; in the context of a problem-solving task, where View Transformation Interaction is used to explore the problem data-set using a game-based simulation using a 2D and a 3D visual representation.

> *Research Question 8:* Independently from the representation, do Locus of Control, Self-efficacy, Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences have a compound effects with View Transformation Interaction, whereby according to the level of the different measures, there will be a significant effect to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?

> *Research Question 9:* Does the representation – 2D and/or 3D and View Transformation Interaction have compounding effects with Locus of Control, Self-efficacy, Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences, whereby according to the level of the different measures, there will be a significant effect to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?

1.4 Thesis Outline

This thesis consists of eight chapters. Below is the outline content of the remaining seven chapters following this one.

Chapter Two: Literature review

This chapter provides a review of the state of the art regarding the characterisation and measurement of insight, as well as a human-centred account of interaction in terms of value and benefits in the information visualisation and VA fields. This chapter also covers insight provenance, and taxonomies as well as the current approaches in insight-based evaluation and experiments, moreover it also outlines the current research regarding individual differences in the information visualisation domain.

Chapter Three: Methods

This chapter describes the methods and approach taken in this research to answer the research questions. It then depicts the key construct underlying the experiments, followed by a description of the research and the experimental design of this investigation. Finally, this chapter describes the choice of statistical analysis methods in relation to the research questions.

Chapter Four: Population characterisation and preparation

This chapter provides an outline of the population sample demographics and previous experience of the sample with the experimental settings. Then this chapter describes the different individual differences distributions and characterisations for the Aptitude-by-Treatment Interaction profiling. Finally, this chapter describes the insight categorisation used in the Visual Mapping Interaction experiment and the data collection aspects of the View Transformation Interaction experiment.

Chapter Five: Visual Mapping Interaction experiment

This chapter describes the analysis and the findings of the Visual Mapping Interaction experiment addressing objectives one and two of this thesis. Initially this chapter outlines the experiment statistical analysis, followed by the validation of the assumption of the experiment using the NASA-TLX workload assessment. Then the chapter covers the interactivity main effects before covering the compounding effects between the interaction treatment and individual differences.

Chapter Six: View Transformation Interaction experiment

This chapter describes the analysis and the findings of the View Transformation Interaction experiment, addressing objectives three and four of this thesis. As in the previous chapter, initially this chapter outlines the experiment statistical analysis, followed by the validation of the assumption of the experiment using the NASA-TLX workload assessment. Then the chapter covers the interactivity main effects and compounding effects between the interaction treatment and individual differences. Concluding with the findings of the interaction treatment using different visual representations – 2D and 3D.

Chapter Seven: Discussion

This chapter examines critically the research undertaken and its findings. It also discusses the interpreted results in the context of the research objectives.

Chapter Eight: Conclusions

This final chapter concludes and highlights the main contributions to the body of knowledge. Also, this chapter provides an outline of the potential areas of further development and future research.

Chapter 2. Literature Review

The purpose of this chapter is to establish the state of the art regarding the definition of insight as a concept and how to measure and evaluate it. Interaction is then evaluated in the context of VA, and the relationship it has with insight. Finally, this review looks at the growing research of the effects of individual differences in visualisation.

2.1 Insight 14
- 2.1.1 Definition of insight 14
- 2.1.2 Characteristics of Insight 14
- 2.1.3 Evaluating and Measuring Insight 18
- 2.1.4 Insight Provenance 22
- 2.1.5 Insight: a summative overview 24

2.2 Interaction 25
- 2.2.1 Interaction and insight 25
- 2.2.2 Interaction and Games 30
- 2.2.3 Interaction: a summative overview 31

2.3 Individual Differences 32
- 2.3.1 Individual Differences in Information Visualisation 32
- 2.3.2 Individual differences : a summative overview 35

2.1 Insight

2.1.1 Definition of insight

In their research and development agenda for VA, Thomas and Cook [5] specify the VA grand challenge as 'Enabling Profound Insights'. More recently Keim *et al.* [18] describe that deriving insight from massive, dynamic, ambiguous, and often conflicting data is one of the goals of VA. In this context insight has two distinct concepts, one taken from the cognitive sciences defined as spontaneous insight or the 'eureka' effect [28], and another known as knowledge building insight taken from the visualisation community, defined as a unit of discovery [29]. It is thus important to study insight from these two point of views as they can both be considered as distinct aspects.

2.1.2 Characteristics of Insight

Reiman [30] categorises insights, which he calls 'eureka slips', into strategies used in finding these insights. These strategies are defined as try (trial and error), read (user manual), ask (in person or over the phone), help (online or application help), stumbled across (serendipitous insight), notice (observation of an other user), email (method requesting help or public notification) and other. This has benefits in describing insights with finer granularity, but they cover a narrow spectrum for the use mainly in software applications. Additionally, this research starts to link insight to its provenance as an outcome of a task. On the other hand these definitions lack in crispness as some strategies used overlap such as the read and help categories, whereby user manuals can be often embedded within the help files of the application.

Saraiya *et al.* [29, 31] describe a thorough classification of insights and whilst the study is in the biological domain, its use can be considered domain and application independent. The characteristics of insights defined are: the number of observations and their accuracy, timing elements such as time to first insight, serendipitous aspects described as directed vs. unexpected, ability to generate new avenues of enquiry (hypothesis), domain value, breadth vs. depth and categorisation. The latter aspect, insight categorisation is also generic in nature and describes four main aspects: overview, detail, cluster and pattern. In their controlled experiment they asked participants to equationte questions about the data before the exploration. This approach helped establish expectations and goals about the data, which allows the capture of the serendipity aspects of insight. The timing elements allow the observation of how insight progresses over time. They asked participants to estimate the amount of insight gained at regular intervals, this approach enabled them to contrast and compare actual vs. perceived number of insights, thus providing some level of validation to the count. The interesting finding in this study is that experts and novices performed evenly in terms of insights, using the non-domain tools set in these experiments. Unsurprisingly, they found lack of motivation to be a factor, but despite this factor this study did yield significant results. In their discussion regarding the methodology, Saraiya *et al.* outline that this is a highly labour intensive approach, primarily driven by their use of the 'think-aloud' protocol for insight capture, but other primary concerns were the lack of domain expertise. They conclude that the VA tool has a great influence in the nature of insights and therefore it needs to be chosen carefully. Based on the these findings, this initial experiment was subsequently expanded into a longitudinal study [32]. In this second study they chose two motivated domain experts, with whom they chose the VA tools and examined their exploration of the data over a longer period of time. The main insight capturing method was using a diary analogous to the Rieman 'eureka slips' [30] and regular debriefing meetings. With this study a better understanding of the collective and connective nature of insight were investigated. This led to the ability to inform the design of VA tools for the analysis process in the domain under consideration as well as help future research design in this kind of study.

North [33] builds on the work of Saraiya *et al.* [29], and defines insights characteristics more broadly. Where the main features of insights are described as (1) complex in term of size of the data-set, (2) depth which accounts for the accumulation of insight, (3) qualitative accounting for the imprecise nature, unexpected and (4) relevance to the domain. North argues that significant insights will often rank high in these characteristics. He then discusses the suitability of control experiments to measure insight and the important aspects when considering this approach. North reasons that often, controlled experiments studying insight go against the key insight characteristic described, thus hindering the value of the approach. Substantiating his claim by describing that if the experiments have very precise instructions, they will constrain unexpected insights. Also the need for tasks to be short makes the findings shallow and domain irrelevant.

Additionally as questions have Boolean type answers they are not qualitative in nature. Finally, the simplicity of answers given in these controlled experiments is not conducive to finding relevant and complex insights. Subsequently, he argues that because of these issues it limits the generalisation of the findings beyond the remit of that particular study. Further, this paper challenges whether time related performance is relevant, as perhaps participants start to guess under time pressure. Nevertheless, this critique still values the controlled experiment format, but proposes a different methodology. Recommending increasing the complexity of the data, broadening the answer range and suggesting multiple answers format. Further, he advocates for an open-ended protocol, qualitative insight analysis and emphasis on domain relevance as used in Saraiya *et al.* study [29]. North concludes that there are benefits in using both methods of controlled experiments, In this dual approach, the open-ended experiment preceeds the task-oriented approach, thus enabling the benefit that both approaches offer.

Klein and Jarosz, [34] made a naturalistic study of insight by analysing 120 occurrences of insight in real life settings. They selected the cases, based on the account of a radical shift in the mental model. Klein and Jarosz, used Klein *et al.* [35] data/frame theory of sensemaking to define insight. The data/frame theory evolved from naturalistic decision making research where frames are mental models of a situation informing the decisions. Frames can be also seen as analogous to 'fluid' hypotheses, in that they can still evolve, as more information becomes available. In this context, the sensemaking process objective is to reach congruence between data and frame. There are two main cycles in this process, elaboration and reframing. The elaboration phase initiates from the data to create associated frames and the reframing phase starts from a frame to create a new frames. In Klein and Jarosz's naturalistic study of insight, the radical shifts refer to radical reframing such that resulting insights are more accurate, comprehensive and useful. Thus, defining insights as "*… understanding that caused specific events, seeing new relationships between elements, or identifying new ways to accomplish an outcome*." [34; p. 338].

Insights were then coded based on their provenance features as the key categorisation criteria. The qualifying insight incidents excluded the occurrences requiring more information to determine the insight provenance (1% of the total) and also incidents where researchers did not reach agreement (2% of the total). Klein and Jarosz, categorised the insights using the features as illustrated in Table 2.1 and analysed the results using Cohen's kappa coefficient as a statistical measure of inter-rater agreement to arrive at each insight provenance. After discussing the rater disagreements, they reached a kappa value above .75, which is considered excellent. The results of this naturalistic study, contextualise the impasse based problem-solving approach, used in cognitive science laboratory setting investigations. Table 2.1 shows that 55% of insights were gradual, suggesting that sudden insights of the 'eureka' type could be an epiphenomenon of insight. Additionally, 82% of insights were attributed to connection in the data versus filling a gap in the understanding / knowledge. This naturalistic approach to the characterisation of insight give a good grounding of what is relevant in the 'real world'. Thus, insight-based VA studies capturing both the sudden and progressive nature of insight to see new relationships between elements would be grounded in reality.

Feature Name	Feature Description	% Insights in Feature		
1. Connections	The person made a connection between different data points or filled a gap (yes or no).	[Yes] [No]	82% 18%	(98/120) (22/120)
2. Contradictions	The person identified a contradiction in thinking (yes or no).	[Yes] [No]	38% 62%	(45/120) (75/120)
3. Explain away vs. explore	The person tried to explain away the contradiction or else explored it further.	[Explain away] [Explore]	0% 100%	(0/45) (45/45)
4. Suspicious	The person had a suspicious or an open mind-set.	[Yes] [No]	58% 42%	(26/45) (19/45)
5. Understanding vs. action	The insight was about understanding, or understanding plus action.	[Understanding] [Action]	54% 46%	(65/120) (55/120)
6. Individual vs. collaborative	The insight involved individual effort or collaborative efforts.	[Individual] [Collaborative]	68% 32%	(82/120) (38/120)
7. New data	The insight was triggered by new data versus a reorganization of thinking without any new data.	[Yes] [No]	76% 24%	(91/120) (29/120)
8. Sudden vs. gradual	The insight was sudden or gradual.	[Sudden] [Gradual]	45% 55%	(54/120) (64/120)
9. Incubation	There was vs. was not an incubation period.	[Was] [Was Not]	9% 91%	(5/55) (50/55)
10. Search	The insight was vs. was not about how to search for data.	[Was] [Was Not]	13% 87%	(16/120) (104/120)
11. Coincidence	The insight was vs. was not based on noticing coincidences.	[Was] [Was Not]	10% 90%	(12/120) (108/120)
12. Impasse	The person struggled with an impasse (yes or no).	[Yes] [No]	24% 76%	(29/120) (91/120)
13. Surprise	The person was vs. was not surprised.	[Was] [Was Not]	91% 9%	(109/120) (11/120)
14. Accidental	The insight was vs. was not accidental.	[Was] [Was Not]	18% 82%	(22/120) (98/120)

Table 2.1. – Insight Features and Percentage Insights per Feature – Adapted from [35]

2.1.3 Evaluating and Measuring Insight

Insight has become an important performance indicator in the VA community, though challenges remain as to how to evaluate and measure such outputs [16]. Thus a deeper understanding of the evaluation methods currently in use and how to measure insight is needed.

In cognitive science insight are measured via studies that are constructed around creating puzzles that lead to a gridlock situation where participants have to change their assumptions or frame of mind to arrive at the solution, therefore generating an insight by measuring the time it has taken for the solution to be revealed [36–38]. These approaches are biased towards the ‘aha!’ sudden insight phenomena as reviewed earlier in [34], and. Further, reinforced by studies by Metcalfe and Wiebe [39], where they found that insight problems differed from non-insight ones by the sudden and unexpected nature of the solution, these, cognitive science approaches are good, but do not give a full comprehensive picture.

Using a neurological approach, Bowden *et al.*, [40] based their analysis on current limitations in insight research and the inability to detect insight unambiguously. They define a framework for insight studies using a large set of problems that are quick in their resolution and that can be solved with or without insights with an unambiguous solution. The approach uses recent neuroimaging advances in functional magnetic resonance imaging (fMRI) combined with Electroencephalogram (EEG) readings, to map and measure brain activity. With this method, they have created a neurological model that establishes that enables some predictability of insight by analysis of the mapping and activity patterns in the brain. Bowden *et al.* conclude that this kind of neurological approach together with classical cognitive science methods can help understand the brain functions and demystify the origins of insight. This framework has clear tangible benefits, but it also has major drawbacks. This approach can be intrusive and requires very specialised and expensive equipment, hence not the most readily available of method to use. Further, the detection is as good as the problems themselves, thus suffering to an extent from the same insight categorisation restrictions.

Other, objective measures of insight include, Riche [41], who proposes to investigate physiological methods to detect insights. The approaches considered use body sensors to monitor eye position and pupil dilatation as well as heart rate, muscle and brain activity. Riche aims to create a physiological model to predict insight by associating these physiological measures to observed insights and as such also suffers from the same drawbacks than the work by Bowden *et al.* [40]. Additionally physiological measures are hard to interpret unambiguously to assign a psychological value, although the measures are objective the interpretation and inferences towards insight as a psychological phenomena can be subjective.

In the information visualisation field, Lam *et al.* [16] developed an evaluation taxonomy. Insight measurements have been identified as particularly useful in two of the seven scenarios of their classification: ‘Evaluating visual data analysis and reasoning VDAR’; and ‘Evaluating collaborative data analysis CDA’. VDAR studies concentrate on how user generate actionable knowledge from insights, where as CDA studies are focused on the collaborative nature of the analysis. Both these evaluation categories are mainly conducted as either controlled experiments [29, 31, 42] or as longitudinal studies [30, 32, 43, 44] and take different formats either as a case study [45, 46] or as a laboratory experiment [47] and generally use observation as the exploration progresses or post-experiments interview or both to measure insight.

North *et al.* [48], compare different information visualisation evaluation methods using previously evaluated visualisations [49]. The first method is based on a benchmark task, the second is coined the insight method. The benchmark task method is composed of structured tasks executed by levels of complexity and collects the answers to a multiple choice questionnaire and measures the time to provide the answers, as well as their accuracy as dependent variables. Whereas the insight method is an open-ended, think-aloud protocol experiment, where the researcher silently records the timings of the insights. The dependent variables are insight count and post experiment categorisation. Due to the difference in nature of the two methods, the comparison metrics in the context of visualisation are broader and include task taxonomy and associated effort spent in the analysis. North’s *et al.* research found that the number of insights were positively correlated to the time spent on the task; where, more time spent generated more insights. Thus, North *et al.* suggest that limiting the time of the study when performing a task-based method can bias the results, by limiting the number of insights generated. Interestingly, an important breakthrough was that the type of visualisation used facilitated the generation of certain type of insights. Hence there was a relationship between type of insight and visualisation type. Equally depending on the visualisation one or the other method was favoured. Yet, interaction based visualisation acted equally on both methods. Based on the findings of the insight method they hypothesised, interaction played a role in generating more insights. Additionally, participants engaged in the insight method gave immediate feedback during the analysis, wishing there was more interactivity in the visualisation to deepen their analysis. Table 2.2 provides a comparison of the empirical results, showing the difference that evaluation has on the interpretation.

Benchmark task method results	Insight method results	
	More Insights	Less Insights
Fast and accurate	Confirms	Refute
Slow and inaccurate	Refute	Confirms
No difference detected	Expand	Expand
Not tested	Extend	Extend

Table 2.2. – Comparison of Empirical Results, (adapted from [48]**)**

North *et al.* conclude that although the insight method is more time consuming, complex and subjective to analyse, the benchmark task is more complex and time consuming to design and requires deep domain knowledge. This suggests that to reduce the complexity and subjectivity of the insight method, a generalised categorisation would address these issues. Table 2.3 gives a summary comparison of the evaluation methods, showing the benefits and drawback in each method, providing a good and valid comparison of the evaluation of the currently available methods in information visualisation. The results take a dichotomist view, and conclude that there are benefits in both methods, but implying that either one or the other must be used. Nevertheless, a viable approach might be to merge both methods leveraging the benefits of the union.

Comparison factor	Benchmark Task-based method	Insight-based method
Purpose	Evaluate specific research question about task performance	Evaluate insight generated in realistic analytic scenario
Design prep	Prepare benchmark tasks and scoring scheme Better with simple data, tools, tasks	Prepare problem scenario Better with complex data and tools
Experiment design	Benchmark task protocol Form based Time and accuracy Can be multiplexed Short-term study only Longer preparation time	Open-ended protocol Think aloud Capture insights Interaction with user Can be longitudinal Variable procedure time
User tasks	Determined by experimenter	Determined by user (user identifies insights)
Participants	Any users Many users	Expert, motivated users Motivation is detectable Train without biasing
Empirical data analysis	Processing scores data Quantitative statistical analysis	Coding rich insight and usability data Statistical analysis Higher variance Longer analysis time
Primary outputs	Identify tasks supported by a visualization Perceptual, mechanical task efficiency (time, accuracy) Statistical differences Feedback on selected tasks only, ensures coverage of those tasks Low-level tasks	Identify tasks promoted by a visualization Cognitive, interactive learning efficiency (amount of insight) Statistical differences Detects new tasks, ignores unneeded tasks Higher level tasks, user hypotheses, Summary Qualitative feedback and analytic process
Subjective bias	Choice of benchmark tasks and scoring scheme	Coding of insights and categories
Bias threat	Ecological validity	Repeatability

Table 2.3. – Summary of Benchmark Task and Insight Methods [48]

2.1.4 Insight Provenance

The characterisation and measurement of insight are can biased when taken in isolation to their provenance. Provenance gives insight context, defined as *"a historical record of the process and rationale by which insight is derived"* [50; p. 42]. Taking this definition we investigate what kind of low level analytic activity instantiated the insight, thus defining insights in terms of a task- based taxonomy.

In the paper by Amar *et al.* [51], insight provenance can be extrapolated in terms of a task taxonomy, they call it 'analytic primacy'. Analytical primacy is defined as the focus of the users analytic goals to increase the value and utility of information visualisation. They describe the taxonomy based on a thorough analysis of a corpus of 196 data analysis questions scrutinised using an affinity diagraming process. This process was based on existing document analytic literature, by Wehrend and Lewis [52] and Roth and Mattis [53]. The resulting taxonomy has ten primitive analysis tasks types (retrieve value, filter, compute derived value, find *extrema* (min-max), sort, determine range, characterise distribution, find anomalies, cluster and correlate). All these elements are defined using a *pro forma* abstract, using three key terms: 'data case' relating to the data-set; 'attribute' which is the value measured; and 'aggregation function' which is the numeric representation. Although the initial intent was to generate a common vocabulary for information visualisation evaluations, this work contributes to the state of the art in the characterisation of insight, as it helps in defining the insight provenance.

Gotz and Zhou [50] developed a taxonomy to categorise actions based on their semantic intent and was inspired by Activity Theory [54] which is widely used in HCI. Thus bridging the gap between manual and automatic insight provenance capturing challenges, such as inferring high level semantic meaning. The resulting characterisation has four tiers (task, sub-task, action and events). They focus on the action tier, which is the most relevant to the characterisation of insight. Gotz and Zhou described three elements of insight (type, intent and parameters) and the top-level actions contain three categories (exploration actions, insight actions and meta actions) divided into two sub-categories. These sub-categories are visual and knowledge insight actions which describes their origin. This research was validated using their own tool a web-based visual analytic tool (HARVEST) that captured the users behaviour using this taxonomy, providing an automated approach to capture the insight provenance, as an alternative to manual capture.

Yi *et al.* [55] take a user-centric view and argue strongly for a link between sensemaking as defined by Pirolli and Card [56] and data/frame theory described by Klein [35]. In order to analyse the behavioural insight provenance, they conducted an extensive literature survey. The outcome is a four-category grouping for the guiding processes that generate insight. These processes provide: an overview, which focuses on the big picture; then adjusts, which relates to the level of abstraction the user takes on the data by changing perspectives; followed by a pattern detection, which broadens the view into the structural aspects of the data, including trends, frequencies and outliers; the insight generation process matches the mental model, which relates to gaining a deeper understanding and confirmation of hypothesis. They also mention that all these categories are interwoven and are not discrete processes as such. An important finding that stems from this study is the key relationship between user engagement and interaction in generating insights.

2.1.5 Insight: a summative overview

Insight as a concept is multifaceted, and closely related to its provenance. It has a sudden serendipitous element as well as a gradual exploratory nature as units of discovery. Overall insights pertain to a mental vision or perception seeing relationships between elements. The key characteristics can be defined around five axes: relevance, quality, depth, complexity and predictability.

Insight-based evaluation has become the predominant method in the VA evaluation tool-kit. Controlled experiments have been the most common approaches for insight-based studies, supplemented by longitudinal studies with subject matter experts, where the typical setting has been in-depth analysis of large data-set, which are more complex and time consuming. Although neurologic and physiologic approaches are available to measure insight, clear definitions of insight and its interpretation are still required. A common denominator in the measure of insight is the quantitative aspects of absolute counting and timing of insight. Otherwise, there is no overwhelming consensus with regards to the classification of insights, but a multitude of taxonomies, which are derived from both the characteristics of insight and their provenance. In terms of evaluation methods in controlled experiment setting, although a recent study has used an open-ended insight-based evaluation, the majority of studies are benchmark tasks. Otherwise, the open-ended insight studies typically have a longitudinal design. There are opportunities to enrich the open-ended insight evaluation, by including some of the benchmark task elements into it, as at present they are mutually exclusive.

This thesis will use an insight-based evaluation in a controlled experiment setting, which aims at merging both the open-ended and benchmark tasks methods, in order to leverage the benefits of both approaches.

2.2 Interaction

Interaction is at the core of the definition of VA. Further, Thomas and Cook [5], indicate the need for a 'science of interaction' to replace 'representation' in information visualisation, establishing the importance of interactivity as well as the well researched concept of visualisation [57, 58]. Thomas and Cook, view this interactive science from three aspects: the human cognitive and perceptual constraints basis; data manipulation; and transformation or the nature of interaction itself including their visual representations. Stromer-Galley [59] describes in her paper two fundamental distinct areas of research regarding interaction. Firstly, Interaction-as-Product which relates to the interaction between user and the system. Secondly Interaction-as-Process which defines the interaction between users mediated or not by a system. Here interaction is the process that facilitates communication. Aigner [13; p. 18] further expands the latter interpretation by defining Interaction-as-Process more precisely in the context of VA as *" … the process of active discourse of users with the data."*.

2.2.1 Interaction and insight

Pike *et al.* [8] review the state of research with regards to the science of interaction as initially set out by Thomas and Cook [5] as well as recommending the future research. In their view 'interaction *is* the inquiry'. With this statement they imply that the process of inquiry in the analytical reasoning *is* interaction. When considering the elements of interaction, they acknowledge that despite the advances in taxonomical research, there is still a need to further understand the relationship between components in terms of inquiry process and the capacity to generate knowledge. They outline the recommendations for research as: ubiquitous, embodied interaction; capturing user intentionality; knowledge-based interfaces; collaboration; principles of design and perception; interoperability; and interaction evaluation. The latter aspect is of most interest, which is referred to as 'evaluating the cost and benefits of interaction'. In this section it is stated that interaction has rarely been isolated as an experimental variable, thus helping understand the benefits of interaction. Also they mentioned that as building knowledge and generating insights are the key effects of interaction, these should be the key metrics of investigation. They clearly illustrate this by declaring, "*While it is clear that* *visual representations can be informative without interactions … and interaction cannot function alone without visual representations, exactly what kind and degree of benefit is realised by allowing a user to interact with visual representations is still undetermined."* [8; p. 272].

Card *et al.* [1; p.7] describe information visualisation as *"The use of computer-supported, interactive, visual representations of abstract data to amplify cognition"*, and as mentioned earlier VA links this description of amplified cognition as the generation of insight and the key output of the process of interaction. Historically the value of interaction has been based upon empirical comparisons between static and interactive visualisations [60], though now a more theoretical approach is called for. From the cognitive science point of view, Liu *et al.* [61] describe a distributed cognition (DCog) theoretical framework as a way to describe cognition as an emergent property of the interaction by users with the visualisation applications. In the context of Information visualisation, where they challenge the traditional cognition theories, information processing only occurs in the human brain. They argue that DCog amplifies cognition, as users perform better with their tools than without. Hence to analyse the emergent property of interaction one cannot isolate the users from the tools used, but must consider the cognitive system as a whole in the analysis. They also argue that for these reasons the study of DCog is better studied in an ethnographic setting, as opposed to laboratory studies, as one needs to study the problem solving in context of the environment in which the analysis is performed with the real tools used. They challenge the validity of controlled experiments in the laboratory as they typically ignore the situated and social nature of humans. Although, they do consider that laboratory experiments could be effective if they take into account the relationship between external representations and the users internal model. But as yet they had not seen any serious analysis to this end. Longitudinal ethnographic studies also have challenges, but they believe it to be better suited to inform the design of improved systems. Based on this framework they suggest that the research agenda informing the science of interaction should consider amongst other aspects the cognitive coupling, interaction strategies in sense making and analytical reasoning as well as *"How does interaction with visual structures enable turning information into meaningful understanding?"* [61; p. 1178]. Where *meaningful understanding* can be understood as insight. Further they stress the concept of emergence, where the insights emerge from the whole system comprising a tight coupling between users and the system.

Yi *et al.* [62] in their paper describe an interaction taxonomy that aims at bridging the identified gap between interaction centric (as-product) and user task-centric taxonomies. The resulting taxonomy is based on the user intents and can be related to the concept of insight provenance as discussed earlier, but with the specificity of interaction as the key mechanism. The method employed was to survey the existing literature and systems, from which they extracted the different interaction techniques, which they then categorised. The resulting taxonomy has the following seven categories:

- Select: mark something as interesting;
- Explore: show me something else;
- Reconfigure: show me a different arrangement;
- Encode: show me a different representation;
- Abstract/Elaborate: show me more or less detail;
- Filter: show me something conditionally; and,
- Connect: show me related items.

These intent-based interaction categories provide a good foundation to establish the insight provenance from the interaction point of view. Thus would help in defining the value of interaction from an insight-based evaluation framework to assess the VA tools.

Lam [57] describes a framework of interaction costs inspired by Norman's [63] seven stages of action. These can be considered as the different steps in the process of interaction. To create this framework, Lam collected interaction related usability problems in 32 user studies. Figure 2.1 Illustrates this framework.

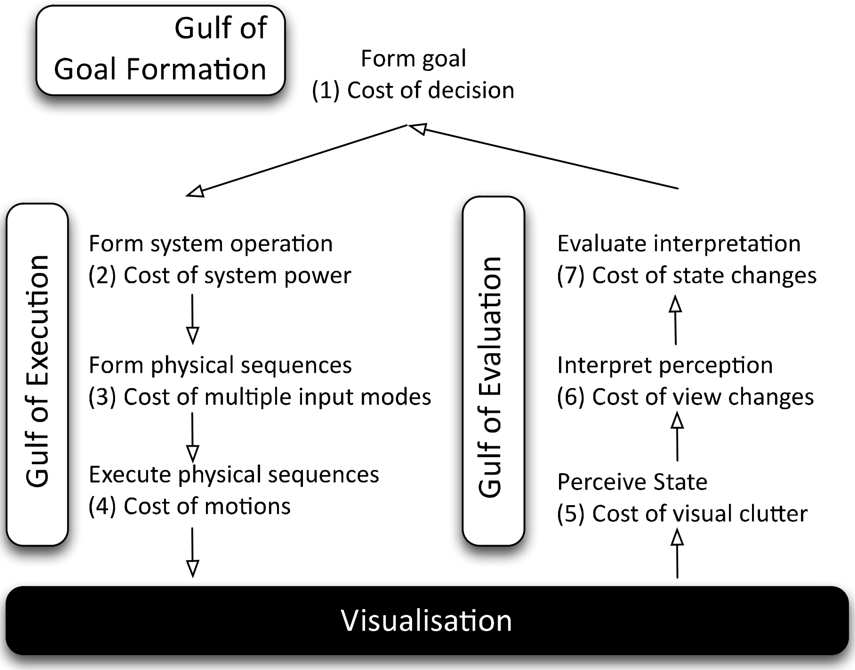


Figure 2.1 – Framework of Interaction Costs Adapted from [57]

The resulting seven costs as illustrated in Figure 2.1 are:

1. Decision cost to form goals relates to the possibility of loosing the context of the data when exploring a subset;
2. System-power cost to form system operations relates to the definition of the actual operations to perform in relation to the possible complexity of the tool;
3. Multiple input mode cost to form physical sequences refers to the interaction choices the user has in relation to the state in which the user is and the possible confusions thus created;
4. Physical-motion costs to execute sequence defines the actual HCI action required to achieve the user's intent;
5. Visual-cluttering cost to perceive state relates to the typical visualisation challenges of large data-sets where the visual real estate is limited and different tool artefacts can hinder the user's perception;
6. View-change cost to interpret perception refers to the users expectation and how the tool may mismatch them and hinder the interpretation; and,
7. State-change costs to evaluate interpretation relates to the contextualisation of the detail in the bigger picture and connection to other data subsets.

All the above costs in the process of interaction provide an adequate evaluation framework to assess the value of the different tools by establishing the costs they impose upon the user. Lam argues that the aim of the different tools should be to narrow the different gulfs defined, which are the gulf of execution and evaluation as defined by Norman [63] in addition to the gulf of formation that Lam defines as the difficult task that establishes the precise intent the user has on the data. Regarding the latter, she is not sure whether it is the role of the information visualisation community to help users in this challenge, but in the context of visual analytics the study is definitely in scope of the work of this thesis.

Liu and Stasko [64] in their paper give a top-down perspective of mental models, visual reasoning and interaction. They focus on the relationship between internal (mental model) and external representations (visualisation). Thus, they define the concept of mental model in information visualisation, as *"… a functional analogue representation to an external interactive visualisation system"* [64; p. 1001]. In this definition the mental models preserve the structural and behavioural properties of the external system, as well as the schematic and semantic information about the data. Additionally, they argue that a mental model of interactive visualisation is constructed in the short-term memory (STM) for reasoning in a particular problem. Further from a developmental perspective, they describe the interactive process into four internal-external interchanges (internalization, processing, augmentation and creation) asserting that interaction *is* a central part in the cognitive process. They argue that a holistic view on interaction should be founded on user intent. Taking into consideration the two primary human cognitive limitations regarding mental modelling, (i.e. limited working memory (STM) and limited accuracy of the information held), they propose that the interaction has three primary purposes: external anchoring; information foraging; and cognitive offloading. This view is in line with the DCog view discussed previously in Liu *et al.* [61]. They also discuss the implication for evaluation, such as the need to have protocols such as 'think-aloud' and 'verbal analysis'. Furthermore they give particular emphasis on the need to account for individual differences, highlighting that very little work has been done in this area. They believe that aspects such as visual memory, special cognitive ability and learning style are key aspects that require further research.

2.2.2 Interaction and Games

Keim *et al.* [18] when describing the spatio-temporal domain, show how VA is playing an increasing role in decision-making. This area of VA uses 2D and 3D simulation models in a 'serious games' environment. Decision makers interact in an immersive game, simulating the situations and challenges may be confronted with in real life. These simulations enable the use of VA for problem solving in a safe environment. In this context, serious games can be considered as a sub-domain of VA and also analogous to e-learning where interaction is an important aspect of the overall experience and benefits [65]. Not a lot of research has been done in this sub-domain and more would be beneficial.

Bown *et al.* [66] describe four case studies of the use of 'serious games' within a VA context (sustainable urban planning, police firearms training, soil science and cancer systems biology). In these contexts, they highlight the need to consider the interactivity evaluation from the human perceptual and cognitive processes point of view. In terms of perception they describe the concept of immersion, which is assessed using physiological measurements such as EEG and skin conductance. Additionally, they suggest that immersion can also be measured by physical observations, such as eye-movement tracking. These physiological and physical approaches as discussed in section 2.1.3 regarding measuring insight, have challenges in terms of costs, intrusiveness and reliability with regards to open-ended insight-based evaluations. In terms of cognitive processes, the performance metrics are the same as the information visualisation domain, using task performance as the measurement focus and an insight-based approach to serious game analysis could bring similar evaluation benefit as those found in information visualisation.

Diakopoulos *et al.* [67] argue that data-driven games can be analytical and insight generating experiences. They build their views on the work by Pike *et al.* [8], discussed in the previous section where in the context of games, it discusses the accumulation of insights by transforming the view of the data. Highlighting the potential benefits of using game constructs as systems involving data relationships to extract insights and understand their interactive provenance.

Salem and Zimmerman [19] in their book about game design fundamentals define games as interactive systems, having what they call a 'core mechanic' the key to interactive elements of a game. This interaction core mechanic engages the user's cognitive and psychological processes.

Previous research [68], [69] analysing 2D and 3D representation in information visualisation have concentrated on the intrinsic nature of the representation and studied the value and benefits in terms of information visualisation. More research into the interactive aspects is needed, by taking interaction as the comparison metric between visual representations.

2.2.3 Interaction: a summative overview

To analyse the emergent properties of interactions (namely insights), one must consider the whole cognitive system (human-visualisation) in the analysis. Laboratory experiments can be effective when the analysis aims are the appreciation of the role of interaction with visual structures in generating meaningful understanding (insight), and the relationship between external representations and user internal models are considered. Understanding user intent behind interaction, equates to understanding the interaction that drives the insight, leading to the concept of insight provenance. Thus as interaction is central to reasoning it must be analysed with the user intent in mind. Also further research is required in areas of visual memory, cognitive abilities and learning styles to gain a more detailed understanding of the role of interaction in the reasoning process.

Under the view that interaction ***is*** the inquiry, there is the need to research further its cost and benefits capacity to generate knowledge and generate insights. Thus, this thesis aims at expanding these research aspects by considering interaction as an independent variable.

The use of serious games has clear benefits in the VA domain, where interaction plays an essential part. Insight-based research in serious games based VA, has had little research to date. Investigation into the benefits of view transformation interaction within an insight-based evaluation methodology would contribute to this area of VA.

2.3 Individual Differences

HCI has a long history of accounting for individual differences in user analysis research [21]. In the field of visualisation the interest in individual differences is recent.

2.3.1 Individual Differences in Information Visualisation

Conati and Maclaren [70] in their paper explore the role of individual difference for the various cognitive abilities (visual memory, spatial visualisation, perceptual speed, disembodiment, need for cognition and learning style) with regards to performance in using target visualisations. In their experiment they designed the tasks according to the taxonomy by Amar *et al.* [51], also discussed in the insight section of this review. The results found only one significant effect attributed to a cognitive ability (perceptual speed) with regards to the visualisation effectiveness. When investigating further into the accuracy with target visualisations, they discovered that other cognitive abilities could be used as a predictor of performance (need for cognition, special visualisation and learning style).

Previous work by Allen [71] has shown similar results, associating user performance to design features on individual differences in the cognitive ability for search tasks, showing that both compensatory and capitalisation matching were present. A performance increase was observed for participants with lower cognitive ability. This increase was attributable to the augmentation benefits the system provided. Moreover users with higher cognitive ability got greater benefits with features of the system that demanded greater cognitive resources. Additionally, the results also suggest that users do not self-adapt to the systems features that best suited their cognitive capabilities, thus leading to the need to adapt the tools.

Ziemkiewicz and Kosara [72] investigated the individual differences compatibility with the preconceived visual metaphor preferences and associated performance. The individual differences studied were taken from the Mini-IPIP Big Five personality [73] covering Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism constructs. The results confirmed that compatibility had an influence on performance, but decreased with users with high scores in openness and spatial cognitive abilities, or user with no *a* *priori* preference. A surprising result was that response time compatibility effect was only true for women. A strong relationship of visual and verbal metaphors in participant's comprehension of the task was present, but as in Allen [71], they did not find signs of adaptation. Ziemkiewicz and Kosara concluded that when evaluating new systems, gender, spatial ability and personality are important aspects to consider. Additionally, they recommend taking users preconceived preferences with care, as they do not show unambiguously true preferences.

Green and Fisher [9–11] undertook two studies comparing procedural learning with two interfaces, an information visualisation application and a web table. Their ongoing aim is to advance research towards a personal equation of interaction, which takes full account of individual differences in a predictive manner. In their studies, they investigated three main psychometric measures, Locus of control (LoC), IPIP 20-item Big Five Neuroticism and IPIP 20-item Big Five Extraversion [73], questioning whether these had a significant effect in performance and to what degree these relate to the number of insights reported. Green and Fisher define insight as knowledge gained from either the content or the ontological relationship. Table 2.4 outlines the key results on these studies. Additionally, in [9], they found that for inferential tasks, LoC was the best predictor of performance. Contrary to expectation, externally oriented LoC participants had better performance scores in inferential tasks and when the task became more complex made fewer mistakes with information visualisation tool than with the web table tool.

	Completion Times	Errors	Insights
Interface	Faster times in Web Table	Fewer errors in Web Table	More Insight in Information visualisation Application
Locus of Control	Internal LoC faster times	None	External LoC more insights
Extraversion	More extraverted faster times	None	Less extraverted more insights
Neuroticism	More neurotic faster times	None	Less neurotic more insights

LoC: Locus of Control

Table 2.4. – Summary of Green and Fisher Results [9–11]

In [11], when covering the next steps in the personal equation of interaction, they describe how individual differences can improve the interpretation of visually enabled analysis. They conclude that the best performers in terms of completion time had low information usage, a dislike for ambiguity, low extraversion and low need for cognition scores. Additionally, these high performers were more prone to have their behaviour dictated by their emotions. These studies, lead the way to define user profiles using cognitive tasks instead of user group membership. In their discussion, they call attention to the need to further evaluate the individual differences with regards to insight as a measure. They highlight the lack of clear consensual definitions of insights is a key challenges in insight-based evaluations, where clear definition of insight is paramount to enable analytical comparison between studies. In their knowledge-insight perspective, they found significant differences in relation to the type of visualisation, but do not report any correlation with individual differences. Additionally, they mentioned that further research is required to understand the effects of learning style as a factor for visualisation.

Ziemkiewicz *et al.* [12], in their study, focus on the influence of LoC on performance according visualisation style building on previous research by Green *et al.* [9, 10]. They hypothesise that the differences found in Green *et al.* work regarding inferential tasks performance in the information visualisation tool, is due to layout rather than the task. In their study, they used, four layouts: basic tree view; bordered tree; indented boxes; and, nested boxes. These are used in search and inferential tasks, based on four data-sets as part of their experiment. The key finding supporting their hypothesis was that participants with an external LoC performed better in the nested box view which is the layout used by the information visualisation tool in Green *et al.* [9]. They conclude that LoC is a robust measure and can be directly related to the performance of data exploration tasks, where each group performed differently according to the visualisation layout. Although they cannot infer causality, they argue that external LoC users may be more willing to adapt to visualisation types, thus explaining these results. They link their findings to distributed cognition views discussed above in Liu *et al.* [61] and works by Cassidy and Eachus [74] where external LoC is linked to surface learning. Ziemkiewicz *et al.* hypothesise that perhaps internally focused user ability to adapt is a benefit when using visual analytics systems, where a reliance on external systems are required. Externally focused users, may find it difficult to use visual analytic tools, as they may need to align their internal views to the external representations of the VA system. Based on these results and hypothesis, they recommend that VA system designers should take into consideration LoC relating to the explicit nature of the visualisation layout. More broadly, when the audience may have a more externally focused LoC, it might be useful to deviate from the classical Tufte's ink-to-data ratio considerations [75], by making the layout hierarchical features more explicit to help the exploration process. On the other hand for more internally focused LoC user, they hypothesised that for users with pre-existing mental models such as experts, a stricter ink-to-data ratio design would be more beneficial.

Chen [76] investigated individual differences in spatial-semantic virtual environments finding that experience of the environment had the most significant effect on performance rather than the cognitive abilities studied (spatial ability and associative memory). The experiments compared spatial and textual settings, where the spatial user interface was developed as 3D visual representation of a network diagram. Chen concludes that more research is required to fully understand the interaction effect with the virtual worlds.

Chen and Toh [77] and more recently Hauptman and Cohen [78] investigated the learning style aspects of individual differences for virtual reality (VR) environments, using different learning style instruments. Whereas Hauptman and Cohen used the VARK multimodal learning preferences [79] and Chen and Toh the Kolb learning style inventory. Their findings differ, Chen and Toh adopted reported that learning style had no effect on the VR performance whereas Hauptman and Cohen in their more recent study, reported learning style effects on performance. These latter results promote the strength of VARK as a multi-dimensional and scaled instrument, as per Miller's [80] recommendations when choosing a learning style instrument. Moreover, Hauptman and Cohen cannot fully explain their results and recommend further research to uncover the verbal and non-verbal impact on performance.

2.3.2 Individual differences : a summative overview

Research suggests that individual users do not adapt their cognitive abilities to the visualisation tool. Further, studies show the potential and need to adapt the visualisation tools according to the task and the individual differences to improve performance. Although the adaptation is limited at present, future research mapping further the individual differences will improve this capability. Insight-based evaluations in this domain require a well-defined definition of insight, in order to enable inter-study comparisons and quantify contributions. Further research is needed to demonstrate the impact of individual differences such as learning styles and self-belief cognitive constructs on the outcomes of interaction. These advances will benefit the design of future VA tools, and interface individualisation.

2.4 Thesis Objectives and Research Questions

Following the literature review, the objectives of the thesis can now be put into context with respect to previous studies. Here, each of the objectives of this thesis are outlined and justification given for their inclusion in the study. Where appropriate the ‘research gap’ reported previously is stated explicitly.

2.4.1 Objective One

Investigate the effects of Visual Mapping Interaction (VMI) in the context of performing an analytical task using information visualisation.

Rational for Objective One

The visual analytic agenda [5] calls for a science of interaction outlining in particular the need to look at the nature of interaction as data transformation and manipulation. Further, in a recent review of the state of the art of the science of interaction in the context of VA, Pike *et al.* [8] suggest that further research is required. In their view interaction ***is*** the enquiry, and the research advances suggests to aim at further understanding the inquiry process and its capacity to generate knowledge beyond the visualisation aspects previously researched [57, 58]. Cognition science research into information visualisation interaction [61, 64], calls for more research into the science of interaction to investigate the cognitive coupling between interaction and analytical reasoning. This additional research would lead to understanding how the visual structures turn information into insights. Moreover, research aiming at defining a ‘personal equation of interaction’ [9–11] and related research [12], use procedural and inferential tasks to define different mental effort for the task evaluation.

2.4.2 *Objective Two*

Investigate the compounding effects of VMI with performance-related psychometric measures (LoC, SE, and SA) and the VARK model of learning styles in the context of performing an analytical task using information visualisations.

Rational for Objective Two

Although individual differences research in HCI is well established, in the field of information visualisation the attention is recent. There is a good basis of research [70–72] that have investigated many aspects of individual differences such as Locus of control, spatial visualisation and learning style to mention but a few, but have all been focused on the visualisation aspects. Moreover, research aimed at defining a 'personal equation of interaction' [9–11], call for further research into the learning style effects as a factor in interactive information visualisation. Still this research [9–12], has so far looked at the visual structures and not at the interaction aspects per se. Thus has not addressed the need to take interaction as the independent variable that is called for by Pike *et al.* [8]. Similarly, the learning style aspects have not been investigated. Additionally, existing visualisation research investigating various learning style models [76–79] suggest that the VARK model of learning styles multidimensional and scaled instrument has good results in uncovering the verbal and non-verbal aspects of performance.

2.4.3 *Objective Three*

Investigate the effects of view transformation interaction (VTI) in the context of a problem-solving task, where VTI is used to explore the problem data-set using a game-based simulation using a 2D and 3D visual representation.

Rational for Objective Three

Keim *et al.* [18] discussed the increasing role serious games have in visual analytics as a decision-making aid tool used within a simulation environment. In this context, serious games can be considered as a sub-domain VA and e-learning, where interaction is an important aspect of the overall experience and benefits [65]. Moreover, as objective one focused on the visual mapping interaction and taking the information visualisation reference model developed by Card *et al.* [2], another relevant area where human interaction occurs as illustrated in Figure 1.1 is interacting with the view transformations.

The viewpoint control described [2] is considered a view transformation in information visualisation. viewpoint interaction is a core function of a serious game-based simulation. No research has been found addressing this aspect with regards to the VA agenda [5, 8] calling for a science of interaction. Further, research such as [67] taking a 2D representation of a serious game, calls for a further research in understanding the effects of representation in gathering insights. Additionally, research [68][69] analysing 2D and 3D representation, have looked at taking interaction as the comparison metric between visual representations and the benefits in terms of information visualisation, however this is still an under-researched area.

2.4.4 Objective Four

Investigate the compounding effects of View Transformation Interaction (VTI) with performance-related psychometric measures (LoC, SE, and SA) and the VARK model of learning styles; in the context of a problem-solving task, where VTI is used to explore the problem data-set using a game-based simulation using a 2D and a 3D visual representation.

Rational for Objective Four

Objective two, outlines the rational in the VMI context, and within the information visualisation reference model [2], VTI would benefit from the same research. Additionally, adding to the body of knowledge regarding individual differences in the VR environment [76–79] and game-based VA.

[Intentionally left blank]

Chapter 3. Method

This chapter describes the methods and approaches taken to respond to the research questions. Initially, defining the key underlying constructs of the experiments, followed by a description of the research and experimental design used in this investigation. Then this chapter details the Visual Mapping and View Transformation interaction experiments. The chapter concludes with a description of the tools and statistical techniques used to address the research questions and the associated statistical power calculations.

3.1 Introduction 42
  3.1.1 Individual differences 42
  3.1.1 Aptitude-by-Treatment Interaction 45
  3.1.2 Workload Assessment 46
3.2 Experimental Design Overview 49
  3.2.1 Participants 49
  3.2.2 Design 49
  3.2.1 Pre-study 51
3.3 VMI Experiment 53
  3.3.1 Dependent Variables 54
  3.3.2 Experiment Setting 58
  3.3.1 Experiment Procedure 62
3.4 VTI Experiment 66
  3.4.1 Dependent Variables 67
  3.4.1 Experiment Setting 68
  3.4.2 Experiment Procedure 70
3.5 Statistical Analysis 74
  3.5.1 Power Analysis 78

3.1 Introduction

The key aims in this thesis stem from the visual analytic agenda [5] and the recent review of the state of the art in the science of interaction in VA by Pike *et al.* [8]. In order to address the research questions outlined in the literature review, interaction is investigated as an independent variable, using an insight-based evaluation methodology. Furthermore, interaction is analysed from the visual mapping and at the view transformation levels, as described in the information visualisation reference model [1] described in the introduction chapter.

For the visual mapping interaction experiment, this investigation aims at quantifying the effect of interaction in terms of insight generation, accuracy and the effect on mental effort in an analytical task. As for the view transformation interaction, the aim is to evaluate the insights gained into the problem set and the difference interaction has in terms of performance (measured as a ratio of accurate insights over the overall number of insights) on 2D and 3D visual representations. Additionally in both experiments, building on research on individual differences in information visualisation [9–12] and VR [76–78], this thesis aims to explore the compounding effects of individual differences with interaction processes. The individual differences investigated were the VARK model of learning style preferences and psychometric measures linked to problem solving performance, namely locus of control, self-efficacy and self-assessment.

As reviewed in the literature, insight is the key purpose and output in VA, and insight-based evaluations have become the standard approach to assess new tools, techniques and processes in this domain. Where in essence the key characteristics of insight are defined around five axes: relevance, quality, depth, complexity and predictability [29], [33].

3.1.1 Individual differences

The literature review in the previous chapter covered the state of the art of individual differences research in VA from the information visualisation perspective. Very little has been done in VA with regards to learning style investigation, where the primary research has been focused on education related fields such as e-learning.

Psychometric Measures

This research considers Rotter's locus of control (LoC) [23] to enable comparisons with previous research. LoC a very well studied and popular measure in psychological studies, it suggests how much a person believes they are in control of the events in their lives. The scale categorises individuals from external to internal LoC, based on the degree of belief that events are controlled internally (the individual) or externally (environment, outside the individual) [24]. This measure has been studied in many different contexts; for instance, it has been found that users with a more internal LoC were more effective at work [81]. Recent work in visualisation [9–12] has found that it also has similar significant effects in information visualisation. Rotter in his work specified that the scale was a gradient of generalised expectancies and not the cause of the behaviour. Also, it was important to understand that healthy users would typically be in the middle of the scale with a leaning towards internal focus.

This thesis aims to expand the scope of psychometric measures studied in information visualisation, by investigating measures related to internal and external perception of performance. Particular attention was given to the length of the assessment forms and the associated risks with regards to lack of response, bias and accuracy of the responses [82]. Hence, to minimise the length of the psychometric survey, only the well-researched LoC measure was kept from previous studies to enable comparison and to allow for other psychometric measures to be considered. Thus the Big Five Neuroticism and Big Five Extraversion were excluded from this study. The additional psychometric measures consider the internal and external perception of performance. These are self-efficacy, which measures the attitude towards goals and challenges which is also a high level indicator of self-belief of good performance; and, self-acceptance, which is concerned with confidence in personal decision-making and asserting one's own viewpoint.

SE measures have been shown to relate amongst other factors to levels of motivation and resilience to adversity [25]. Also, SE measures have been correlated to work-related performance. This measure was part of Bandura's social cognitive theory [25] as an *'agentic perspective'* which implies the human capacity for control over their lives and nature. Bandura argues that SE interacts with the environment in a predictive manner according to the high or low levels of SE, where high levels of SE are indicators of self-belief of good performance.

Self-acceptance research [26], has outlined that high scores of SA were associated with more resiliency to stress and higher effectiveness. SA is part of the California Psychological Inventory (CPI) which has been used in leadership assessment [83] and also in assessment of gifted individuals programs [84], but at present has not been researched within HCI and VA in general.

All these psychometric measures were evaluated using the international personality item pool (IPIP) [85]. IPIP reports the following Cronbach's alpha reliability statistics. LoC has a Cronbach's alpha (α = .86) and the SE measure (α = .81) both considered good, and the SA (α = .78), was acceptable, borderline good [86]. The actual Cronbach's alpha reliability, measuring the accuracy and dependability of the questionnaire used to assess the individual differences of research participants in this study, is outlined in Chapter 5 in more detail. All the questionnaires were structured as a 5-point Likert scale, asking the participants to assess the accuracy of the statements presented.

Measures such as the Myers–Briggs Type Indicator (MBTI) have been considered as they would be complementary in learning style studies [87], but this metric is commercially available only which in addition to being costly, the use of the materials requires accreditation. Hence, MBTI has not been included in this research study.

Learning Preferences

There are many different learning style models; the most popular are the Kolb Learning Style Indicator (K-LSI) [88], Gregorc Style Delineator (GSD) [89], the Felder–Silverman Index of Learning Styles (FS-ILS) [90], the VARK Questionnaire [91], and the Dunn and Dunn Learning Style Index (DD-LSI) [92]. Some of these are commercially available (K-LSI, DD-LSI and GSD), where the others are freely available (FS-ILS and VARK). The K-LSI and the DD-LSI have some commonalities and classify learners into four bipolar modes that are based on behavioural aspects of the learning process. The DD-LSI is more extensive in its schema and considers five stimuli (environmental, emotional, sociological, perceptual and psychological) and measures learners on a twenty point scale. The FS-ILS, measure learners on strengths and preferences along five bipolar continua based on how learners absorb and process information. The VARK questionnaire was based on a sensory model, built on the neuro-linguistic programming work [93] and measures learners on their perceptual preferences and strengths.

In this research, the primary focus was perceptual aspects of the learning process. Although, DD-LSI, FS-ILS and VARK consider perceptual aspects, only DD-LSI and VARK were comprehensive in their coverage, and VARK was the only freely available learning preference model and was chosen for this study for this reason. Additionally, Miller [80] in the description of the benefits and challenges of evaluating computer-based instruction using learning styles, concludes that using the correct learning style instrument is key, recommending multi-dimensional and scaled measurements such as VARK.

In the VARK preference model the V is the visual preference, which refers to information presented in a symbolic manner without words, such as graphs and charts. The A is aural preferences, which relates to taking the information presented in audio. The R is the read-write preference that categorises people who absorb information in written format. The K is kinaesthetic preference relating to a practical learning preference. Additionally, the premise of the VARK preference systems is that people are mostly multi-modal, with preferences. This approach makes scores in each perceptive modality valid on its own, as they're not bipolar in nature. The learning preferences profiling standardisation is still undergoing, but a study by Leite *et al.* [13] found preliminary support for the validity of the VARK scores with reliability estimates of .85, .82, .84, and .77 for the V, A, R, and K subscales respectively.

3.1.1 Aptitude-by-Treatment Interaction

To investigate individual differences, aptitude-by-treatment interaction [94] is a well established methodological approach for this purpose, particularly in the education domain [95]. Also, more recently Chen and Toh [77] used ATI to study the VR effects on learning based on the users learning styles. ATI is multi-factor in nature; it examines the effects of individual differences (aptitudes) on learning outcomes according to the instruction method (treatment). In this research the treatment considered is the ability to interact or not with the visual mappings or view transformation as described in the introduction chapter.

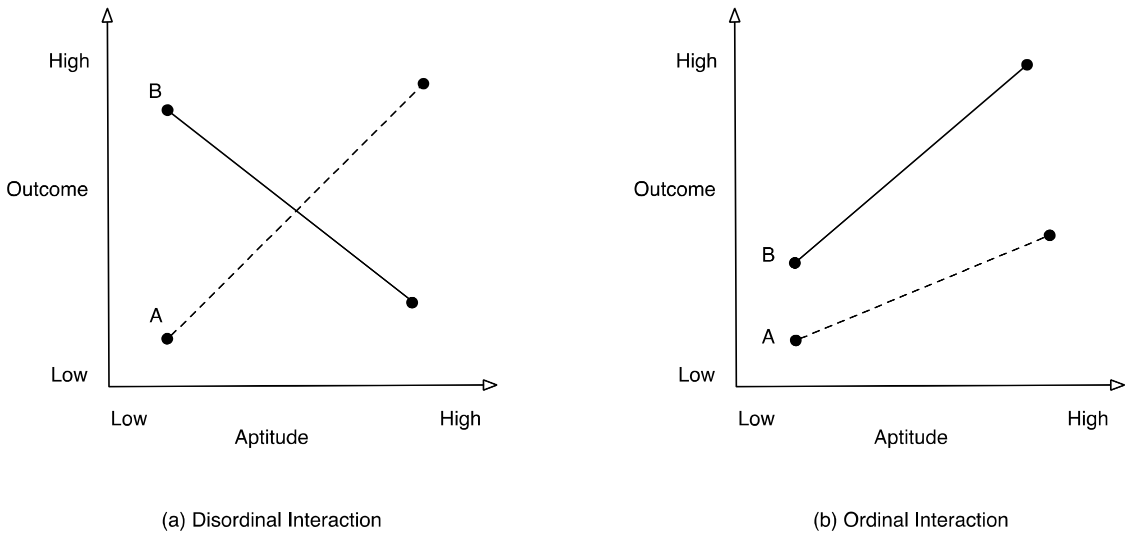


Figure 3.1 – ATI Types

Jonassen and Grabowski [96] describe the ATI into two types of interactions see Figure 3.1. In the disordinal interaction (a), participants with low aptitudes scores performed higher with treatment B than A, but participants with high aptitudes scores had the reverse effect, whereby they performed better with treatment A than B. In the ordinal interaction (b), both the low and the high aptitude scored participants performed better with treatment B than A. Thus, results from an ATI method can help understand what were the levels of aptitude that yield the desired outcome. Therefore, this method can be used to select or advise users based on the individual differences, to increase their performance outcome.

3.1.2 Workload Assessment

Workload assessment methodologies are widely used in human factors research and HCI, in order to assess how difficult participants find the different tasks performed. In this study, workload assessments are used to provide validation with regards to the choice of experiments and the procedure used.

The HCI field uses subjective and objective methods for workload assessments. The principal objective methods are based on physiological measures, measuring heart rate and perspiration for example. In this thesis investigation the objective measures were discarded, as these approaches would have been too intrusive. The subjective approaches have the two major reliable and validated [97] methods: the Subjective Workload

Assessment Technique (SWAT) [98] and the NASA Task Load Index (NASA-TLX) [99]. An evaluation by Rubio *et al.* [100] of the SWAT and NASA-TLX found that they were equivalent and highly correlated methods, and this evaluation recommended to use NASA-TLX, when assessing the workload of an individuals on a specific task. The recommendations are based on the following five factors. Intrusiveness (1), both methods are found to be equivalent as both uses questionnaires to make the measurements. Sensitivity (2), the NASA-TLX was found to be more sensitive in evaluation by Rubio *et al.* . Convergent validity (3), looked at the Pearson correlation factor for the mental workload in relation to the global workload, and both SWAT and NASA-TLX, were highly correlated ($p < .001$). Concurrent validity (4), looked at the Pearson correlation factor for the performance workload in relation to the global workload, and both SWAT and NASA-TLX were also highly correlated ($p < .001$). Finally, diagnosticity (5), this evaluation looked specifically at how the mental workloads discriminated between tasks, here also the two workload assessment methods performed very similarly, but the SWAT had a high discriminant power.

In the experiments performed in this study, the workload assessment needed to evaluate the individual participants workload on the specific task of insight generation and gathering. Based on the recommendation of Rubio *et al.* [100], the workload assessment method used in this investigation is the NASA-TLX. Thus, this method is used to measure the difficulty experienced by participants in the different interaction treatments of the experimental settings described later in this chapter.

Title	Endpoints	Description
Mental Demand	Low / High	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low / High	How much physical activity and coordination was required (e.g. keyboard combination, mouse touch, mouse keyboard combination etc.) Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low / High	How much time pressure was felt due to the rate or pace at which the tasks or task element occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good /Poor	What is the perception of success at accomplishing the goals of the task (answering the post-experiment questionnaire). What was the satisfaction level of the performance in accomplishing the task?
Effort	Low / High	How hard was the task to accomplish the level of performance attained?
Frustration Level	Low / High	What was the felt level of insecurity, discouragement, irritation, stress and annoyance **versus** security, gratification, contentment, relaxation and complacence?

Table 3.1 – NASA-TLX Rating Scale Definitions Adapted from [99]

The NASA-TLX measures the experiment on six key metrics mental, physical, temporal demands and performance, effort and frustration on a (0 – 10) scale. Then, the participants give a weighting to the different measures in pair-wise fashion define their relative importance in order to provide an overall workload score. Table 3.1 outlines the definition of these metrics.

3.2 Experimental Design Overview

The experimental design consists of two distinct experiment settings. The first experimental setting addresses objective one and two of this thesis aiming at investigating the effects of VMI in the context of performing an analytical task using information visualisation, and the compounding effects VMI has with individual differences in the same context. The second experimental setting (divided into two experiments) addresses the third and fourth objective of this thesis investigating the effects of VTI as a main effect and the interaction effect with individual differences in the context of a problem-solving task using a game-based simulation with 2D and 3D visual representations.

3.2.1 Participants

The participants were recruited from the general university population of undergraduates, postgraduates, staff and relatives aged 18 – 65. No other restrictions were imposed on the selection of participants, in order to obtain an as wide as possible spread of individual differences. As soon as a prospective participants expressed an interest in the research study, they were sent by email the experiment's participant information sheet (see Appendix 2), explaining the experiment and the steps that it comprises as well as what was expected of the participants. In the same email the participants received the ethical consent form for signing (see Appendix 1) which was collected on the day of the study.

3.2.2 Design

One of the objective of this thesis is to determine the effects of interaction upon insight generation and accuracy. As discussed in the literature review, controlled experiments are the most common approach to conduct this kind of insight-based study [16]. This research can be categorised according to the taxonomy developed by Lam *et al.* [16], as evaluating visual data analysis and reasoning VDAR, and the methodological approach taken was a mixed-design set of two controlled experiments with interactivity as the independent variable with two levels (interactive and non-interactive). The first controlled experiment (VMI) investigates interaction with visual structures in an analytical task. The second examines the interaction with view transformation within a 2D representation and a 3D representation in a game-based simulation problem-solving task. There are a few differences with the VMI an VTI experiments environment and setting. The VMI studies the interaction effects with the context of VA as it intersects with information visualisation. The VTI experiment goal is to investigate interaction in the context of VA, where it intersects with game-based simulations. These two settings enable a broader exploration of the effects of interaction in the VA field. Sections 3.3 and 3.4 of this chapter describe these experiments in further detail. The key dependent variable is insight. Additionally, the insights were categorised in all experiments in terms of accuracy (accurate or inaccurate) and for the VMI experiment also by mental effort (inferential or procedural), which was analysed as within-subjects factor. Furthermore, to understand the effects of individual differences, a group differences design was used using the ATI [94] methodology to do the analysis.

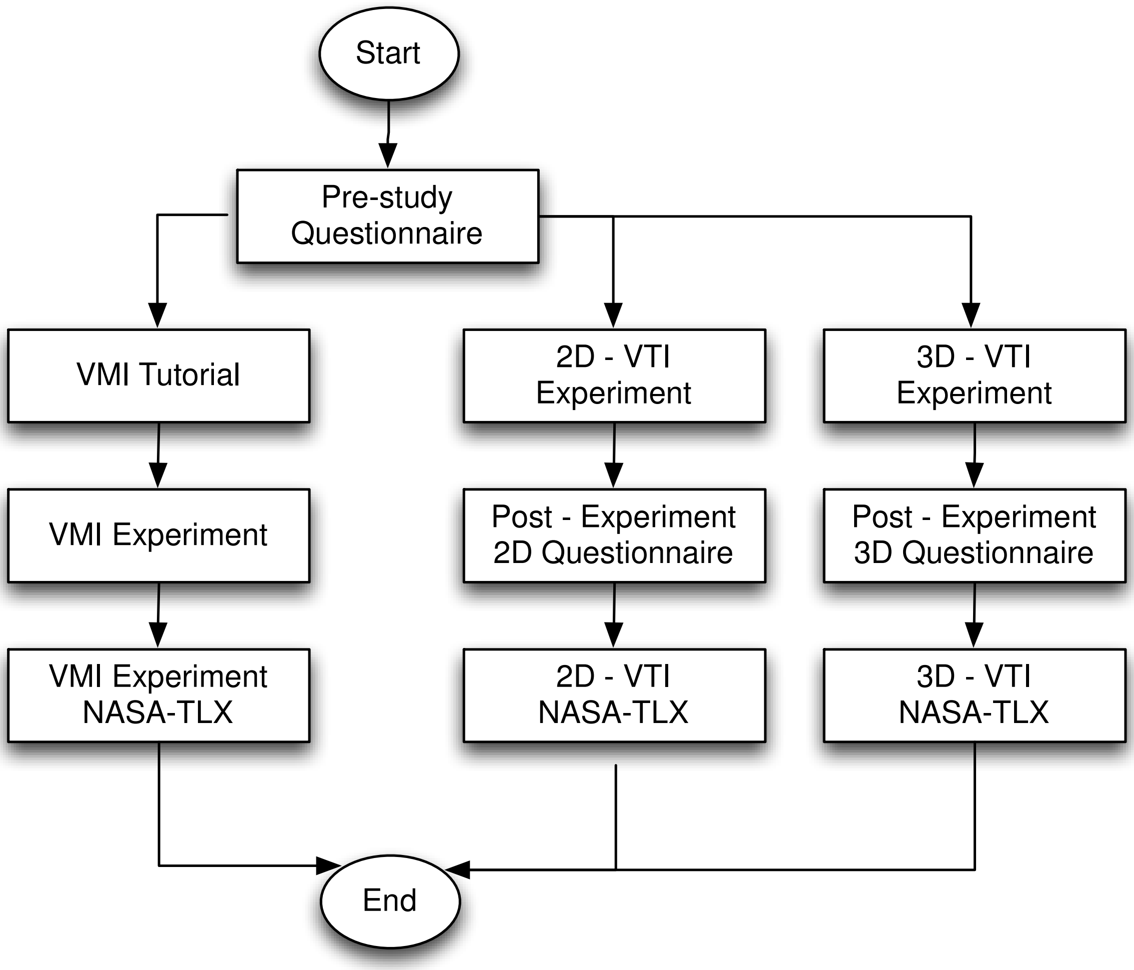


Figure 3.2 – Experimental Design Overview

Figure 3.2 illustrates the experimental design overview of the three experiments of this thesis. The main independent variable (interactivity) was allocated randomly by assigning participants into an interactive (IG) and non-interactive group (NIG) for the three experiments; also the allocation was separated by gender to avoid gender bias in the groups and the visual representation randomised to prevent learning bias. All experiments were independent, and were performed by the same participants in single session. Prior to the experiments all participants filled an pre-study online questionnaire. The VMI experiment contained three parts. Starting with a tutorial task, then the actual experiment and concluding with a NASA-TLX assessment. The VTI experiments had three parts, starting with the experiment, followed by a post-experiment questionnaire and ending with a NASA-TLX assessment, Table 3.2 provides the detail procedure of the experiments.

3.2.1 Pre-study

The participants complete the pre-study questionnaire (see Appendix 6). This questionnaire covers general data about the participants such as age, sex, degree studying or obtained, an assessment of their expertise at gaming and information visualisation applications. Also, the participants filled a Rotter's locus of control [23], a SE and, a SA assessment, as well as a learning style VARK evaluation [79].

All the individual differences tests were performed online prior to the experiment. All psychometric questionnaires were structured as a 5-point Likert scale, asking the participants to assess the accuracy of the statements. These test were taken from the international personality item pool (IPIP) [85] and the three metrics measured were: LoC (20 questions); SE (10 questions) and SA (10 questions), the full set of questions see Appendix 7 for LoC and Appendix 9 for SE and SA. These metrics were chosen in accordance to previous research in the domain [10, 101], and relevance to the research questions. The learning preferences used the VARK model of learning styles. The questionnaire comprised 16 questions where participants had to select none or all answers that best explained their preferences see Appendix 8 for the full list of questions.

Step	Description
	Pre-experiment
1	Send Experiment briefing (by email)
2	Send ethical consent forms
3	Fill Online Participant Pre-study Questionnaire
4	Fill Online Locus of Control Questionnaire
5	Fill Online VARK Questionnaire
6	Fill Online Self-Efficacy and Self-Acceptance Questionnaire
	Experiments
1	Collect and Sign ethical consent forms
2	Interactivity group assignation
3	Read Study experiments briefing
4	Watch introduction of Portal Game
5.a	Perform Short simple Visual Analysis interactive Tutorial (IG)
5.b	Perform Short simple Paper and PDF Data Analysis Tutorial (NIG)
6	Countdown 99-0 in steps of 3 in writing
7	Fill questionnaire NASA-TLX
8.a	Play Flash 2D version of Portal (IG)
8.b	Watch video walkthrough of Flash 2D version of Portal (NIG)
9	Countdown 99-0 in steps of 3 in writing
10	Fill 2D post-experiment Questionnaire
11	Fill questionnaire NASA-TLX
12	BREAK
13.a	Play Flash 2D version of Portal (IG)
13.b	Watch video walkthrough of Flash 2D version of Portal (NIG)
14	Countdown 99-0 in steps of 3 in writing
15	Fill 2D post-experiment Questionnaire
16	Fill questionnaire NASA-TLX
17.a	Perform Interactive VA – Mercer Island data-set (IG)
17.b	Perform Paper and PDF Analysis – Mercer Island data-set (NIG)
18	Countdown 99-0 in steps of 3 in writing
19	Fill questionnaire NASA-TLX

Table 3.2 – Experimental Procedure

3.3 VMI Experiment

In the VMI experiment, interaction is isolated by allowing participants to interact with the visual structures resulting from the visual mapping. Interaction is the independent variable studied to understand the effect it has on the generation of insights, their accuracy and also how it would affect the mental effort required for insight generation. Within this context, this experiment also looks at the conjoint effects VMI has with individual differences studied in a group comparison setting. Figure 3.3 highlights the elements of the information visualisation reference model that are part of the VMI experiment.

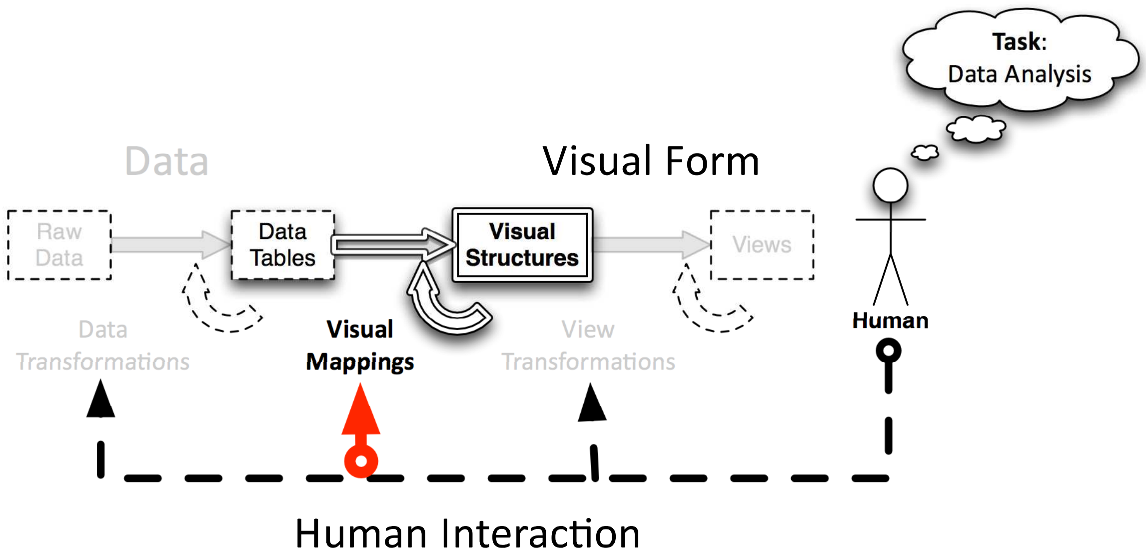


Figure 3.3 – VMI Elements of the Information Visualisation Reference Model

The purpose of this experiment is to address objectives one and two of this thesis. This experiment is based on an insight-based evaluation where insight is the key DV and participants interact with the visual structure in a data analysis task. The data collection concentrates on counting insights, defining their accuracy and categorising them according to the mental effort required to generate them. All participants are profiled according to their individual differences in terms of psychometric measures (LoC, SE and SA) and learning style (VARK). This experimental setting will allow the quantification of the benefits of interaction in terms of insight generation, accuracy and mental efforts required to generate them. Additionally, it is possible through the ATI method assess the compounding effects individual differences may have with interaction.

3.3.1 Dependent Variables

The VA literature characterises insights in terms relevance, predictability, depth, complexity and quality.

The design of this experiment did not intend to measure insight relevance, as this is domain specific, and would have required the recruitment of a sufficiently large pool of experts to validate the experiment.

In terms of predictability, the experiments in this research do not measure this aspect, but equally does not hinder it, as the approach is a time limited open-ended experiment.

Complexity and depth in this investigation are combined into a measure of mental effort derived from the insight provenance.

Finally, quality aspects are taken into account from an accuracy point of view. This research recognises insights in their simplest form, at the lowest level of interpretation using their provenance and mental effort to categorise and count them.

Number of insights

The number of insights were captured by means of an electronic document that the participants filled as they performed their exploration in the data analysis experiment.

Characteristic of insight

Based on the insights captured and recorded by the participants during the data analysis experiment, the insights were categorised according to their provenance using the taxonomy developed by Amar *et al.* [51] in a 'jeopardy' fashion, taking the categorised insight as an answer to the low-level component of analytic activity thus defined. Table 3.3 describes this taxonomy adapted from Amar *et al.* in a 'jeopardy' fashion and are described in more detail using the *pro forma* abstract used by Amar *et al*. The *pro forma* abstract uses three key terms for the consistency, *data case* as entity in the data-set, *attribute* as a value measured for all data cases in the data-set and *aggregation function* as a numeric function that represents a set of data cases (e.g. sum, mean, etc.)

ID	Category	Pro Forma Abstract:	Illustrative Insight Examples (Based on the data analysis data-set)
1	Retrieve Value	What are the values of attributes {X, Y, Z, …} in the data cases {A, B, C, …}?	• The price of a 2-bedroom house on Barlowe Avenue, in December 2009 was $27,000. • The price per square foot of 4-bedroom house on Avalon Road, in January 2008 was $28.
2	Filter	Which data cases satisfy conditions {A, B, C, …}?	• There are only 3-bedroom or smaller houses for less than $80,000. • The most expensive house in the south region is $234,000.
3	Compute Derived Value	What is the value of the aggregate function F over a given set S of data cases?	• The median price of a house in the north region is $113,000. • There were 34 houses sold in the middle region.
4	Find Extremum	What are the top/bottom N data cases with respect to attribute A.	• The most expensive house is $455,000. • The smallest house is 720 square feet.
5	Sort	What is the sorted order of a set S of data cases according to their value of attribute A?	• The regional order by number of sales is South, North and Middle. • 2008 had more sales than 2009.
6	Determine Range	What is the range of values of attributes A in a set S of data?	• In the north region the houses range from 2 to 5 bedrooms. • Overall the properties range from $1 to $533 per square foot.
7	Characterise Distribution	What is the distribution of values of attribute A in a set S of data cases?	• Most properties are on the coast. • The northern region had over 50% of the sales in 2009.
8	Find Anomalies	Which data cases in set S of data cases have unexpected / exceptional values?	• There is 2 houses with 0 bedrooms. • There is one house at $533 per square foot; excluding this outlier the average is $120 per square foot.
9	Cluster	Which data cases in a set S of data cases are similar in value for {X, Y, Z, …}?	• All condominiums are within a 1 square mile radius. • The top 10 most expensive houses are in the Walnut neighbourhood.
10	Correlate	What is the correlation between attributes X and Y over a given set S of data-sets?	• Houses in the southern region are 20% more expensive than the northern region. • January has the highest number of sales.

Table 3.3 – Insight Categories

Mental Effort Based Insights Grouping

Previous research [9–12] has studied iterative procedural versus inferential learning using different visualisation interfaces. Procedural tasks considered from a bottom up perspective were tasks that require little conscious mental effort, due to their automatic, and repetitive nature. Inferential tasks were about drawing conclusion from the data, and require more conscious mental effort, as used in reasoning activities such as induction, deduction, and comparison [17].

Using these definitions, the categories of insight previously defined were grouped into these two parent categories based on the conscious mental effort required to obtain the insights.

ID	Procedural	ID	Inferential
1	Retrieve value	5	Sort
2	Filter	7	Characterise Distribution
3	Compute Derived Value	8	Find Anomalies
4	Find Extremum	9	Cluster
6	Determine Range	10	Correlate

Table 3.4 – Mental Effort Category Grouping

Table 3.4 illustrates the grouping of the different insight categories into procedural or inferential categories

Accuracy of insights

All insights were individually checked for accuracy in the data-set. The accuracy was assesses in a binary manner, thus the answers provided were either accurate or inaccurate within the data-set.

Based on the different categorisations above, the resulting DVs of interest are:

- Total insight
- Total accurate insights
- Total inaccurate insights
- Total procedural insights
- Total inferential insights
- Total accurate procedural insights
- Total accurate inferential insights
- Total inaccurate procedural insights
- Total inaccurate inferential insights
- Overall score

The overall score variable was created in order to assess the overall performance, and was generated using equation (3.1)

$$Overall\ Score = \frac{Accurate}{Accurate + Inaccurate} \cdot 100 \tag{3.1}$$

In order to analyse the individual differences and interaction compounding effects yield, two insight scales were created based on the key DV of insights factors i.e. mental effort (procedural or inferential) and accuracy (accurate or inaccurate). The first scale is based on the accuracy yield defined as:

- Overall accuracy yield
    - total accurate insights – total inaccurate insights
- (Mental Effort Factor) accuracy yield
    - total accurate (Mental Effort Factor) insights
      – total inaccurate (Mental Effort Factor) insights

Where the mental effort factor has two levels (inferential and procedural). The construct of these scales is such that for the accuracy yields, the interpretation is that a negative value means that there were more inaccurate insights than accurate ones. Also, increases in this accuracy yield, signifies an increase in accuracy.

The second scale is based on a mental effort yield defined as:

- Overall mental effort yield
    - Total inferential insights – total procedural insights
- (Accuracy Factor) Mental effort yield
    - total (Accuracy Factor) inferential insights – total (Accuracy Factor) procedural insights

Where the accuracy factor has two levels (accurate and inaccurate). For these mental effort yields, negative values indicate that there were more procedural insights, than inferential ones. An increase in this mental effort yield signifies, that the number of inferential insights has increased. Thus, the interpretation is that higher scores indicate higher mental effort.

Both these scales – accuracy and mental effort, can be considered analogous to a gain in accuracy or mental effort, as the scores increase on the scales.

3.3.2 Experiment Setting

Apparatus

The experiment ran on a Mac Pro, running OSX 10.7.0 and a combination of three screens, 1 LCD and 2 projectors as extended displays. Tableau reader from Tableau Software [102] was used in the data analysis experiment VA for the IG. The NIG had PDF and paper versions of the same visualisations. The screen configuration was the same for both groups. The tableau and the PDF-based visualisations were displayed on the extended projector screens and the insight capturing sheet on the LCD display.

Data-set

The data-set used in the VMI experiment was based on a real estate data for 2008 and 2009 for the island of Mercer, WA (USA). The data points had the following fields:

- Date of purchase
    - Day of the week
    - Day
    - Month
    - Year
- Address
- Island area
    - North
    - Mid
    - South
- Home type
    - Condo
    - House
    - Townhouse
- Sale Price
- Square footage
- Price per square foot
- Number of bedroom
- Number of Bath

The tutorial used another geographical area in the US with the same data fields.

Tableau Reader

Tableau reader from Tableau Software [102] was used for the interactive data analysis. Figure 3.4 provides a screenshot of the Tableau Reader main configuration visualisation tab. The application was configured using six tabs: a main visualisation tab; a median values tab; 2008 data-set tab, 2009 data-set tab; a 2008 *maxima* tab and a 2009 *maxima* tab.

The first tab contained the main visualisations, composed of three geo-located colour-coded maps of the island of Mercer with distinct glyphs for each house type. The colour codes used in the different maps were: sale price; bedroom numbers and bath numbers. This main visualisation tab also contained three bar charts, a line chart and a bubble scatter plot. The first bar chart depicted the number of record by month for 2008-09 on the x-axis and island area and house type on the y-axis, each bar was also colour-coded by sale price. The second bar chart showed the number of records for 2008-09 by day of the week with the days of the week also colour-coded. The third bar chart illustrated the number of record for 2008-09 by month and colour-coded day of the week, it also included a monthly median number aggregated for both years. The line chart illustrated the year to date sales in USD by year and month, colour-coded by island area. Lastly the bubble scatter graph, depicted the sale price on the x-axis and the square footage on the y-axis, the bubble sizes represented the number of baths and the number of bedrooms was colour-coded. For the bubble scatter graph the average for both the x- and the y-axis were represented by a line.

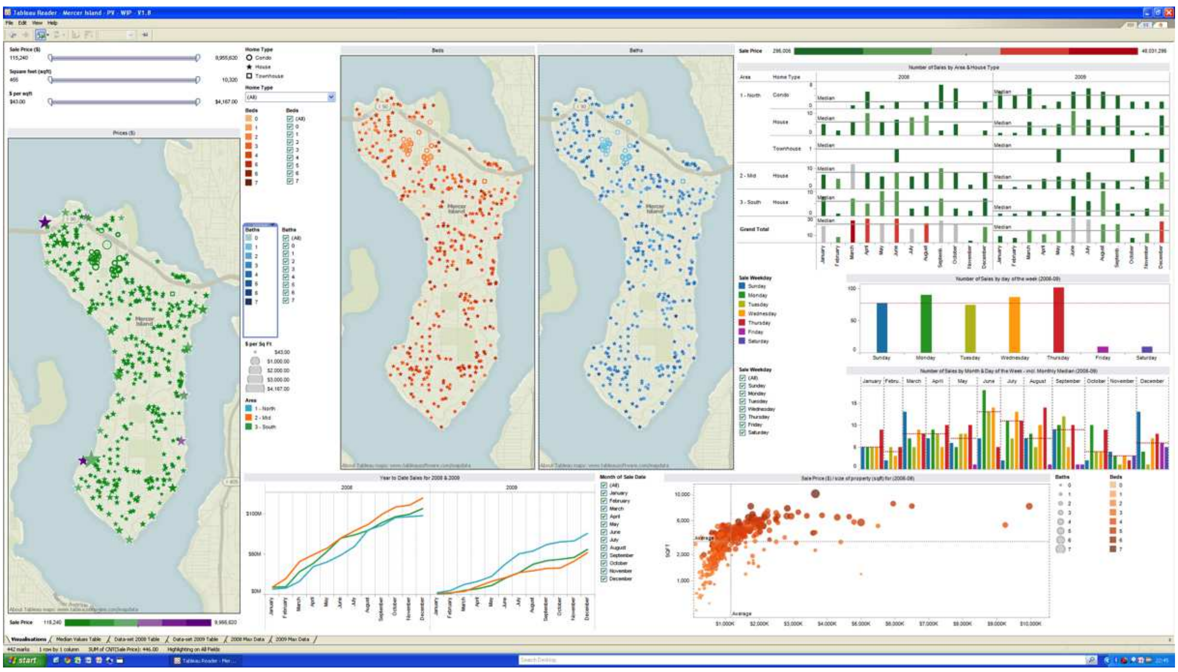


Figure 3.4 – Screenshot of Tableau Reader – Mercer Island Data-set

The second tab contained a larger version of visualisation tan sale price, geo-located colour-coded map of the island and a median values table. The median values table provided the number of records, sale price, square footage and price per square foot by house type and number of bedrooms and baths, categorised by island area with a grand total category as well.

The third and fourth tabs provided respectively all the records for 2008 and 2009. The data were presented as a table with sale price, square footage and price per square foot were grouped by island area and the month, day, day of the week, address, number of bedrooms and baths were simple columns.

The fifth and sixth tab specified respectively the *maximum* value for 2008 and 2009 for the sale price, square footage and price per square foot by month, number of bedrooms and baths and house type.

The interaction configuration settings in tableau for this data-set were configured to enable the highlighting of data (the rest of the data remained grey) by the selecting any field in the legend or tables or selecting an items or a group of items in the visualisations. This highlighting feature would not filter any data out. Tableau was also configured to enable filtering on one or multiple fields, by selecting the categorical data (day of the week, number of bedrooms and baths and house type) fields individually with check boxes or dropdown menu and brushing ranges on the continuous data (sale price, square footage and price per square foot by month). This filtering capability, would change the aggregation of data for the median and average values and the different temporal aggregations such as by month, and day of the week. Also all the tabs were linked when filtering and/or highlighting. The more complex interaction such as clusters, trends and temporal patterns analysis, were inferred by using filtering and highlighting alone.

PDF and Paper

All the visualisations and tables available in Tableau reader were available un-filtered and un-highlighted in PDF. The 2008-09 data-set table and the maxima tables for 2008-09 were also available on paper format.

This non-interactive setting was design to deny the possibility to filter or highlight data. Thus any aggregation such as median or average value calculation temporal aggregation by day of the week or month were not available for sub-sets of the data. Figure 3.5 provides an illustrative sample of the PDF and printed visualisations.

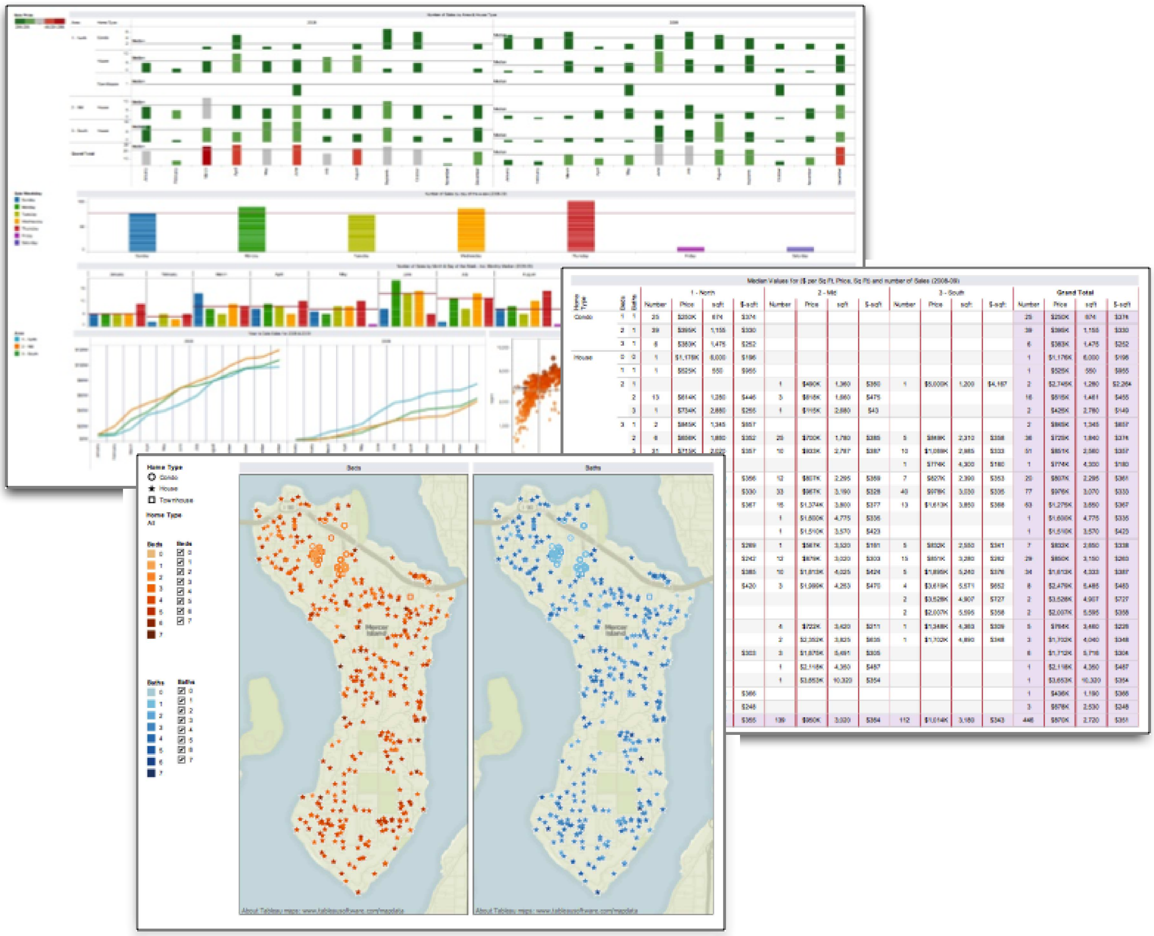


Figure 3.5 – Non-interactive PDF Samples – Mercer Island Data-set

3.3.1 Experiment Procedure

Tutorial

The aims of the tutorial were to familiarise the participants with the apparatus (multiple displays), the visualisation, tables, application (Tableau reader), questionnaires and procedure in general for all experiments for the data analysis task. The data-set used for the tutorial was different than the one used in the data analysis experiment, but both were in the same domain (real-estate) with the same variables and terminology. In the tutorial the IG used Tableau reader from Tableau Software [102] and the NIG used PDF and paper versions of the same visualisations. Both groups used a combination of three screens, 1 LCD and 2 projectors as extended displays. The task objective for the participants in both groups was to find in the visualisation and tables as many of the pre-defined insights provided in a printed grid (see Appendix 4) in 10 minutes. The pre-defined insights grid was designing as a two dimensional matrix of the following rows and columns: quantity and ranges; north, mid and south island areas; sale price; square footage; price per square foot; number of beds and temporal patterns. These predefined knowledge / insight snippets covered all the categories described in Table 3.3, illustrated bellow is an example of this matrix regarding quantity and ranges.

Overall:

- Total of 277 properties sold

Regarding the northern area:

- 65 records in the north
- only 1 condo in the north

Regarding the mid island area:

- 127 records in mid island
- The mid island has more than half the condos of the data-set

Regarding the southern area:

- 85 records in the south

Regarding the square footage:

- The square footage (sq ft) range is 640 – 4,494
- The majority (273 out of 277 records) are less than 2,350 sq ft

Regarding the bedrooms:

- The bedroom range is 0 – 5
- The majority (270 out of 277) is in the 2-4 bedroom range
- There is 5 records with 0 bedrooms
- 3 bedroom houses are the most popular (total 190 out of 277)

Regarding temporal patterns:

- The total weekly median number of records is 14
- January has the highest proportion of 2 bedroom properties

Regarding the price:

- The price range is \$2,500 – \$495,000

Regarding price per square foot (\$/sq ft):

- The range is 1 – 271 \$/sq ft
- 271 properties are less than \$120/sq ft

VMI Experiment

In the VMI data analysis experiment, all participants in both groups had 15 minutes to explore the data for insights. The data-set was based on the island of Mercer, WA (USA) real estate data for 2008 and 2009. The configuration in the experiment was the same as in the tutorial. The IG used Tableau reader from Tableau Software [102] and the NIG used PDF and paper versions of the same visualisations. Both groups used a combination of three screens, 1 LCD and 2 projectors as extended displays. On the LCD screen was displayed the an empty insight capturing matrix with the same design as the pre-defined insights grid used in the tutorial. The task objective for the participants in both groups was to find in the visualisation and tables as many of the insights as possible. These were self-reported by typing their findings into the computer in the empty matrix (see Appendix 5). The visualisations and tables were the same in both experimental settings, this cancelled the significance of visualisation type effect that previous studies have shown [9],[72], creating the same effects for both for treatments.

To motivate their exploration participants were asked to decide where they would like to live or not, based on the data available. The insights were defined to the participants, as insights into the data that they found of interest in terms of the following broad characteristics:

- General
- Detail
- Temporal or other, pattern or clusters

After the exploration task, the participants were asked to count backwards from 0 to 99 in minus three (–3) increments in writing on a pre-printed grid (see Appendix 3). This approach, has been proven to prevent rehearsal [103] and in way act as a short-term memory (STM) dampening.

Finally, the participants were asked to complete a NASA-TLX form, to assess the experiment in terms of workload. The NASA-TLX measured the experiment based on six key metrics, mental, physical, temporal demands and performance, effort and frustration. In the context of the VMI experiment the workload assessment was use to check the validity of the experiments in terms of independence of the interaction treatment as an independent variable. The assumption was that both interaction treatments (interactive and non-interactive) must not have significantly different overall task workload.

Post-experiment, the resulting grid was analysed by counting the insights generated, confirming their accuracy and assigning a task and mental effort category as defined in Table 3.4. This resulting processed data-set, constituted the corpus for the statistical analysis.

3.4 VTI Experiment

In this experiment interaction is isolated from the view transformation point of view. Interaction as an independent variable is investigated to appreciate the effect it has on the gathering of insights and their accuracy. In this experiment the effect of the visual representation is also investigated with the VTI context, thus assessing the performance of the participants in gathering accurate insights. As with the VMI experiment, this experiment also looks at the conjoint effects VTI has with individual differences analysed in a group comparison setting. Figure 3.6 highlights the elements of the information visualisation reference model that are part of the VTI experiment.

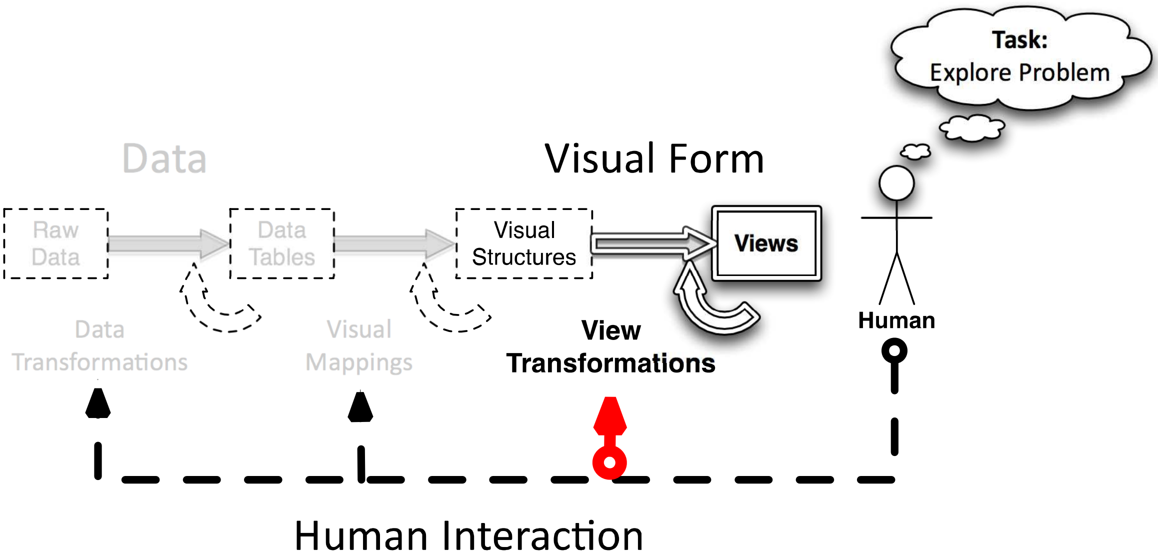


Figure 3.6 – VTI Elements of the Information Visualisation Reference Model

The purpose of this experiment is to address objectives three and four of this thesis. As in the VMI experiment, VTI is also an insight-based evaluation experiment with insight as the main DV. In the VTI setting the insights are collected by means of a post-experiment questionnaire. In the VTI experiment the participants interact by changing the views in the visual structures in order to perform an exploration of the problem-solving data-set using two different visual representations – 2D and 3D. Although the participants performed a problem-solving task, the post-experiment questionnaire evaluated their insights regarding the environment in which their performed their problem solving. The same individual differences profiling used in the VMI experiment is used to assess the compounding effect these have with interaction using the ATI method.

3.4.1 Dependent Variables

The VTI experiment follows a more classical insight-based evaluation experimental setting. In terms of relevance the same issues as VMI apply with regards to the lack of expert access. Hence the use of a game-based simulation enabled the selection of a wider participant pool.

With regards to predictability, this investigation uses a post-questionnaire as the data collection mechanism, and does not measure the predictability aspects of the insights gained.

The complexity and depth are measured by the nature of the questions in the post-experiment questionnaire, relating to the characteristics of the problem solving data-set rather than specifics of the solutions of the problem-solving task.

Finally, quality aspects are taken into account from an accuracy point of view with the addition of undefined category, where participants can opt to answer explicitly that they do not know the answer to the question in the post-experiment questionnaire.

Number of insights

The number of insights were the total number of insights identified by answering the post-experiment questionnaire, which excludes the all insights related to 'don't know' answer category.

Accuracy of insights

The post-experiment questionnaire responses were evaluated for accuracy as accurate or inaccurate insights for each visual representation (4 variables) and additionally two variables (one per representation) with the number of questions they did not know. These variables were normalised as a percentage of the total number of questions. The 2D post-experiment questionnaire had 71 questions and the 3D questionnaire 64. Furthermore, three score variables were created to assess the overall performance, generated using equation (3.2), where i is alternatively the 2D and 3D representation as well as the combined 2D and 3D.

$$Score_i = \frac{Accurate_i}{Accurate_i + Inaccurate_i} \cdot 100 \tag{3.2}$$

The resulting DVs of interest for this analysis were:

- Total percentage of accurate 2D insights
- Total percentage of inaccurate 2D insights
- Total percentage of unidentified 2D insights
- Total percentage of accurate 3D insights
- Total percentage of inaccurate 3D insights
- Total percentage of unidentified3D insights
- Overall percentage of accurate insights
- Overall percentage of inaccurate insights
- Overall percentage of unidentified insights
- 2D Score
- 3D Score
- Overall score (combined 2D and 3D score)

3.4.1 Experiment Setting

Apparatus

The experiment ran on a Mac Pro, running OSX 10.7.0 and used a 27 inch Apple LCD screen. The game-based simulation were a Flash version of the game of Portal from Valve Corporation [104] for the 2D representation and the 3D first person game for the 3D representation.

Game-based Simulations

The choice of the game was guided by the following requirements:

- Interaction centric problem solving
- Minimal interaction noise, meaning that the game environment should not clutter the problem solving process by excessive interactions that were unrelated to the problem resolution.
- System behaviour based problem solving, i.e. the resolution of the problem should be related to understanding the system (environment) behaviour.
- Simple and clear aims and goal

The game of portal from Valve Corporation [104] satisfied these requirements and additionally it had a 2D version implemented in Adobe Flash. One of the benefits of using the portal platform was that both 2D and 3D representations share the same simple rules i.e. a physics simulation with the objective of solving a series of puzzles with the goal of going from a starting point to an ending point. These puzzles were solved, by using a teleporting device and manipulating simple objects. The puzzles used were bespoke, thus still is a puzzle for the experienced participants who had previous experience with the game.

The 3D game takes the form of a 3D single player, first person, action video game and the 2D game is an Adobe flash version of the same game. Figure 3.7 illustrates the simple goal of the game. In this illustration the character needs to cross the gap, it creates a blue entrance portal, then jumps into it and exits through the orange portal at the velocity it entered the blue portal. Thus, obeying the laws of physics the character is projected onto the other side.

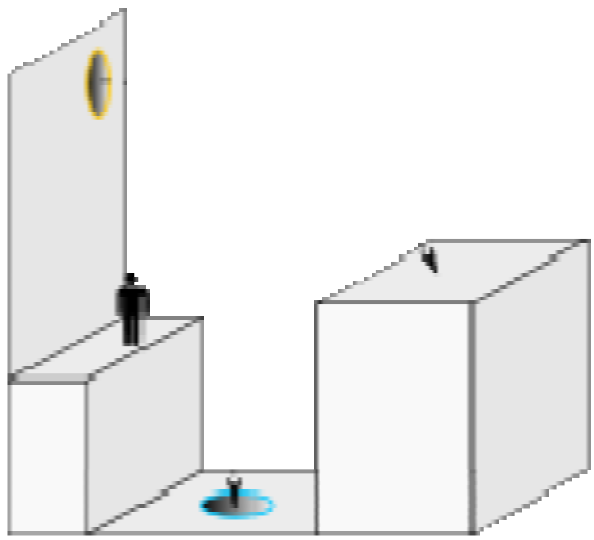


Figure 3.7 – Portal Illustration

The IG played the game, and the NIG watched a video walkthrough of the same level been played.

3.4.2 Experiment Procedure

As illustrated in Table 3.5 the order of the VTI settings regarding the 2D and 3D representation were alternated to minimise confounds transferable learning effects. Thus creating 2 sub-groups within the IG and NIG groups.

Group	1st Experiment	2nd Experiment
Interactive 1	2D Game	3D Game
Interactive 2	3D Game	2D Game
Non-Interactive 1	2D Game Walkthrough	3D Game Walkthrough
Non-Interactive 2	3D Game Walkthrough	2D Game Walkthrough

Table 3.5 – Order of View Transformation Experiments 2D and 3D Representation

In the 2D Flash version of portal experiment, the participants experienced 9 levels of the game. The IG were given the instruction to go to the exit of the level to complete it, by means of using the teleporting device and objects such as cubes, when and if appropriate. Figure 3.8 provides a screenshot of the 2D Portal game.

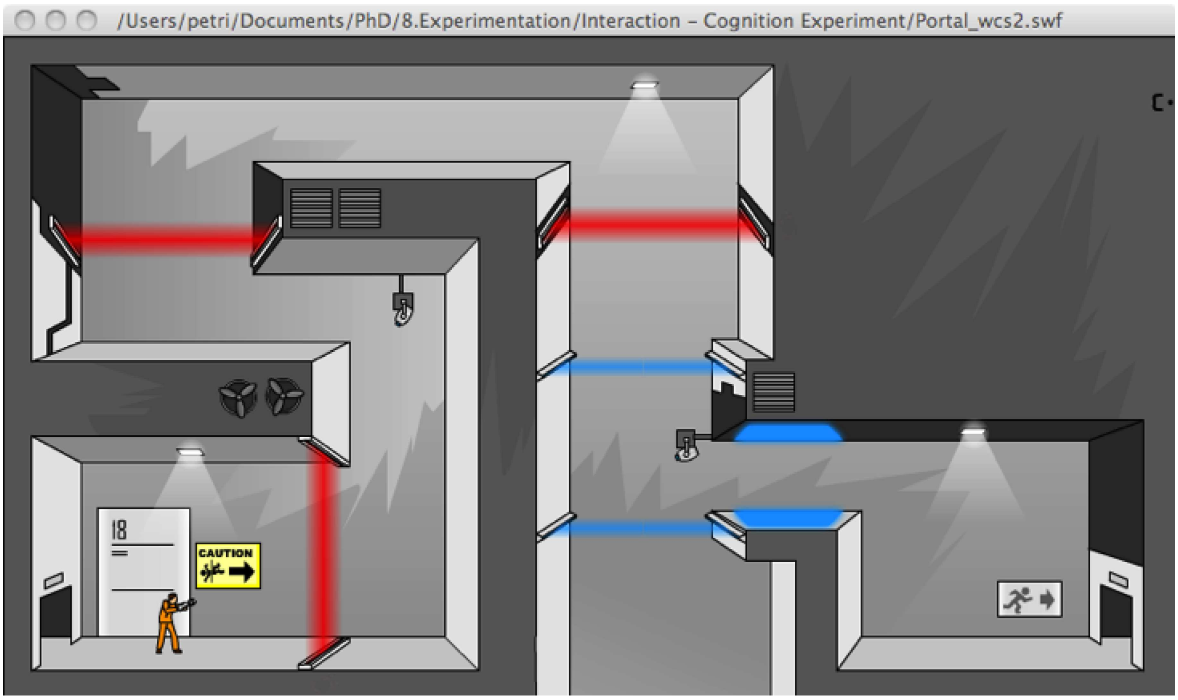


Figure 3.8 – Screenshot of 2D Portal Game

There was a time limit of 2 minutes per level; the time was carried over if the level was performed quicker. This time limitation was in place in order for the interactive participants to experience every level.

For participants who were not familiar with gaming in general, they were given practical assistance. The protocol for this practical assistance was, that they would ask an action centric question, then the researcher would answer briefly, by providing the keyboard and mouse combination needed to perform the requested action.

The participants were made aware that the emphasis of the experiment was for them to experience all the levels, and that actual performance or ability to solve the puzzle was not measured. The emphasis was on exploring the problem space, the environment and related features.

In the 3D portal experiment, the participants experiences 3 levels of the game. This time the time limit was 4 minutes. For the participants inexperienced in gaming, the same practical assistance as in the 2D experiment was given. Also in this experiment the emphasis was on exploring the problem space, the environment and related features. Figure 3.9 provides a screenshot of the 3D Portal game.

Figure 3.9 – Screenshot 3D Portal Game

The NIG in both the 2D and 3D experiments watched a video of the same levels of the game being played to completion. They were asked to observe the not only how the puzzles were solve but also the problem set and environment.

After each experiment similarly to the VMI experiment, the participants were asked to count backwards from 0 to 99 in minus three (–3) increments in writing on a pre-printed grid to prevent rehearsal.

Next, they were asked to fill a questionnaire about the systems behaviour and observations. This approach assessed the participants insights in terms of breadth vs. depth as defined by Saraiya *et al.* [29, 31] and North [33]. Participants were encouraged to give their best guesses, if they did not know the answer, alternatively, they could choose to select the ‘don’t know’ option. Figure 3.10 provides an example of the kind of questions the participants were asked Appendices 12 and 13 provide respectively the full 2D and 3D questionnaire.

What color is the energy ball target when activated?
Give best guess answers, if not sure.

- Orange
- Green
- Random
- Blue
- Other:

About cubes. Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Can go thought portals.			
Need to carried by character to go through portals.			
Can be held by character whilst creating a portal (with a beam).			
Create a challenging hazard.			
Can be pushed.			
Can be placed on buttons without been carried.			
Hold buttons.			
Enable character to reach higher.			
Enable portals to stay open.			
Open gates.			
Allow character to speed up.			
Protect character from harm.			

Figure 3.10 – Sample 2D and 3D Post-experiment Questionnaire

Finally, the participants were asked to complete a NASA-TLX form, to assess the experiment in terms of workload, measuring the experiment on six key metrics mental, physical, temporal demands and performance, effort and frustration. In the context of the VTI experiments the workload assessment is used to check the validity of the experiment in terms of independence of the interaction treatment as an independent variable. For the VTI experiment, the investigation analysis is centred on the post-experiment questionnaire; the important aspect in this experiment setting is that the performance in these post-experiment questionnaires is comparable. The assumption is that the performance workload element of the scale is not significantly different for both interaction treatment groups. The overall task workloads is expected to be significantly different as the tasks are clearly of unequal difficulty (i.e. the video walkthrough a lot easier that playing the game), but when answering the post experiment questionnaire, the participants must have an equal workload in terms of the performance workload.

3.5 Statistical Analysis

All the experiments were analysed using IBM SPSS Statistics Version 19.0.0, figures and tables were created using JMP 10.0.0 from SAS and Microsoft Excel 14.2.1. All programs ran under Mac OSX version 10.7.4. For the data collection Google Docs Forms was used and Microsoft Excel VBA macros and in cell formulae were used to calculate the different derived data used in the analysis such as the NASA-TLX workloads and the various individual differences scores.

To choose the appropriate statistical analysis method, the decision trees illustrated in Figure 3.11 and Figure 3.12 were used. These decision trees are based on the number of independent variables, number of DVs, factor design (related or unrelated) and whether the DVs meet the parametric test assumptions.

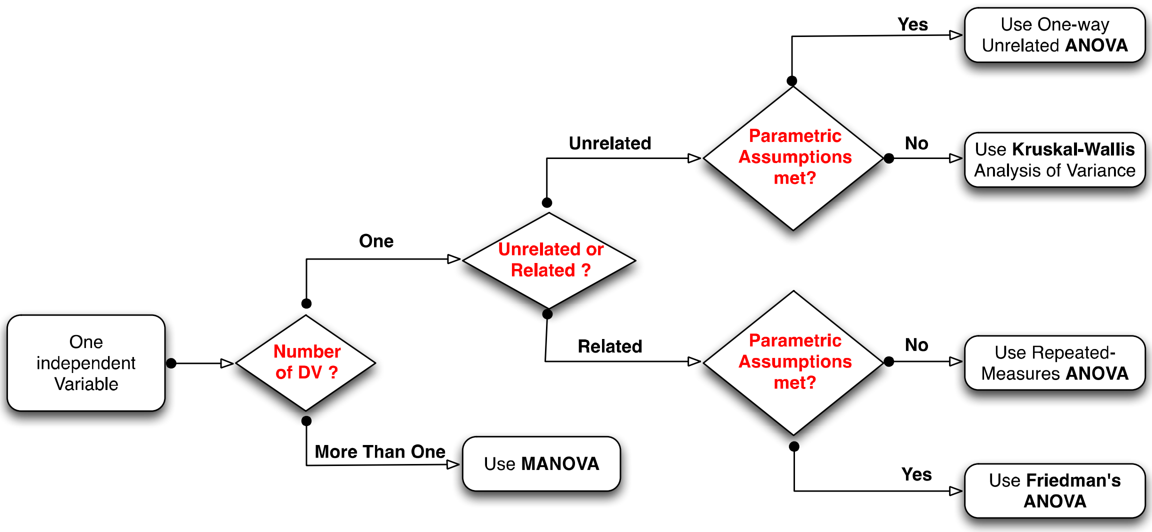


Figure 3.11 – Statistical Analysis Decision Tree for One Independent Variable Adapted from [105]

The factor design, refers to the participants and the independent variable. Unrelated factor design means that the DV analysed were collected for different participants and for the related design the same participants were used.

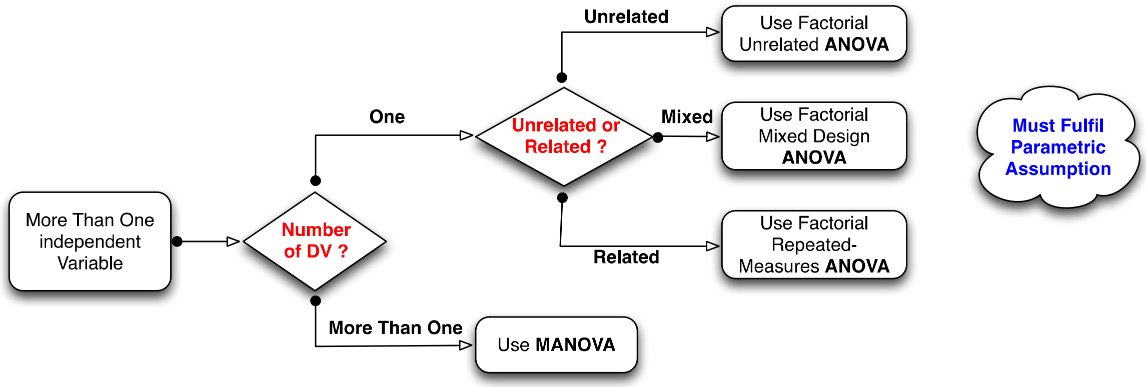


Figure 3.12 – Statistical Analysis Decision Tree for More Than One Independent Variable Adapted from [105]

The assumption of parametric test are:

- Normally distributed data
- Homogeneity of variance
- At least interval data
- Independence
- Homogeneity of covariance (MANOVA only)

All DVs are independent and can be considered interval data where intermediate values have a meaningful interpretation. To test normality the Kolmogorov-Smirnov (K-S) test was used, where normality is asserted by non-significant results (i.e. p ≤ .05). To test homogeneity of variance or heteroscedasticity assumptions, levene's test was used, where the DVs have homogenous variances when the test is not significant (i.e. $p \leq .05$). For the MANOVA the additional assumption is the homogeneity of covariance tested by a Box's M test of equality of covariance matrices, where homogeneity of covariance is asserted by a non-significant results at $p \leq .001$ [106]. MANOVA and ANOVA are fairly robust to violation of normality [107], thus transformations were used on the data, in an effort to meet the parametric assumption, when and only if the homogeneity of variance was also violated. Furthermore, the residuals of the analysis were tested for normality using the K-S test, thus when the residuals were normal, this reaffirmed the results [108, 109] in particular when there was an acceptable violation of normality.

When the violations of the parametric assumption remained post transformation, then a non-parametric equivalent test (Kruskal-Wallis test or Friedman's ANOVA) was performed on the untransformed data on an exploratory analysis basis without correcting significance p-values [110, 111].

All results are reported at a p-value of $p < .05$. The parametric test *post-hoc* analysis, had a Bonferroni correction applied to protect from Type I error, so that all effects were then reported at a k^{th} level of p significance, where k is the number of post-hoc test executed. The ANOVA analysis *post-hoc* tests are t-tests and for the MANOVA the post-hoc analysis is done using univariate ANOVAs. Other *post-hoc* test included a non-parametric Jonckheere's test to uncover any significant trends in relation to the individual differences were applicable, this trend test was used for all significant parametric and non-parametric results.

Effects sizes were reported using pearson's correlation coefficient r where $r > .10$ is a small effect, $r > .30$ is a moderate effect and $r > .50$ is a large effect [112, 113].

The calculation of the effect size varies with the analysis, for ANOVAs r is calculated using equation (3.3), where F is the probability distribution defined as the ratio comparing the variance between groups and within groups and df_R is the degrees of freedom of the residuals. To calculate the effects for the t-test the equation (3.4) where t is the t statistic. For the non-parametric tests the effect sizes are calculated using equation (3.5), where z is the z-score and N the number of records in the sample.

$$r = \sqrt{\frac{F(1, df_R)}{F(1, df_R) + df_R}} \tag{3.3}$$

$$r = \sqrt{\frac{t^2}{t^2 + df_R}} \tag{3.4}$$

$$r = \frac{z}{\sqrt{N}} \tag{3.5}$$

Table 3.6 specifies the different statistical analysis used to investigate the various research questions in this thesis.

Objective One: Investigate the effects of Visual Mapping Interaction in the context of performing an analytical task using information visualisation.	
Research Question 1: Does Visual Mapping Interaction affect the number of insights generated and their accuracy, when compared to an equivalent non-interactive task?	**Analysis:** Oneway ANOVA for each DV
Research Question 2: When insights are categorised based on mental effort (inferential for high and procedural for low mental effort), does Visual Mapping Interaction have an effect on the number and accuracy of insights generated in each mental effort category?	**Analysis:** Oneway ANOVA for each DV
Objective Two: Investigate the compounding effects of Visual Mapping Interaction with performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles in the context of performing an analytical task using information visualisations.	
Research Question 3: Do Locus of Control, Self-efficacy, and Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences have compound effects with Visual Mapping Interaction, whereby according to the level of the different measures, there will be a significant effect on the generation of insights and their accuracy?	**Analysis**: MANOVA
Research Question 4: When categorising insights based on mental effort, do individual differences (LoC, SE, SA, V, A, R, and K) have a compounding effect with VMI with regards to the generation and accuracy of insights?	**Analysis**: MANOVA
Objective Three: Investigate the effects of View Transformation Interaction in the context of a problem-solving task, where View Transformation Interaction is used to explore the problem data-set using a game-based simulation using a 2D and 3D visual representation.	
Research Question 5: Does View Transformation Interaction affect the number and accuracy of insights identified in a problem data-set represented in a game-based simulation, when comparing to an equivalent non-interactive task?	Analysis: MANOVA
Research Question 6: Does the view representation – 2D and/or 3D, have an effect on the number and accuracy of insights into a problem data-set represented in a game-based simulation?	**Analysis:** Repeated Measures ANOVA
Research Question 7: Does the representation – 2D and/or 3D and View Transformation Interaction have an interaction effect with regards to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?	**Analysis:** Repeated Measures Factorial ANOVA

Table 3.6 – Research Objectives and Question with Associated Statistical Analysis

Objective Four: Investigate the compounding effects of View Transformation Interaction with performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles; in the context of a problem-solving task, where View Transformation Interaction is used to explore the problem data-set using a game-based simulation using a 2D and a 3D visual representation.	
Research Question 8: Independently from the representation, do Locus of Control, Self-efficacy, Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences have a compound effects with View Transformation Interaction, whereby according to the level of the different measures, there will be a significant effect to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?	Analysis: MANOVA
Research Question 9: Does the representation – 2D and/or 3D and View Transformation Interaction have compounding effects with Locus of Control, Self-efficacy, Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences, whereby according to the level of the different measures, there will be a significant effect to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?	**Analysis:** Repeated Measures Factorial ANOVA

Table 3.6 – Research Objectives and Question with Associated Statistical Analysis (Cont.)

3.5.1 Power Analysis

This thesis investigates a variety of research questions using different statistical analysis approaches; therefore several *a priori* power calculations were performed to define the suitable sample size. All power calculation were done using G*Power version 3.1.4. [114].

Oneway ANOVA

In the VMI experiment when comparing two groups (interactive and non-interactive) with a power = 80%, $\alpha = .05$ in order to detect a large size effect f = .40 the sample size required was 52 and for a medium size effect f = .25 the sample was 128.

MANOVA

For the VMI experiment the there was 5 DV's and 28 different groups (7 individual differences measures x 2 ATI levels x 2 interactivity treatments) with a power = 80%, $\alpha = .05$ in order to detect a large size effect $f^2 = .35$ the sample size required was 56 and for a medium size effect $f^2 = .15$ the sample was 84.

For the VTI experiment the there was 4 DV's and 28 different groups with a power = 80%, $\alpha = .05$ in order to detect a large size effect $f^2 = .35$ the sample size required was 56 and for a medium size effect $f^2 = .15$ the sample was 112.

Repeated Measures ANOVA

Research question six compares two within factors measures (2D and 3D) with a power = 80%, $\alpha = .05$ in order to detect a large size effect *f* = .40 the sample size required was 15, for a medium size effect *f* = .25 the sample was 34, and for a small size effect *f* = .10 the sample was 199. These results assume a correlation amongst measures of *r* = .50 as the actual value is unknown before the experiment was performed.

Repeated Measures Factorial ANOVA

In the VTI experiment, when comparing two within factors measures (2D and 3D) and 28 different groups with a power = 80%, $\alpha = .05$ in order to detect a large size effect *f* = .40 the sample size required was 84, and for a medium size effect *f* = .25 the sample was 112, These results assume a correlation amongst measures of *r* = .50 as the actual value is unknown before the experiment was performed.

[Intentionally left blank]

Chapter 4. Population Sample Characteristics and Preparation

The first section of this chapter describes the participants' demographics and previous experience with the experimental setting. Then this chapter describes the participant's individual difference distributions and the ATI profiling used in the analysis. After that a description is given of the insight characterisation and grouping for the VMI experiment as well as the data collection questionnaire of the VTI experiment.

4.1 Participants 82

4.1.1 Description 82

4.1.2 2D and 3D Previous Experience 82

4.1.3 Individual Differences 83

4.1.4 ATI Methodology Profiles 87

4.2 VMI Experiment Data Preparation 90

4.2.1 Insights Categorisation 90

4.2.2 Mental Effort Based Insights Grouping 91

4.3 VTI Experiment Data Preparations 93

4.3.1 Post-Experiment Questionnaire 93

4.1 Participants

4.1.1 Description

The study was composed of 44 participants, 25 males and 19 females; age range was 18 to 56 years with a median value of 24 years old. Most were students (31), 17 undergraduates and 14 postgraduates spread across 14 departments. The remainder (13), were university staff and professionals. The group allocation was assigned randomly into an interactive and a non-interactive group. The procedure was as follows; the first male and the first female participants were assigned to the IG. Then, the subsequent participants were assigned to alternate groups of NIG and IG by gender. The participants booked their participation slots online, without prior knowledge of the allocation procedure. This resulted in two evenly split groups of 22. The IG was composed of 13 males and 9 females, and the control NIG had 12 males and 10 females. During the experiment 2 participants (1 male in the non-interactive track and 1 female in the interactive track), did not complete the task finding it too difficult. Subsequently, the analysis considers 42 participants, removing these 2 participants who did not complete the experiment from the sample, thus creating two groups with 21 participants.

4.1.2 2D and 3D Previous Experience

When examining participant's previous experience with 2D gaming, the data shows that overall 92.9% of participants (39 participants) had previous experience with 2D games. The split by group was 90.5% for the NIG (19 participants) and 95.2% for the IG (20 participants). For the 3D gaming experience, overall 52.4% (22 participants) had previous experience. The group split was 61.9% for NIG (13 participants) and 42.9% for IG (9 participants), there is a 19% difference between NIG and IG groups in experience, but fortunately in favour of the NIG, who will not be interacting in the 3D representation, and thus will not bias the interaction findings. When asking specifically whether participants had played the 2D and/or 3D Portal; the data shows that overall 81.0% of participants had no previous experience of either the 2D or 3D Portal games (32 participants), only 4 participants (9.5%) had experience with the 2D Portal game, 9 (21.4%) had experience with the 3D Portal game and 3 (7.1%) of both 2D and 3D Portal games, all of which were in the NIG. When interpreting the results, this potential bias was taken into consideration.

4.1.3 Individual Differences

The participants' locus of control (LoC) *score*s ranged was from 55 to 90 out of a theoretical range of 20 to 100, the mean was 74.24 ($SD = 9.09$); this was inline with previous studies of this kind [12]. The Cronbach's alpha reliability estimate for the LoC in our sample was $\alpha = 0.816$. SE measure scores ranged from 26 to 48 out of a theoretical range of 10 to 50, the mean was 38.07 ($SD = 5.38$). The Cronbach's alpha reliability estimate was $\alpha = 0.742$. SA measure scores ranged from 21 to 48 out of a theoretical range of 10 to 50, the mean was 38.71 ($SD = 5.40$). The Cronbach's alpha reliability estimate was $\alpha = 0.825$. All Cronbach's alpha reliability [86] estimates were above 0.7, which constitutes a good reliability level [115]. Table 4.1 provides the description statistics by interaction grouping of the different individual differences.

	LoC	SE	SA	V Score	A Score	R Score	K Score
Non-Interactive-Group							
Median	74	37	38	7	8	8	8
Mean	72.62	36.95	38.29	6.76	7.57	7.86	8.10
Std Dev	8.74	5.72	6.57	2.66	2.96	3.21	3.00
Min	59	26	21	2	2	2	3
Max	86	48	48	11	12	15	13
Interactive-Group							
Median	77	38	40	8	8	9	8
Mean	75.86	39.19	39.14	7.62	7.76	8.48	7.76
Std Dev	9.36	4.88	4.03	2.64	2.70	2.54	2.28
Min	55	30	31	3	3	4	4
Max	90	48	45	12	13	13	11
Overall							
Median	77	38	39.5	7	8	8	8
Mean	74.24	38.07	38.71	7.19	7.67	8.17	7.93
Std Dev	9.09	5.38	5.40	2.65	2.80	2.88	2.64
Min	55	26	21	2	2	2	3
Max	90	48	48	12	13	15	13

LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **V**: Visual, **A**: Aural, **R**: read-write, **K**: Kinaesthetic

Table 4.1 – Individual Differences Description Statistics

Regarding the learning preferences, when we consider each of the VARK questions as a 5-point Likert scale, where 0 was no modality chosen and 1 to 4 was one to four modalities, the Cronbach's alpha reliability estimate for the VARK questionnaire in our sample was $\alpha = 0.824$, which was a good reliability level.

Table 4.1 gives the arithmetic description statistics for the VARK scores by modality for the sample, where V, A, R, and K were the learning preferences as described in section 0. The procedure used to determine the participants individual learning preferences was described by Fleming as the standard scoring [116]. The learning profile was determined based on the overall sum of scores for all modalities and a lookup table that defines a stepping distance. The individual scores were sorted in descending order and profile was established by calculating the difference between the 1st and 2nd score then 2nd and 3rd and lastly 3rd and 4th, if the difference was smaller or equal to the stepping distance both modalities were retained, the 2 consecutive score difference was greater that the stepping distance. See Appendix 10 for a copy of the official procedure and the lookup table.

Example:

Visual (V)	Aural (A)	Read-write (R)	Kinaesthetic (K)	Total
8	4	5	9	26
2nd	4th	3rd	1st	

Table 4.2 – VARK Profiling Example

The stepping distance for a total score of 26 is 2, according to the official scoring (Appendix 10). $K_{Score} - V_{Score} = 1$, is < 2, K and V were retained. $V_{Score} - R_{Score} = 3$, is > 2, that is the end of the process.

The single modality preference was defined as the highest modality score when the difference between the first and the second highest scores was greater than the stepping distance. Additionally a strength value (Mild, Strong and Very Strong) was defined for the uni-modal preferences by another lookup table based also on the total sum of scores see Appendix 10. Table 4.3 outlines all the possible profiles.

Type of Preference	VARK Profile		Number of Profiles
Single preferences	Visual Aural Read-write Kinaesthetic	- Mild, Strong and Very Strong. - Mild, Strong and Very Strong. - Mild, Strong and Very Strong. - Mild, Strong and Very Strong.	12
Bi-modal preferences	VA, VR, VK, AR, AK and RK		6
Tri-modal preferences	VAR, VAK, ARK and VRK		4
All four modes preferred	VARK		1
	Total		23

Table 4.3 – VARK Profiles

Out of the possible 23 profiles as illustrated in Table 4.3, the participants were distributed into 12 categories outlined in Table 4.4 of which 50% fall into the all four modes VARK profile. Further, Table 4.4 describes the preferences frequencies and percentages for each modality blending all the strengths into one modality frequency total.

Type of Preference	Preference	VARK Profile	N	% of Total
Single preferences	V	Mild V	2	4.8%
	A	Strong A	1	2.4%
	R	Mild R Strong R	3 1	9.5%
	K		-	0%
Multi-Modal preferences	Bi-modal	VA	2	83.3%
		VR	-	
		VK	1	
		AR	1	
		AK	-	
		RK	-	
	Tri-modal	VAR	-	
		VAK	1	
		ARK	3	
		VRK	1	
	All four modes	VARK	26	
		Total	42	

N: Number of participants, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table 4.4 – VARK Preference Frequencies

The distribution of the profiles in this sample was very different from the overall distribution found in Fleming's research [117], N = 62094 used as the baseline illustrated in Figure 4.1.

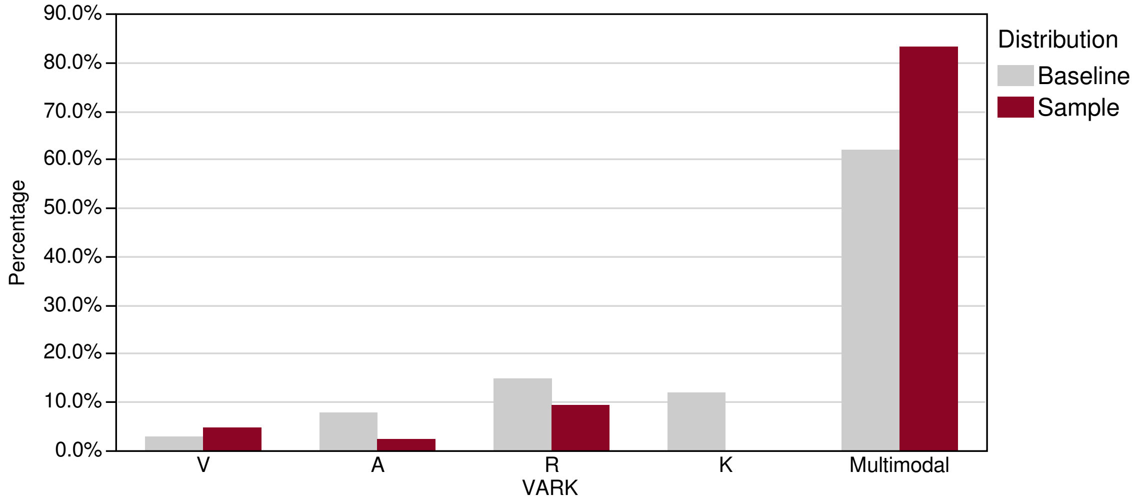


Figure 4.1 – VARK Preferences Distributions

In order to compare this study's sample, Fleming proposed a reduction in complexity, by defining profiles with some modality, thus creating a uni-modal distribution. The key assumption was that when combined, the individual preferences get weaker and that all modalities contributions sum was 1. This approach gives a score of 1 to single preferences, half to bi-modal preferences, a third for tri-modal and a quarter to all four modes profiles. Table 4.5 outlines the resulting simplification.

Learning Preference	Total	%
Visual (V)	10.67	26.5%
Aural (A)	10.33	26.5%
Read-write (R)	12.33	27.7%
Kinaesthetic (K)	8.67	19.3%
Total	42	

Table 4.5 – Simplified VARK Preference Frequencies

Figure 4.2 illustrates the resulting simplified distribution for this study, which was more closely aligned with Flemings simplified results for N=62094 [117] baseline. Thus, findings indicate a higher level of confidence for the analysis of the profiles effects in this study.

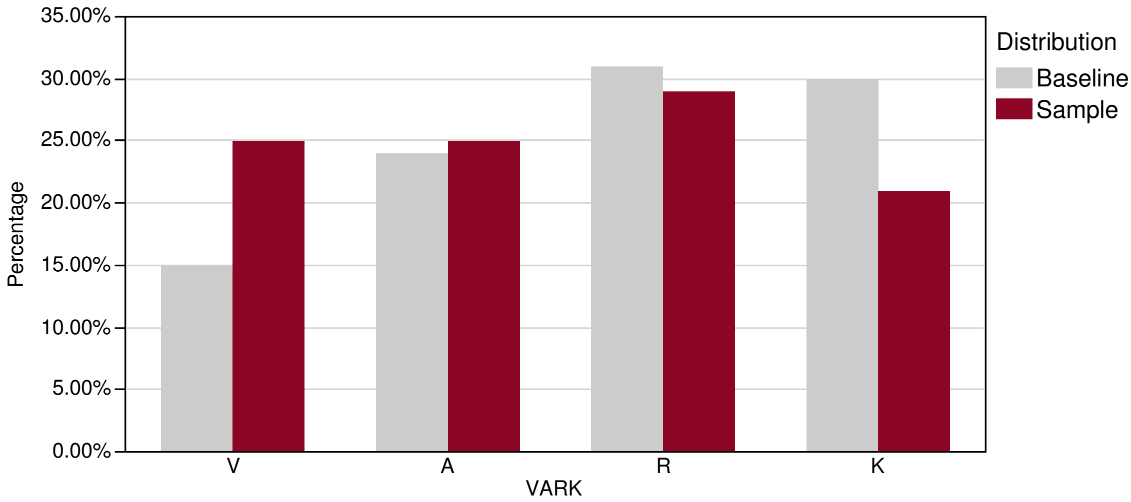


Figure 4.2 – Simplified VARK Preferences Distributions

4.1.4 ATI Methodology Profiles

In order to prepare the sample for the ATI approach and considering the size of the sample (*N* = 42) all the individual differences were classified as above and below mean. For Locus of Control (LoC) measurement, this dichotomy further signifies that the participants with below mean measures have a tendency for an external LoC, and the participants with above mean measures were more internally focused LoC. For all other individual differences (SE, SA, V, A, R, and K), these were simply referred as high and low levels of the measurement. The outcome of this categorisation was captured in Table 4.6, where it can be observed that the different cells are fairly even in size. In the rest of the thesis this categorisation will be referred as ATI categorisation (ATI-C), further using ATI-C (X) format where X is the individual difference (e.g. ATI-C (R), is the above / below mean categorisation for the read-write learning preference).

	Non-Interactive-Group		Interactive-Group		Overall	
	Below Mean	Above Mean	Below Mean	Above Mean	Below Mean	Above Mean
LoC	11	10	9	12	20	22
SE	12	9	11	10	23	19
SA	11	10	8	13	19	23
V	13	8	9	12	22	20
A	8	13	9	12	17	25
R	14	7	10	11	24	18
K	8	13	9	12	17	25

LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table 4.6 – Individual Differences ATI Categories Participants Numbers (N)

When parametric test assumptions were violated and a non-parametric test was available, the individual differences for ATI were categorised into high, mid-level and low groups. In order to have the most balanced sample sizes per cell possible, the data-set as a whole was divided according to their distance from the mean. Participants with scores more than one-half standard deviation from the mean were classified as high, and those with scores less than one-half standard were classified as low. The rest of the participants were classified as mid-level. Thus, the medium category represented 34% of the sample, and the high and low categories 33%. The resulting split was defined in Table 4.7 and the outcome by category and interactivity group was outlined in Table 4.8.

Furthermore, the individual differences were analysed into two distinctive groups one composed of psychometric measures LoC, SE, and SA and the other composed of the VARK learning preferences. This is justified by looking at Table 4.9 describing the correlations between the different individual differences, where two groups can be clearly differentiated. One group composed of LoC, SE and SA were all significantly inter-correlated $p < .01$ and a second group made of the learning preferences score (V, A, R, and K), of which most are uncorrelated (only the visual preference was significantly correlated to the kinaesthetic preference $r = .35$, $p = .05$).

Measure	Low (≤ M – half SD)	Mid-level (M – half SD < & < M + half SD)	High (≥ Mean + half SD)
LoC Scale	(External) < 69.70	≥ 69.70 & < 78.79	(Internal) ≥ 78.79
SE Scale	< 35.38	≥ 35.38 & < 40.76	≥ 40.76
SA Scale	< 36.01	≥ 36.01 & < 41.41	≥ 41.41
Visual Preference Scale	< 5.87	≥ 5.87 & < 8.52	≥ 8.52
Aural Preference Scale	< 6.27	≥ 6.27 & < 9.07	≥ 9.07
Read-write Preference Scale	< 6.73	≥ 6.73 & < 9.61	≥ 9.61
Kinaesthetic Preference Scale	< 6.61	≥ 6.61 & < 9.25	≥ 9.25

M: Mean, **SD**: standard deviation

Table 4.7 – Individual Differences – Low, Mid and High Grouping

Groups	Non-Interactive			Interactive			Totals		
	Low	Mid	High	Low	Mid	High	Low	Mid	High
LoC Scale	8	7	6	6	10	5	14	17	11
SE Scale	7	7	7	7	8	6	14	15	13
SA Scale	7	7	7	5	9	7	12	16	14
Visual Preference Scale	7	9	5	5	8	8	12	17	13
Aural Preference Scale	8	6	7	5	10	6	13	16	13
Read-write Preference Scale	8	7	6	4	8	9	12	15	15
Kinaesthetic Preference Scale	6	9	6	5	7	9	11	16	15

Table 4.8 – Number of Participants by Low, Mid and High Grouping by Individual Difference

	LoC	SE	SA	V	A	R	K
LoC	1						
SE	.43**	1					
SA	.40**	.62**	1				
V	.03	.03	-.05	1			
A	.03	-.14	-.03	.07	1		
R	.02	.15	-.06	.16	.16	1	
K	.06	.06	.24	.35*	.29	.29	1

*p < .05, ** p < .01
LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table 4.9 – Individual Differences Pearson's Correlations

4.2 VMI Experiment Data Preparation

4.2.1 Insights Categorisation

In this experiment, insights were captured using an electronic document. Participants captured their insights as they performed their exploration in the data analysis experiment. The results were analysed post-hoc. The insights were assessed for their accuracy and categorised according to their provenance. The categorisation uses the taxonomy developed by Amar *et al.* [51] in a 'jeopardy' fashion (what question does this insight answer?) as described in the characteristic of insight part of section 3.3. In this taxonomy the answers were the low-level component of analytic activity thus defined, and therefore captured as low-level insights. Table 4.10 describes these insights categories into further details with the associated pro-forma abstract question. Note that, when determining the insight accuracy, for the category *determine range (ID-6)*, if the range was half accurate (i.e. one of the extremities was accurate), then half a score of low-level insight was allocated.

ID	Category Description	Pro-Forma Abstract
1	Retrieve value	What are the values of attributes {X, Y, Z…} in the data cases {A, B, C…}?
2	Filter	Which data cases satisfy conditions {A, B, C…}?
3	Compute Derived Value	What is the value of the aggregate function F over a given set S of data cases?
4	Find Extremum	What are the top/bottom N data cases with respect to attribute A.
5	Sort	What is the sorted order of a set S of data cases according to their value of attribute A?
6	Determine Range	What is the range of values of attributes A in a set S of data?
7	Characterise Distribution	What is the distribution of values of attribute A in a set S of data cases?
8	Find Anomalies	Which data cases in set S of data cases have unexpected / exceptional values?
9	Cluster	Which data cases in a set S of data cases are similar in value for {X, Y, Z…}?
10	Correlate	What is the correlation between attributes X and Y over a given set S of data-sets?

Table 4.10 – Insight Categories

4.2.2 Mental Effort Based Insights Grouping

Previous research studies [9–12] have investigated iterative procedural, versus inferential learning using different visualisation interfaces. Procedural task were considered from a bottom up perspective, as tasks that require little conscious mental effort, due to its automatic, and repetitive nature. Whereas, inferential tasks, were about drawing conclusion from the data, and require more conscious mental effort, using reasoning activities such as induction, deduction, and comparison [17].

Based on these definitions, the categories of insight previously defined were grouped into these two parent categories. This categorisation was based on the conscious mental effort required to obtain the insights, Table 4.11 illustrates the resulting grouping as procedural or inferential categories.

ID	Procedural	ID	Inferential
1	Retrieve value	5	Sort
2	Filter	7	Characterise Distribution
3	Compute Derived Value	8	Find Anomalies
4	Find Extremum	9	Cluster
6	Determine Range	10	Correlate

Table 4.11 – Mental Effort Category Grouping

Table 4.12 outlines the number of insights in this resulting categorisation, and by interaction grouping (non-interactive or interactive), accuracy (accurate or inaccurate) and mental effort (procedural or inferential).

	Non-Interactive-Group		Interactive-Group		Overall	
Category ID	**Inaccurate**	**Accurate**	**Inaccurate**	**Accurate**	**Inaccurate**	**Accurate**
Procedural Insights						
1	2	29	0	30	2	59
2	6	1	20.5	1	26.5	2
3	4	30	3	67.5	7	97.5
4	1	11	7	32	8	43
6	14	16	10	14	24	30
Category Sub-Totals	27	87	40.5	144.5	67.5	231.5
Procedural Sub-Totals	114		185		299	
Inferential Insights						
5	0	5	0	0	0	5
7	1.5	32.5	8	45	9.5	77.5
8	2	5	1	14	3	19
9	0	1	0	1	0	2
10	2	1	4	3	6	4
Category Sub-Totals	5.5	44.5	13	63	18.5	107.5
Inferential Sub-Totals	50		76		126	
Category Totals	32.5	131.5	53.5	207.5	86	339
Group Totals	164		261		425	

Table 4.12 – Mental Effort Category Grouping Detail

A preliminary analysis shows that insights were overall distributed into 70% procedural versus 30% inferential. Also overall, the IG generated 61% of the insights. Additionally accurate insight represented 80% of the total and the same proportion was maintained across IGs. Whereas for the accurate insights, the inferential insight seemed to be more accurate as a whole. The proportion was 77% and 85% for the procedural and inferential insights respectively. When looking at the group crossovers in terms of total insights, the split between interaction groups by mental effort type was roughly the same as a 60/40 split. For the procedural mental effort the split was 62% (IG) and 38% (NIG) and for the inferential mental effort, 60% (IG) and 40% (NIG). Whereas, the accuracy split was slightly different, the accurate insights were 76% (NIG) vs. 78% (IG) for the procedural insights and 89% (NIG) vs. 83% (IG) for the inferential mental effort. This seems to suggest that for the procedural mental effort the accuracy was better for the IG, the reverse was true for the interferential mental effort.

4.3 VTI Experiment Data Preparations

4.3.1 Post-Experiment Questionnaire

In the view transformation interaction experiment, using 2D and 3D Portal games, the data were gathered in a post-experiment questionnaire, and the analysis was based on the accuracy of the responses in this questionnaire; the participants were assessed on their breadth and depth of understanding respectively of from 2D and 3D environments. The aim in this experiment was to assess the level of insight participants got into the 2D and 3D representations of the system behaviour by their ability to interact or not with the view transformation. In these post-experiment questionnaires, participants had a choice of answering the question or mark that they did not know. When they answered correctly or incorrectly the insight was marked accurate or inaccurate respectively, and when the participants marked the question as unknown, the insight associated was coded as unidentified.

This analysis aims at looking at the significant effects in the type of representation (2D or 3D) and accuracy of the insights into the data-set (accurate, inaccurate or unidentified). Across the interaction groups and individual differences.

In order to test the validity and reliability of the 2D and 3D post-experiment questionnaire the answers were coded as described by Table 4.13 and the Cronbach's alpha was calculated with the resulting coding.

2D-3D Answer	Code
Accurate	1
Inaccurate	2
Unidentified	3

Table 4.13 – Post-experiment 2D-3D Answers Coding

The Cronbach's alpha for the 2D questionnaire was $\alpha = 0.864$ overall and when dividing the questionnaire by interactivity treatment, $\alpha = 0.866$ for the NIG and $\alpha = 0.777$ for the IG. For the 3D questionnaire the Cronbach's alpha was $\alpha = 0.800$ for the NIG and $\alpha = 0.885$ for the IG. All Cronbach's alpha reliability estimates were above 0.7, which constitutes a good reliability level [115].

[Intentionally left blank]

Chapter 5. Visual Mapping Interaction Experiment

This chapter describes the analysis and the findings of the Visual Mapping Interaction experiment, addressing objectives one and two of this thesis. Initially this chapter outlines the experiment statistical analysis. This is followed by the validation assumption of the experiment using the NASA-TLX workload assessment. Then the chapter covers the interactivity main effects before covering the compounding effects between the interaction treatment and individual differences.

5.1 Statistical Analysis 96

5.2 NASA-TLX Analysis.. 97

5.3 Interaction Effects Analysis 98

5.4 Individual Differences and Interaction Compounding Effects Analysis 102

5.4.1 Overall Metrics Analysis 102

5.4.1 Yield Analysis 107

5.5 Summary 116

5.1 Statistical Analysis

The objectives in the VMI experiment are to Investigate the effects of Visual Mapping Interaction and the compounding effects VMI has with the performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles in the context of performing an analytical task using information visualisations.

Initially to analyse the VMI main effects, a oneway analysis of variance (ANOVA) on all the DVs of interest was performed. Then to analyse the compounding effects of the interaction treatment with individual differences, the examination was divided into two sections, one analysing the overall metrics and the other analysing the accuracy and mental effort. In the overall section, the overall score was analysed using a univariate factorial ANOVA and the overall yields were analysed using a multivariate analysis of variance (MANOVA). In the individual yield analysis section a factorial ANOVA was used to test the accuracy yield in order to analyse DVs based on these yield effects between and within participants. The mental effort yield could not meet the parametric assumption satisfactorily thus multiple uncorrected Kruskal-Wallis test were performed on an exploratory basis [110, 111] to explore the possible effects individual differences have within the different interactivity treatments. Also, following any significant parametric test results, a follow-up non-parametric Jonckheere's test was performed to uncover any significant trends in relation to the individual differences.

All the results were reported at $p < .05$ level significance, unless otherwise stated.

5.2 NASA-TLX Analysis

Before the experiment analysis, the NASA-TLX data were analysed to check for experimental bias in order to validate the experiment (see section 3.1.2). The assumption was made that both IG and NIG groups did not have significantly different overall and individual workloads. Thus to check for experimental bias before starting the experiment, a oneway MANOVA was conducted on the NASA-TLX workload data in order to analyse the combined and the individual workloads (mental, physical, temporal, performance, effort and frustration). When checking for compliance with the parametric test assumptions, the normality assumptions were violated, as some of the individual workloads had a significant K-S test result by interactivity group. Nevertheless, levene's test was non-significant for all variables, thus validating the assumption of homogeneity of variances. When testing for homogeneity of covariance Box's M yielded a value of 28.54 and $p = .298$, which was not significant and warranted the acceptance of the homogeneity of covariance assumption. As the MANOVA is robust to the violation of normality, the overall results of the MANOVA can be accepted.

The results showed no significant effects from the interaction treatment on the individual workload metrics, $F(6,34) = .60, p < .05$. For the overall NASA-TLX workload, a oneway ANOVA was performed on the overall workload as the DV. The results again show no significant effects of interaction on the overall workload, $F(1,40) = .13, p < .05$. These findings validate the assumption that both experimental settings were equally demanding in terms of both overall and on an individual workload basis, as measured by the NASA-TLX.

5.3 Interaction Effects Analysis

This section analyses the effects of VMI in the context of performing an analytical task using information visualisation by investigating the following research questions:

Research Question 1: Does VMI affect the number of insights generated and their accuracy, when compared to an equivalent non-interactive task?

Research Question 2: When insights are categorised based on mental effort (inferential for high and procedural for low mental effort), does VMI have an effect on the number and accuracy of insights generated in each mental effort category?

In this context the key DVs of interest are:

- Total insights
- Total accurate insights
- Total inaccurate insights
- Total procedural insights
- Total inferential insights
- Total accurate procedural insights
- Total accurate inferential insights
- Total inaccurate procedural insights
- Total inaccurate inferential insights.

The normality assumptions were violated, having a significant K-S Test, for the following DVs:

- total insights
- total inaccurate insights
- total inaccurate procedural
- total inaccurate inferential insights.

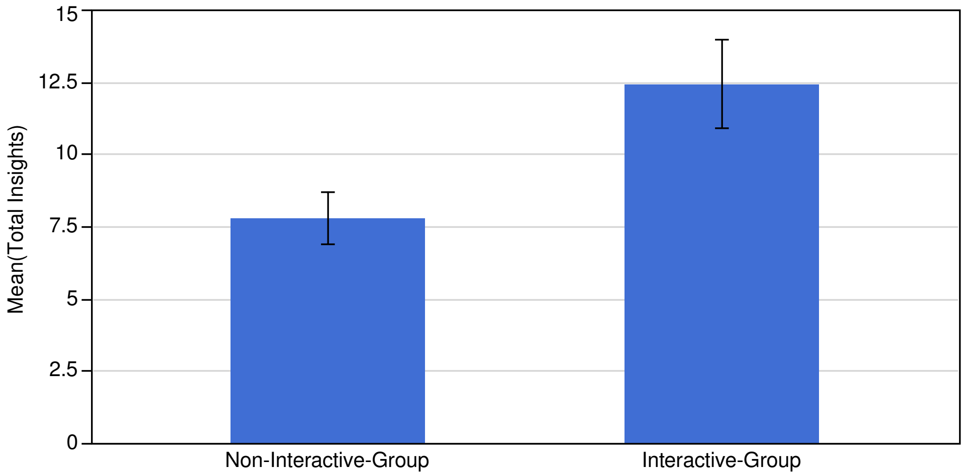
All other DVs fulfilled the normality assumption and all but the total insights had non-significant Levene's test, thus validating the assumption of homogeneity of variances.

In order to fulfil the parametric assumption for the total insight variable that violated both the normality and heteroscedastic assumptions, the variable was transformed using a square root transform described by equation (5.1), where T_i represents the transformed values and V_i the initial values.

$$T_i = \sqrt{V_i}, \forall\, i = 1 \dots N \tag{5.1}$$

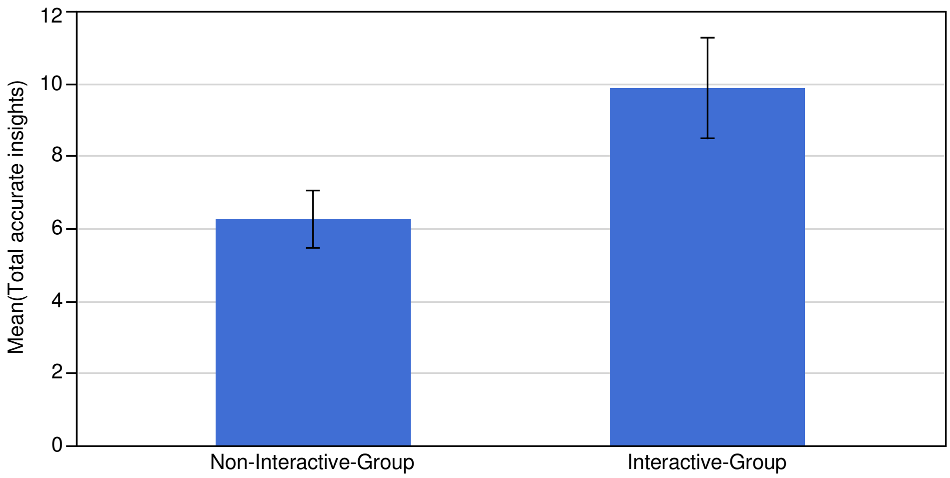
With this transformation the total insight variable met the parametric assumptions. A oneway ANOVA was performed with interactivity as the independent variable. Table VMI.1 in Appendix 14 provides a full statistical description of the un-transformed DVs.

Table VMI.2 in Appendix 14 shows the omnibus oneway ANOVA results. These results show a significant increase in the total number of insights due to the interactivity treatment, $F(1,40) = 6.31$, $r = .37$, (see Figure 5.1). Additionally there is a significant VMI effect increasing the total number of accurate insights, $F(1,40) = 5.02$, $r = .33$ (see Figure 5.2) and also increasing the total number of procedural insights, $F(1,39) = 4.42$, $r = .32$ (see Figure 5.3).



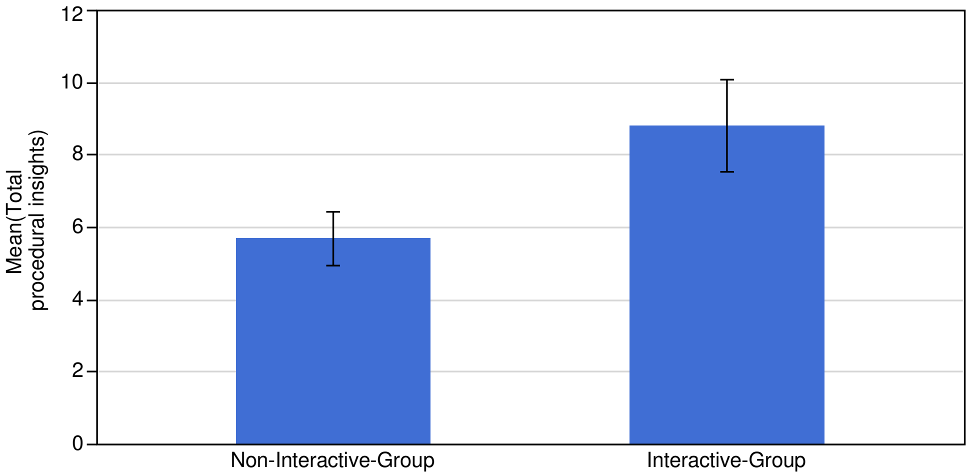
Each error bar is constructed using 1 standard error from the mean.

Figure 5.1 – Total Number of Insights by interaction grouping



Each error bar is constructed using 1 standard error from the mean.

Figure 5.2 – Total Number of Accurate Insights by interaction grouping



Each error bar is constructed using 1 standard error from the mean.

Figure 5.3 – Total Number of Accurate Procedural Insights by interaction grouping

In order to uncover possible relationships between the variables, a Pearson correlation analysis was performed comparing both the total number of accurate insights and procedural insights with the overall number of insights. As shown in Table 5.1 the total number of insights is strongly correlated to the total number of accurate insights $r = .929$, $p < .001$ and also with the total number of procedural insights, $r = .886$, $p < .001$. Additionally, the total accurate insights are also strongly correlated with the total number of procedural insights, $r = .856$, $p < .001$.

		1.	2.	3.
1.	Total insights	1		
2.	Total accurate insights	.929***	1	
3.	Total procedural insights	.886***	.856***	1

*$p < .05$, ** $p < .01$, *** $p < .001$

Table 5.1 – Total Accurate, Total Procedural and Overall Number of insights Pearson's Correlations Results

These results call for further analysis to uncover the effect of accuracy and mental effort with regards to insight generation into more details, which are analysed in the next section regarding the investigation of the compounding effects individual differences have with VMI.

5.4 Individual Differences and Interaction Compounding Effects Analysis

This section analyses the compounding effects of VMI with performance-related psychometric measures (LoC, SE, and SA) and the VARK model of learning styles in the context of performing an analytical task using information visualisations, by investigating the following research questions:

Research Question 3: Do individual differences (LoC, SE, SA, V, A, R, and K) have compound effect with VMI, whereby according to the level of the different measures, there will be a significant effect on the generation of insights and their accuracy?

Research Question 4: When categorising insights based on mental effort, do individual differences (LoC, SE, SA, V, A, R, and K) have a compounding effect with VMI with regards to the generation and accuracy of insights?

5.4.1 Overall Metrics Analysis

This section of the analysis focuses on the following overall metrics:

- Overall Score
- Overall accuracy yield
- Overall mental effort yield

These were analysed in two different groups as overall score required a transformation in order to fit the ANOVA model whereas overall accuracy and mental yield DVs fitted the MANOVA model without transformation.

Overall Score

The normality assumptions and homogeneity of variances were violated for some individual differences by interactivity grouping, having significant K-S and Levene's tests. The overall score was then transformed using a logarithmic transformation described by equation (5.2), where, T_i represents the transformed values and V_i the initial values. Once transformed, the overall score variable complied with the parametric assumptions.

$$T_i = \frac{1}{2} \cdot \ln\left(\frac{101 + V_i}{101 - V_i}\right), \forall\, i = 1 \dots N \qquad (5.2)$$

Table VMI.3 in Appendix 14 provide a full statistical description of the un-transformed DVs.

Table VMI.8 in Appendix 14 shows the factorial ANOVA results for the overall score. The results show a significant interactive effect between interactivity treatment and ATI-C (R), $F(1,32) = 5.73$, $r = .39$ (see Figure 5.4), whereby the participants with above mean read-write preference in the NIG had a significantly higher score.

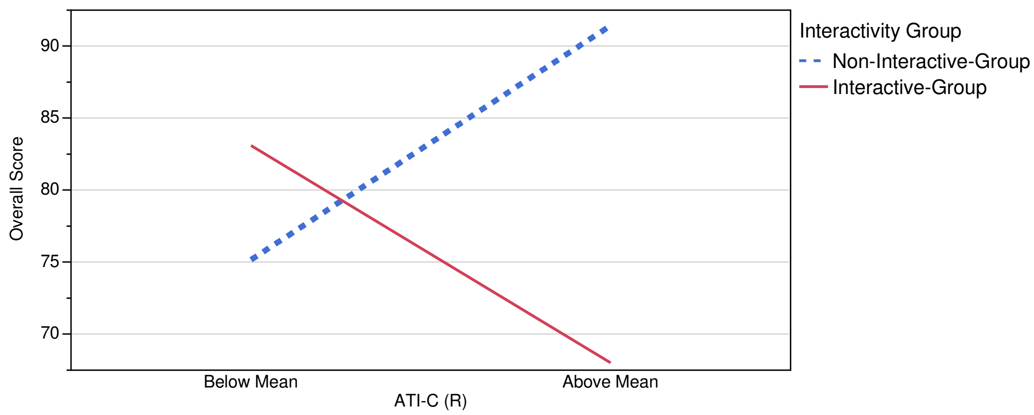


Figure 5.4 – Overall Score by interaction grouping * ATI-C (R)

Additionally a Jonckheere's test was performed to analyse possible trends, uncovering a significant trend for the overall score, in relation to NIG ATI-C (R), $J = 110.50$, $z = 2.58$, $p = .004$ (1-tailed), $r = .56$ (see Figure 5.5).

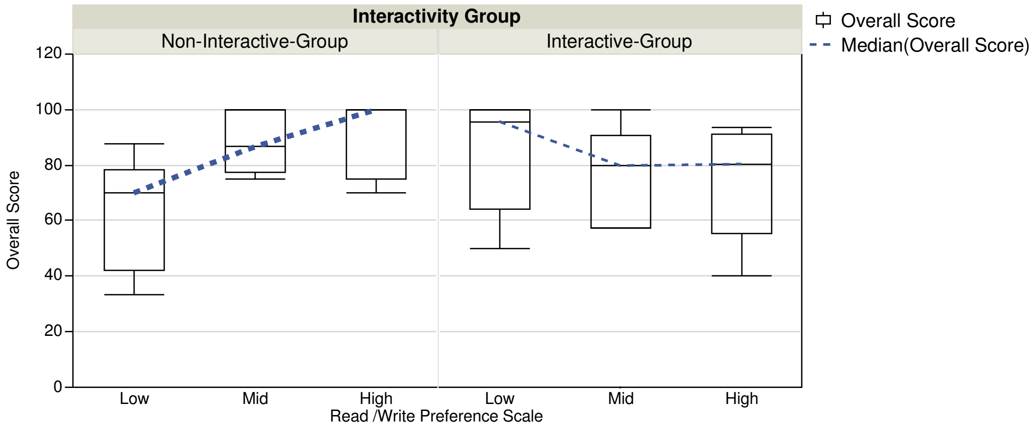


Figure 5.5 – Overall Score by interaction grouping * R Scale

Overall Accuracy Yield and Overall Mental Effort yield

The normality assumptions were violated, as some of the individual differences by interactivity grouping had a significant K-S test result, but levene's tests were not significant, thus the homogeneity of variances assumption was met. Thanks to the MANOVA robustness to normality violation, the MANOVA results could be trusted. When testing for the homogeneity of covariance, Box's M test resulted in a value of 35.57 and $p = .004$ for the psychometric analysis and a value of 23.61 and $p = .004$ for the learning preferences analysis.

Table VMI.6 and Table VMI.5 provide a full statistical description of the overall accuracy and mental effort yield variables.

The MANOVA results show an interaction effect between the interactivity treatment and the ATI-C (SE), $F(2,34) = 3.70$ and also between interactivity treatment and ATI-C (V), $F(2,32) = 4.61$.

The residuals for both variables were tested for normality, non-significant K-S test results for both variables validated these results with violations of the normality assumption.

Separate univariate ANOVA on the outcome variables was performed *posthoc*, the results are in Table VMI.7 in Appendix 14. The results show a significant overall mental effort score difference between interactivity treatment for the participants with above mean SE measures, $F(2,34) = 3.69$, $r = .31$ (see Figure 5.6). Further, Figure 5.7 suggests that the difference is due to an increase in procedural insights for participants with above mean SE measures in the IG, this is highlighted in red in the figure.

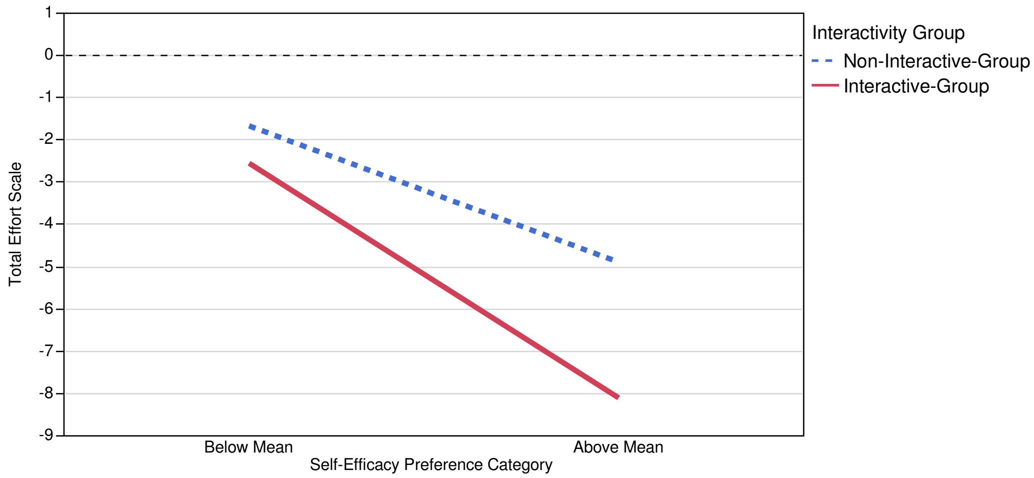


Figure 5.6 – Overall Mean Mental Effort Score by interaction grouping * ATI-C (SE)

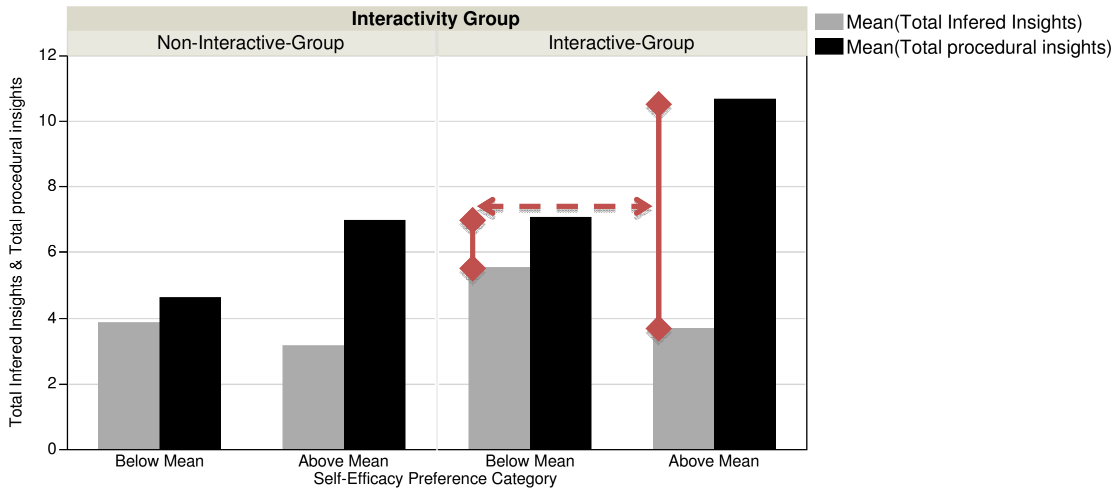


Figure 5.7 – Total Procedural and Inferential Insights

by interaction grouping * ATI-C (SE)

The results also show a significant overall accuracy yield difference between interactivity treatment for the participants with above mean visual preferences, $F(2,34) = 4.61$, $r = .35$ (see Figure 5.8). Moreover, Figure 5.9 suggests that the difference is due to an increase in accurate insights for the participants with above mean visual preferences in the IG, this is highlighted in red in the figure. Additionally, a Jonckheere's test, revealed a significant trend for the overall accuracy yield, in the interactive-group for the V-scale, $J = 103$, $z = 2.06$, $p = .02$ (1-tailed), $r = .45$ (see Figure 5.10). Otherwise, there were no more significant effects between interaction treatment and ATI-C.

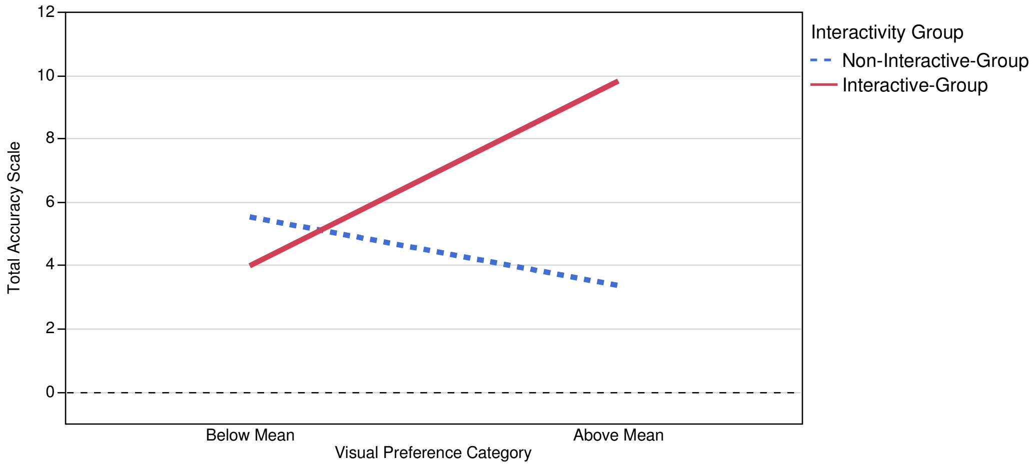


Figure 5.8 – Overall Mean Accuracy Yield by interaction grouping * ATI-C (V)

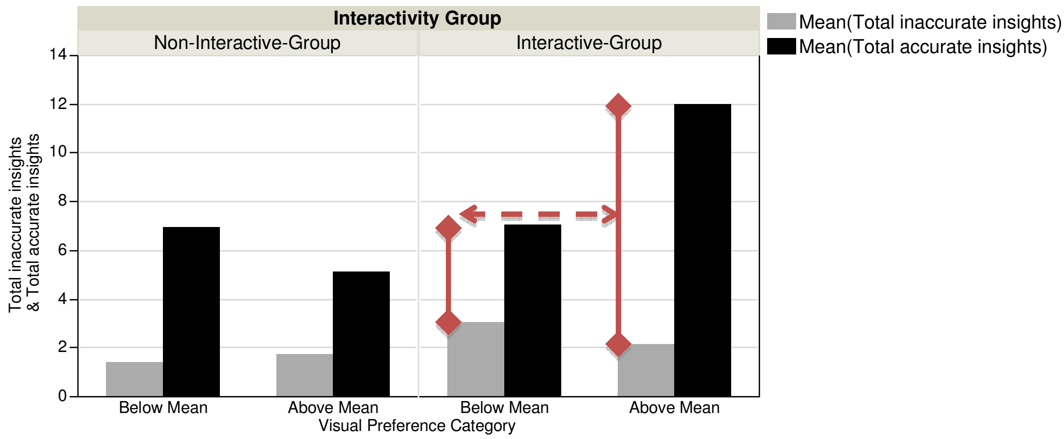


Figure 5.9 – Total Accurate and Inaccurate Insights

by interaction grouping * ATI-C (V)

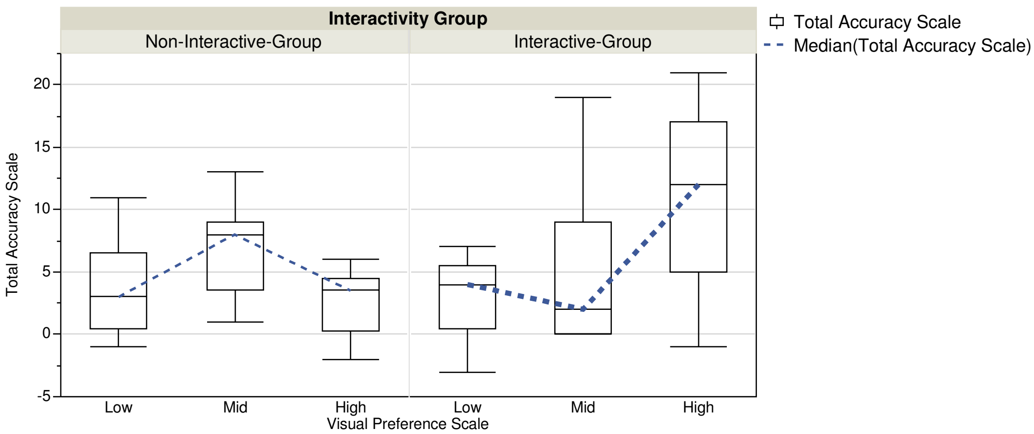


Figure 5.10 – Overall Accuracy Yield by interaction grouping * V Scale

5.4.1 Yield Analysis

This section of the analysis investigates the yields based on the following scales:

- Accuracy yield
- Mental effort yield

In order to analyse the accuracy and mental effort yields, factorial MANOVA assumptions were tested, with interactivity treatment and the ATI-C grouping (psychometric measures and VARK learning preferences alternatively) defined as the between participants factors and with mental effort as the factor with two levels (inferential and procedural) on the accuracy yield and with accuracy as the factor two levels (accurate and inaccurate) on the accuracy yield. The construct of these scales as described in 3.3.1 and as such, for the accuracy, an increased yield signifies an increase in accuracy, and for the mental effort yield an increase in yield is interpreted as an increase in mental effort.

The parametric assumptions checks showed that, both the K-S and Levene's tests were significant, thus the normality and homogeneity of variances assumptions were violated for both yields. Therefore the accuracy yield was transformed using a square root transformation described by equation (5.3), where V_i were the initial values, T_i the transformed values, and the value (-3) was added to enable to re-centre the scale onto zero to facilitate the interpretation.

$$T_i = \sqrt{V_i + 9} - 3, \forall\, i = 1 \ldots N \tag{5.3}$$

The resulting transform still had a significant K-S test result for a few cells, thus violating the normality assumption, but Levene's test was not significant for all groupings, so the equality of variance assumption was maintained for the purposes of the analysis. Hence, although the normality assumption was violated, the validity of the homogeneity of variances assumption and robustness of the ANOVA to violations of normality assumptions warrants the trustfulness of the results of this analysis. As a result the accuracy yield was analysed using a factorial ANOVA.

On the other hand for the mental effort yield, no transformation would allow the validation of the parametric assumptions; therefore multiple Kruskal-Wallis tests were performed for the analysis of the mental effort yield.

Accuracy Yield

The results Table VMI.10 in Appendix 14 show an interaction effect between the interactivity treatment and ATI-C (V), $F(2,32) = 4.61$, $r = .32$ (see Figure 5.11). A separate independent t-test revealed a significant difference for the participants with above mean visual preferences, whereby on average the transformed accuracy yield was higher for the IG ($M = .76$, *Std. E* = .15), than for the NIG ($M = .33$, *Std. E* = .13), $t(30.78) = 2.18$, $r = .36$ this result does not assume equal variances and was adjusted accordingly. This confirms the results illustrated by Figure 5.8 for the overall accuracy yield.

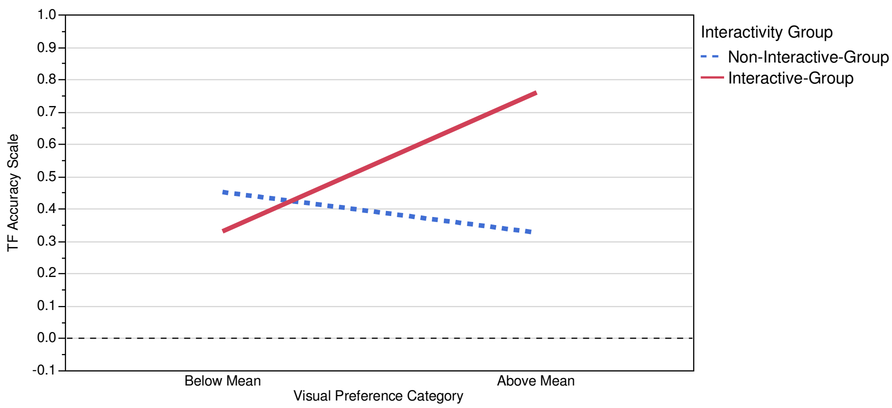


Figure 5.11 – Mean Transformed Accuracy yield by interaction grouping * ATI-C (V)

Table VMI.8 and Table VMI.9 in Appendix 14 provide a full statistical description of the accuracy variable by the different factors (effort, IG and ATI-C)

The results also showed an interaction effect for the accuracy yield, between the type of mental effort and the ATI-C (SE), $F(1,56) = 5.23$, $r = .29$. As separate independent t-test showed that there was a significant difference for participants with above mean SE measures, whereby on average the transformed accuracy yield for the procedural mental effort ($M = .82$, *Std. E* = .18) was greater than the inferential mental effort ($M = .25$, *Std. E* = .07), $t(22.90) = 2.45$, $r = .46$. These results do not assume equal variances and were adjusted accordingly.

Mental Effort Yield

As the mental yield did not meet the parametric assumptions, multiple Kruskal-Wallis tests were performed for the analysis .These tests were performed on the untransformed data on an exploratory analysis basis without correcting significance p values [110, 111] as outlined earlier in section 5.1.

The results Table VMI.11 in Appendix 14, showed no significant differences in the mental effort yield with regards to the interactivity treatment, $H(1) = 0.53$, nor was there a significant difference in the mental effort yield, when splitting by accuracy due to the interactivity treatment, $H(1) = 0.26$ and $H(1) = 0.85$ for the accurate and inaccurate insights respectively.

Individual differences two-way interactions

The results Table VMI.16 and Table VMI.17 in Appendix 14, showed no significant interaction between different interactivity treatment and the individual measures with regards to the mental effort yield.

Psychometric Measures Three-way Interactions

Table VMI.12 provide the statistical description for the psychometric measures. Also note that the key result features are highlighted in red in the figure and numbered accordingly.

The results Table VMI.18 in Appendix 14, shows that for participants with high SE measures, there is a significant difference in the mental effort yield, between the interactivity treatments for the inaccurate insights, $H(1) = 5.48$ (see Figure 5.12 ①).A Jonckheere's test showed a significant trend for the mental effort yield, for the accurate insights, with regards to the SE-scale in the NIG, $J = 35$, $z = -2.48$, $p = .008$ (1-tailed), $r = -.54$ (see Figure 5.12 ②). Additionally, results Table VMI.19 showed that for participants with high SE measures, there is a significant difference in the mental effort yield, between accurate and inaccurate insights, for the NIG, $H(1) = 6.71$, $p = .010$ (see Figure 5.12 ③).

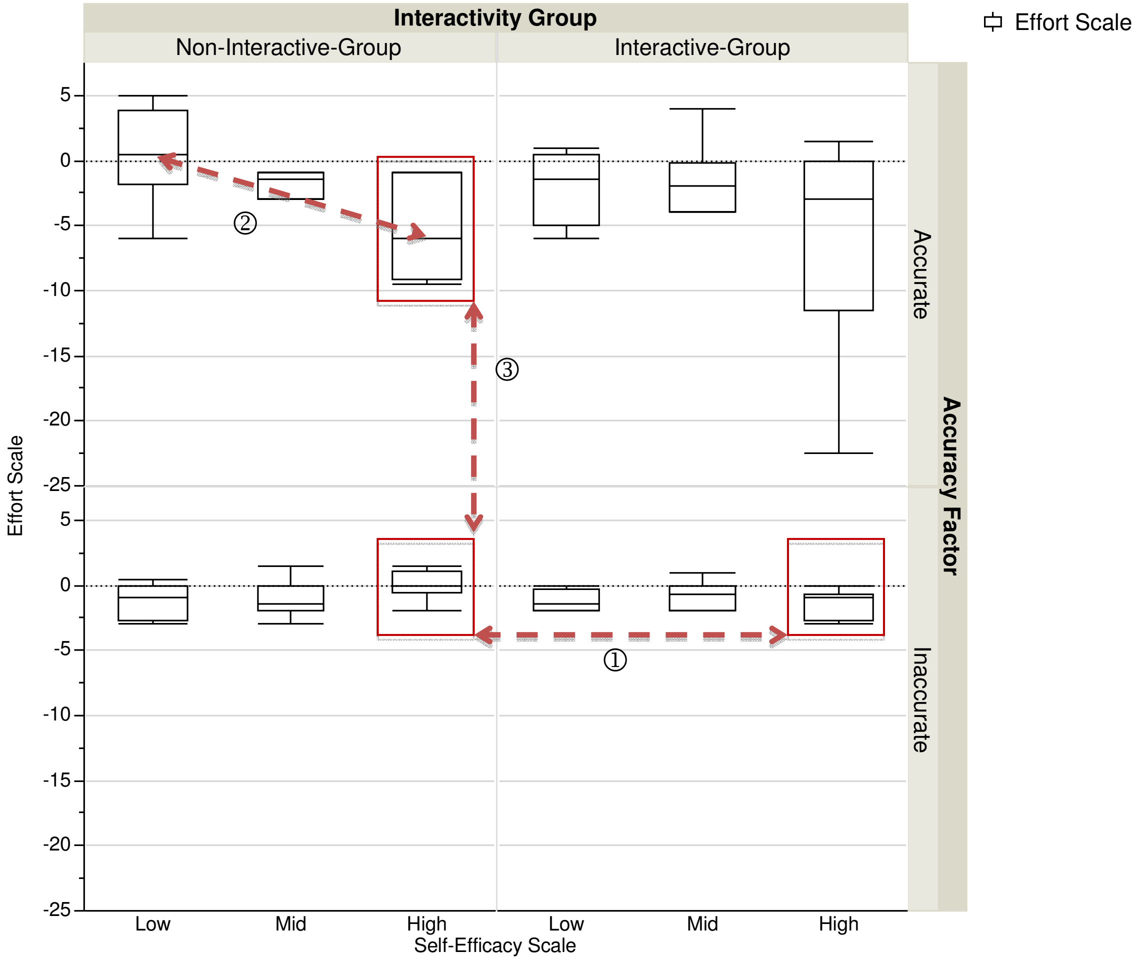


Figure 5.12 – Mental Effort Yield

by Accuracy Factor, Self-Efficacy and Interactivity Treatment

Figure 5.13 ①, suggests that the difference in between interactivity treatments inaccurate insight mental effort yield as illustrated by Figure 5.12 red highlight ①, is due to a larger increase in the inaccurate procedural insights in relation to the inferential insights for the interaction treatment. A Mann-Whitney test reveal that there is a significant difference in the inaccurate procedural insights, U = 8.50, z = –2.23, p = .026, r = .58, (Mdn = 0.00) and (Mdn = 1.50) for non-interactive and interactive treatments respectively. Further Figure 5.13 ②, suggests that the mental effort yield trend in the NIG is due primarily to an increase in number of accurate procedural insights, despite a decrease in number of accurate inferential insights. A single tailed Jonckheere's test on both the inferential and procedural accurate insights for the non-interactive treatment revealed that both trends were significant, J =32.50, z = –1.97, p = .024, r = .53 for the inferential trend and J =66.50, z = 2.59, p = .005, r = .58 for the procedural trend. Additionally, Figure 5.13 ③ also suggests that the difference in mental effort yield in the

NIG for participants with a high level of SE as illustrated in Figure 5.12 ③, is due to a significant increase in the accurate procedural insight generated by these participants.

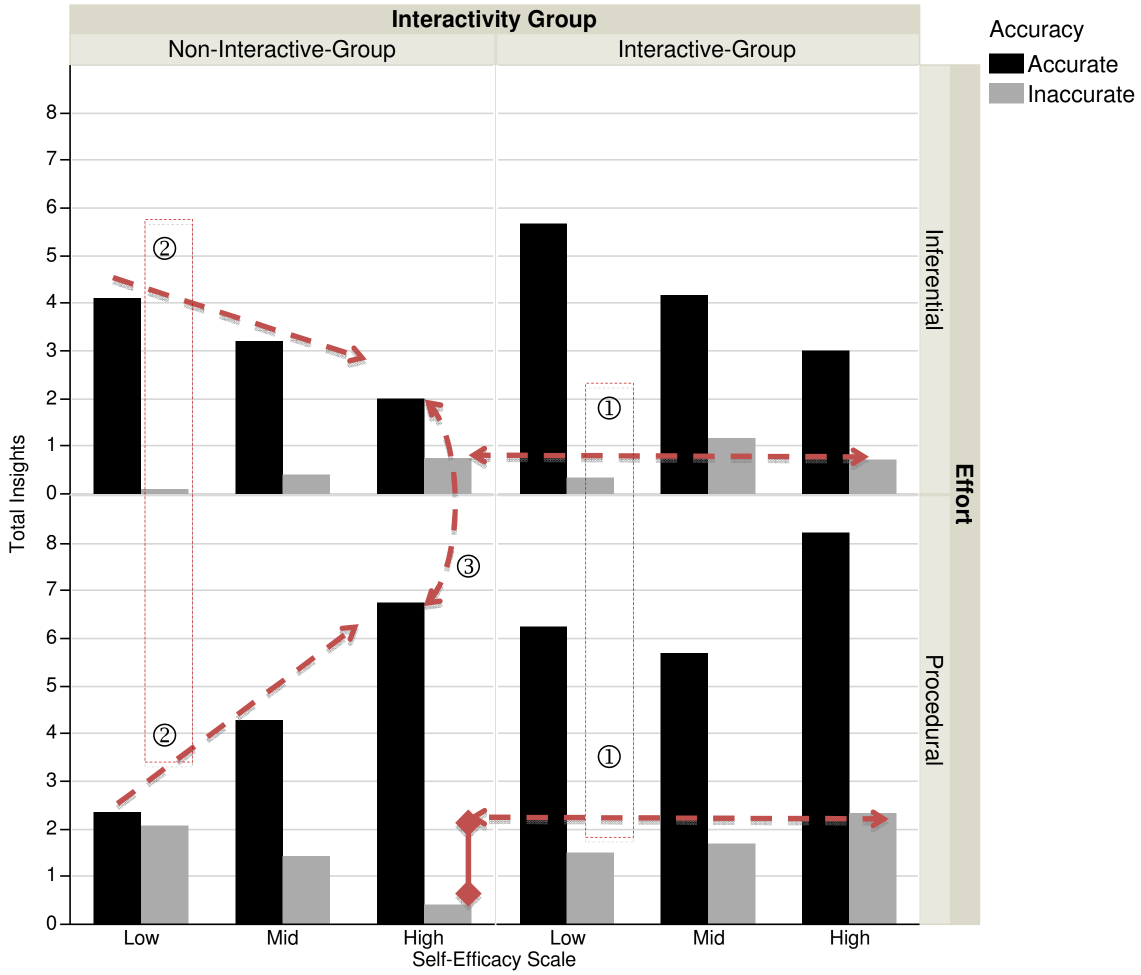


Figure 5.13 – Accurate and Inaccurate Insights by Mental Effort, Self-Efficacy and Interactivity Treatment

Otherwise as presented in Table VMI.18 and Table VMI.19 in Appendix 14 there were no further significant differences in the mental effort yield between interactivity treatments or accuracy factor with regards to individual differences.

<u>Learning preference Three-way Interactions</u>

Table VMI.13 in Appendix 14 provide the statistical description for learning preferences. As previously, note that the key results features are highlighted in red in the figure and numbered accordingly.

Results Table VMI.14 (Part 1) and Table VMI.15 (Part 1) in Appendix 14, showed no significant three-way interaction between interactivity treatments, accuracy factor and the visual or aural learning profiles in the mental effort yield.

Results Table VMI.14 (Part 2) in Appendix 14 shows that for participants with mid-level R-profiles, there was a significant difference in the mental effort yield between accurate and inaccurate insights for the NIG, $H(1) = 4.73$ (see Figure 5.14 ①). Figure 5.15 ①, indicates that the mental effort yield difference in the NIG for the mid-level R participants is mainly due to a larger proportion of accurate procedural insights.

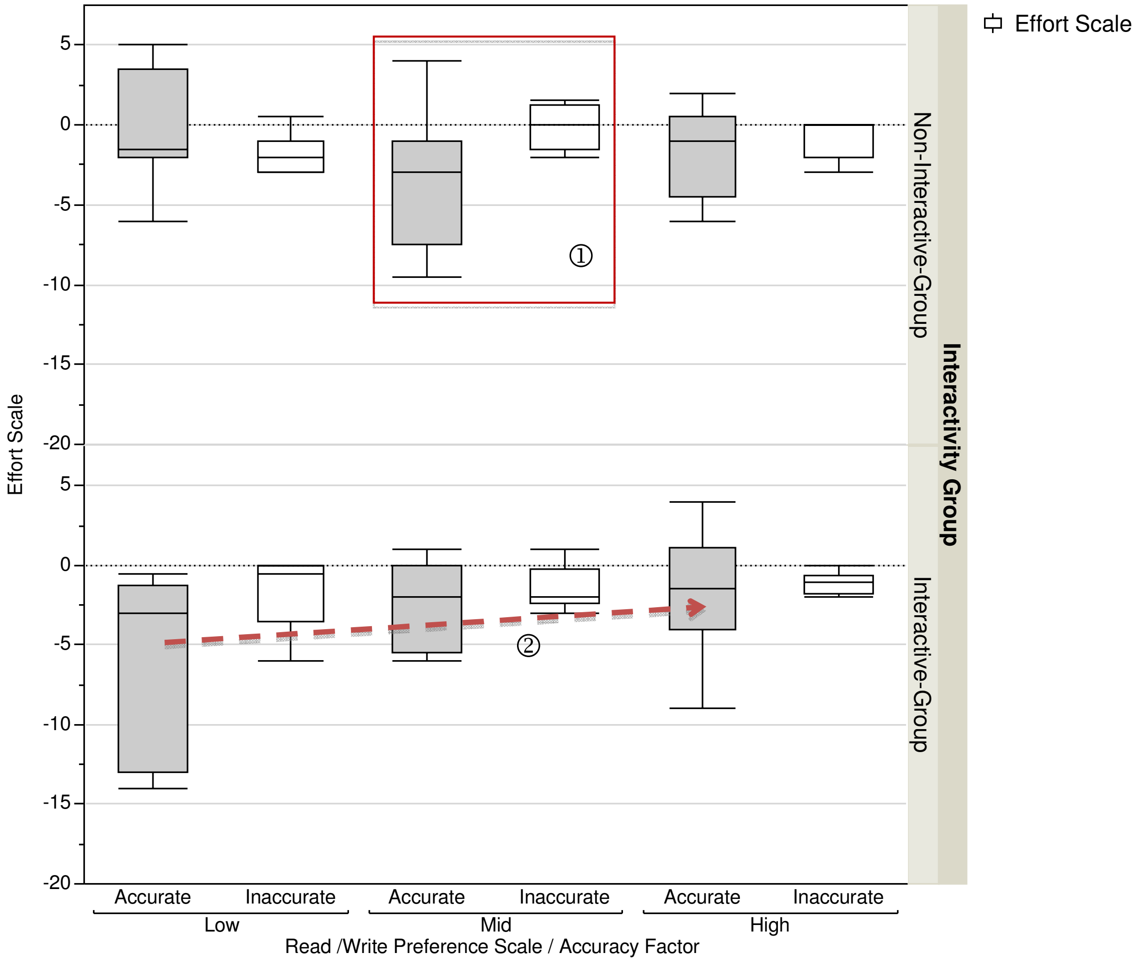


Figure 5.14 – Mental Effort Yield

by, Interactivity Treatment, Read-write Profile and Accuracy Factor

Additionally, a Jonckheere's test revealed a significant decreasing trend for the accurate procedural insights associated to the R-scale for the IG, $J = 42.5$, $z = -1.925$, $p = .027$ (1-tailed), $r = -.42$ (see Figure 5.14 ②). This can be explained by the decreasing trend in accurate procedural insights, as shown in Figure 5.15

②.

A Jonckheere's test, confirmed the significance of this trend, J = 72.00, z = –1.925, p = .027 (1-tailed), r = – .42.

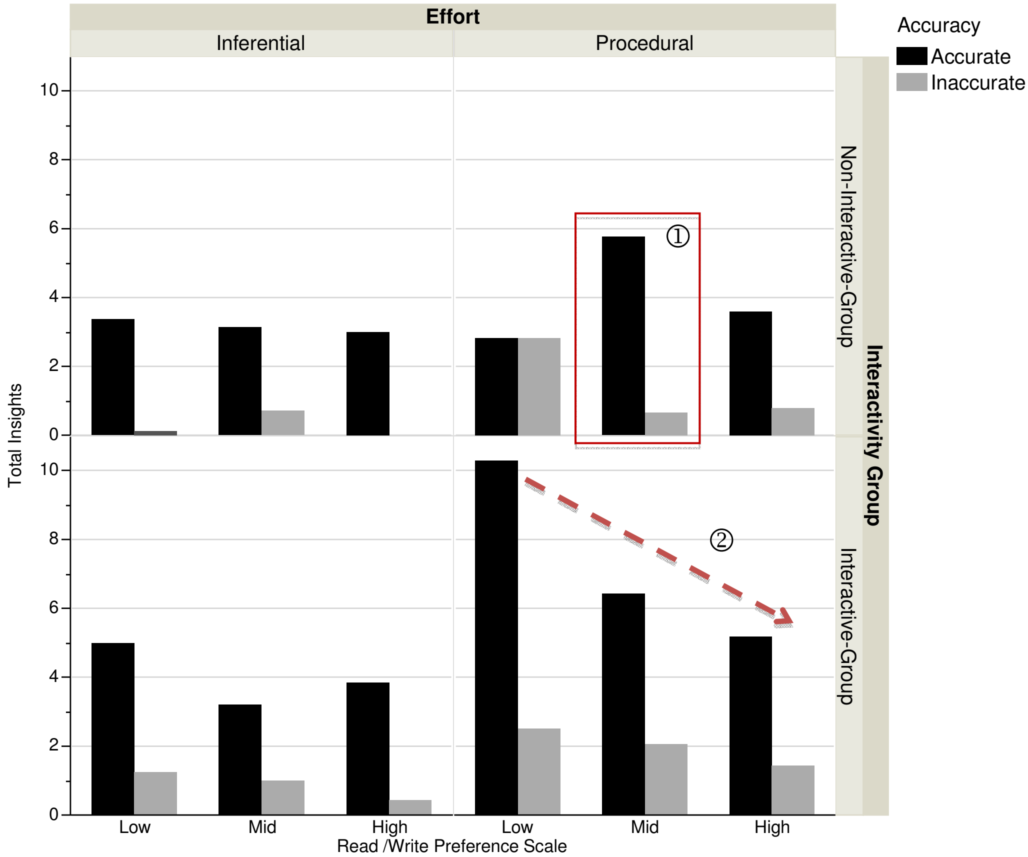


Figure 5.15 – Accurate and Inaccurate

by Mental Effort, Read-write-Scale and Interaction Treatment

Table VMI.15 (Part 2) indicates that for participants with mid-level of K-profiles, there is a significant difference in the mental effort yield between interactivity treatment for inaccurate insights, $H(1)$ = 6.77, p = .009 (see Figure 5.16 ①). Moreover Table VMI.14 Part 2 (Appendix 14) shows that for participants with low K-profiles, there was a significant difference in the mental effort yield, between accurate and inaccurate insights, for the IG, $H(1)$ = 4.29 (see Figure 5.16 ②).

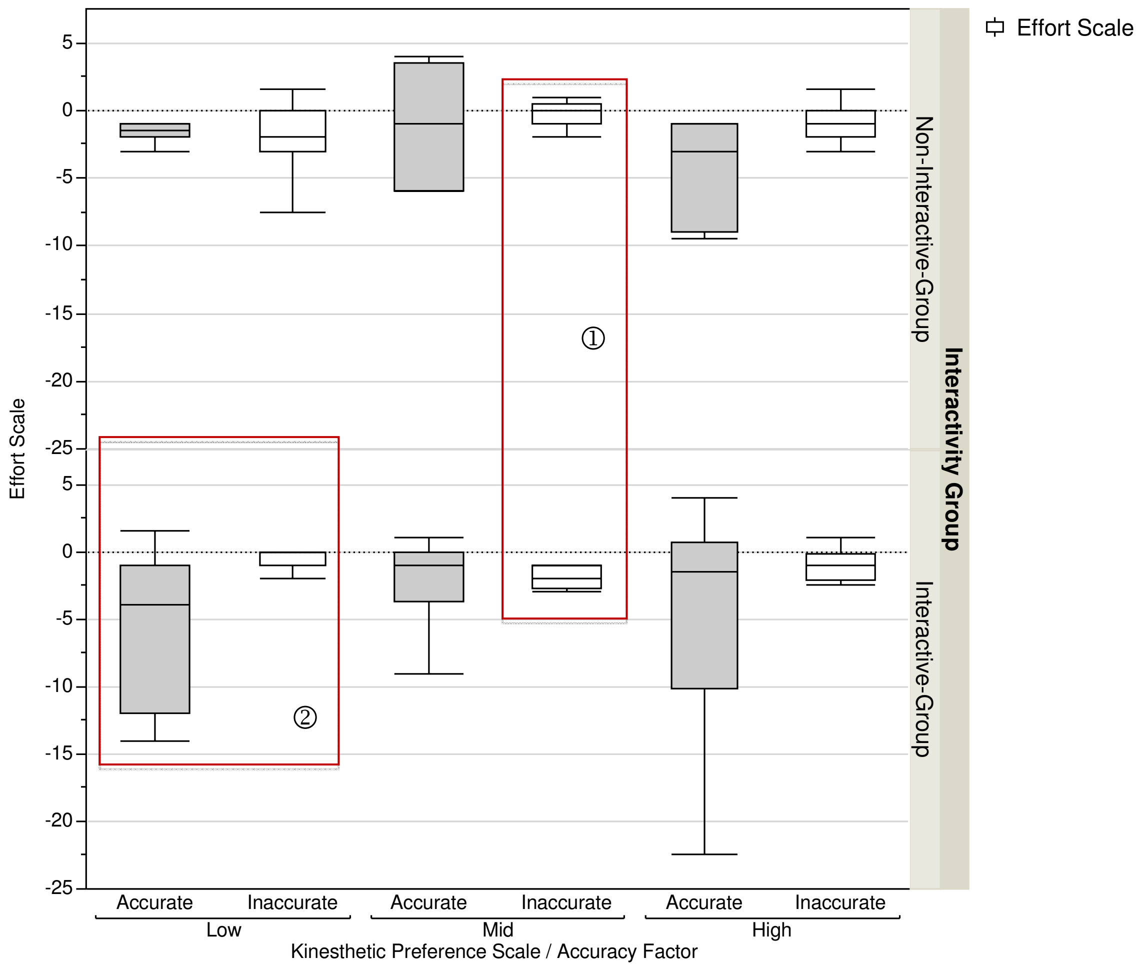


Figure 5.16 – Mental Effort Yield

by, interactivity Treatment, Kinaesthetic Profile and Accuracy Factor

Figure 5.17 ②, indicates that the mental effort yield difference in the IG for the low Kinaesthetic profile participants is mainly due to a larger proportion of accurate procedural insights. Further, Figure 5.17 ① show that the mid-Kinaesthetic profile mental effort yield difference between interaction treatments for the inaccurate insights is mainly due to a larger number of inaccurate procedural insights generated by these participants.

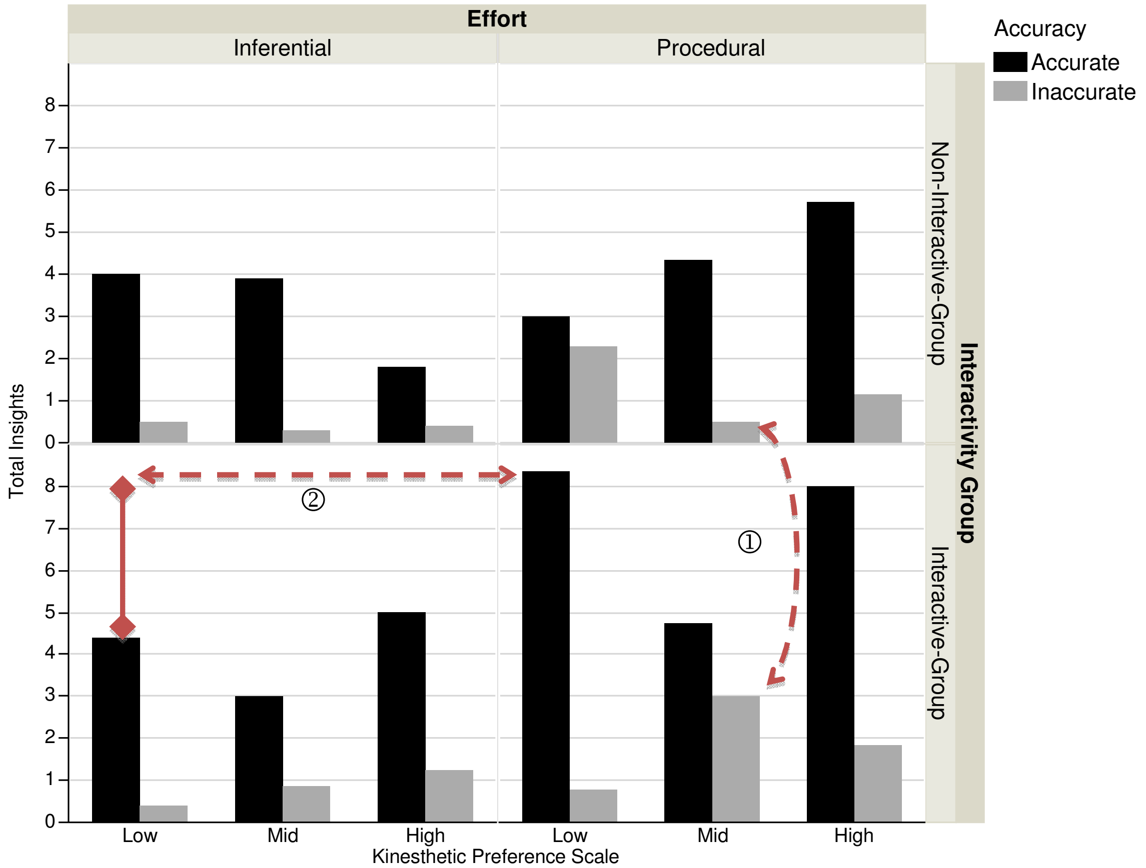


Figure 5.17 – Accurate and Inaccurate

by Mental Effort, Kinaesthetic-Scale and Interaction Treatment

5.5 Summary

VMI, in the context of performing an analytical task using information visualisations, had a significant effect in generating more insights (NIG: *M* = 7.81, *SD* = 4.18 – IG: *M* = 12.43, *SD* = 7.00). When looking at the accuracy of insights, overall VMI generated more accurate insights in comparison to the non-interactive treatment (NIG: *M* = 6.26, *SD* = 3.66 – IG: *M* = 9.88, *SD* = 6.44). With regards to mental effort required to generate insights, the only significant result showed that VMI increased significantly the procedural insights, suggesting that VMI facilitated the generation of low mental effort insights (NIG: *M* = 5.70, *SD* = 3.36 – IG: *M* = 8.81, *SD* = 5.74).

In terms of compound effects between psychometric measures and VMI, locus of control and self-acceptance measures had no significant conjoint effects. Only self-efficacy had a significant effect in relation to the interactivity treatment. For participants with above mean measures of SE, the overall mental effort yield significantly decreased with interaction treatment (NIG: *M* = 73.20, *SD* = 19.84 – IG: *M* = 77.43, *SD* = 16.19), due to an increase in the number of procedural insights (NIG: *M* = 7.09, *SD* = 2.74 – IG: *M* = 10.70, *SD* = 7.57) in relation to the decrease in inferential insights (NIG: *M* = 5.56, *SD* = 3.50 – IG: *M* = 3.71, *SD* = 2.29). Moreover, the mental effort yield for the inaccurate insights was significantly different for participants with measures of SE above half a SD above the overall group mean (NIG: *M* = 0.08, *SD* = 1.20 – IG: *M* = -1.78, *SD* = 1.84), this difference can be explained by a significant difference in the number of inaccurate procedural insights between interaction treatments (NIG: *Mdn* = 0.00 – IG: *Mdn* = 1.50), qualified as a large effect (*r* = .58), whilst the difference in the number of inaccurate inferential insights between interaction treatments is not significant (NIG: *Mdn* = 0.50 – IG: *Mdn* = 0.00). Additionally the results also show that for participants with above mean measures of SE, the accuracy yield for procedural insights were significantly higher than inferential ones (inferential: *M* = 2.23, *SD* = 1.64 – procedural: *M* = 6.16, *SD* = 6.17). The findings show as well as that there is a decreasing trend in the mental effort yield for the accurate insights in relation to the SE measure independently from interaction (Low-SE: *M* = 0.50, *SD* = 3.75 – Mid-SE: *M* = -2.00, *SD* = 0.96 – High-SE: *M* = -5.42, *SD* = 3.72).

The VARK learning preferences had more significant compounding effects with VMI. The accuracy yield follows an increasing trend related to the compounding effects of interaction with an increasing visual measure, whereby the accuracy improves in line with an increase in the visual score (Low-V: M = 3.20, SD = 3.70 – Mid-V: M = 5.86, SD = 6.96 – High-V: M = 10.78, SD = 7.17). On the other hand, the read-write preference had a significant conjoint effect with interaction treatment for the mental effort yield of accurate insights, whereby an increasing mental effort yield trend can be observed as the read-write measures increase (Low-R: M = -6.30, SD = 6.22 – Mid-R: M = -4.44, SD = 7.67 – High-R: M = -1.81, SD = 3.96). This is due to a decrease in the number of procedural insights in relation to the inferential ones whilst the number of inferential insights remained constant. Also the results show an increasing trend in the performance score related to the increase in the read-write measure, for the control group (Low-R: M = 64.77, SD = 19.73 – Mid-R: M = 87.65, SD = 10.45 – High-R: M = 90.00, SD = 14.14). Whereas there is no effect for the IG (Low-R: M = 84.84, SD = 21.39 – Mid-R: M = 69.50, SD = 31.93 – High-R: M = 74.86, SD = 20.15). For the participants with above mean read-write measures, the overall score is higher for the non-interactive participants than the IG (NIG: M = 91.43, SD = 12.15 –IG: M = 68.01, SD = 29.09). Lastly, the results shows a significant difference in the mental effort yield for the participants with kinaesthetic scores below half a SD below overall group’s mean in the interactivity group, whereby the yield for accurate mental effort was lower than that for inaccurate insights (accurate: Mdn = 4.00 – inaccurate: Mdn = 0.00), due to a proportionally higher number of accurate procedural insights in comparison to the inferential insights (inferential: M = 4.40, SD = 3.36 – procedural: M = 8.36, SD = 5.53). Additionally, the participants with mid-level kinaesthetic scores, defined, as within ± half a SD from the group’s mean, had significantly lower mental effort yield for the inaccurate insights in the IG, this can be explained by a significant increase in the number of inaccurate procedural insights due to the interaction treatment (NIG: M = 0.50, SD = 0.84 – IG: M = 3.00, SD = 2.62). Otherwise, there were no compounding effects of the in relation to the aural preferences with interaction.

[Intentionally left blank]

Chapter 6. View Transformation Interaction Experiment

This chapter describes the analysis of View Transformation Interaction in the context of a problem-solving task using 2D and a 3D visual representation of the game of portal as described in chapter 3. Initially this chapter outlines the experiment, statistical analysis and the validation assumption of the experiment using the NASA-TLX workload assessment. Then the chapter covers the interactivity effects and compounding effects between the interaction treatment and individual differences. It concludes with the findings of the interaction treatment using different visual representations – 2D and 3D.

6.1 Statistical Analysis 120

6.2 NASA-TLX Analysis.. 121

6.3 Interaction and Individual Differences Compounding Effects Analysis 124

6.3.1 Overall Percentage Analysis 124

6.3.2 Overall Score.. 126

6.4 Representation Analysis.. 127

6.4.1 Normalised Percentage Analysis 127

6.4.1 Score Analysis 131

6.5 Summary 138

6.1 Statistical Analysis

The objectives for the View Transformation Interaction (VTI) experiment are to Investigate its effects and its compounding effects with the performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles in the context of a problem-solving task. Here VTI is used to explore the problem data-set using a game-based simulation using a 2D and a 3D visual representation.

The analysis is sectioned into two main parts, the first part combines both visual representations (overall) into a combined factor, and the second part analyses specifically the representation aspects comparing the 2D and 3D representations. In the combined section of the analysis, the overall percentages of accurate, inaccurate and undefined insights are analysed using a MANOVA, then the overall score is analysed using a univariate factorial ANOVA. The score and percentages are analysed separately as they are different in nature. In the representation section of the analysis the scores and percentages are also separated, the normalised percentage are analysed using a repeated measures factorial ANOVA where the representation is the within subject factor and the individual differences and interaction treatment are the between subject factor. For the individual representation score section of the analysis, the parametric assumptions could not be met; hence a non-parametric approach was taken for the analysis. Multiple Friedman ANOVA were performed *in lieu* of the repeated ANOVA testing the within subject factor. For the between subject factor a Kruskal-Wallis test was used excluding the individual differences and a Mann-Whitney test was performed with regards to the individual differences splitting by interaction treatment. Note that the Mann-Whitney test and Kruskal-Wallis test are equivalent when there are less than two factors, the benefit in using the Mann-Whitney is that SPSS provides a z-value, thus enabling the calculation of the effect size.

All the results were reported at $p < .05$ level significance, unless otherwise stated.

6.2 NASA-TLX Analysis

Before the analysis the NASA-TLX data were analysed to check for experimental bias. The expectations were that workload was significantly different for both groups (interactive and non-interactive). But, for the item of most importance in this analysis, namely the performance workload, the assumption was that the workload was not significantly different for both groups, thus validating the analysis of the post-experiment questionnaire.

In order to analyse the NASA-TLX combined and individual workloads (mental, physical, temporal, performance, effort and frustration), the parametric test assumptions were checked. The normality assumption was violated, as some of the individual workloads had a significant result when running a K-S test by interactivity group. Nevertheless, Levene's tests on all variables were non-significant suggesting the validation of the assumption of homogeneity of variances. Additionally, a test for homogeneity of covariance was required in this analysis as a MANOVA was used. For the 2D experiment, Box's M gave a value of 19.38 and $p = .334$, which was not significant and warrants the acceptance of the homogeneity of covariance assumption. For the 3D experiment, Box's M gave a value of 40.86 and $p = .002$, which could be considered as not significant at $p < .001$ [106] due to the oversensitivity of Box's M test to deviation from normality. Hence, this warrants the acceptance of the homogeneity of covariance assumption as well.

Table 6.1 show the NASA-TLX measures statistics description for the 2D and 3D representations.

	2D Experiment				3D Experiment			
	NIG		IG		NIG		IG	
	M	SD	M	SD	M	SD	M	SD
TLX Mental Demand	17.71	12.88	25.43	11.43	18.95	13.48	29.90	12.51
TLX Physical Demand	1.43	6.10	2.29	4.19	1.43	6.10	4.43	11.18
TLX Temporal Demand	6.67	7.02	22.14	12.70	4.00	3.33	16.81	10.60
TLX Performance Score	18.05	10.80	12.62	8.33	18.76	9.68	12.67	10.37
TLX Effort Score	19.76	10.55	27.81	13.38	16.52	7.98	24.95	12.04
TLX Frustration Score	10.62	11.80	22.48	15.96	11.48	11.89	18.48	15.33

N: Number of participants, **M:** Mean, **SD:** Standard deviation

Table 6.1 – TLX Measures Statistic Description

In Figure 6.1 the key result regarding the assumptions is highlighted in red i.e. for the performance score there was no significant difference between the workload means.

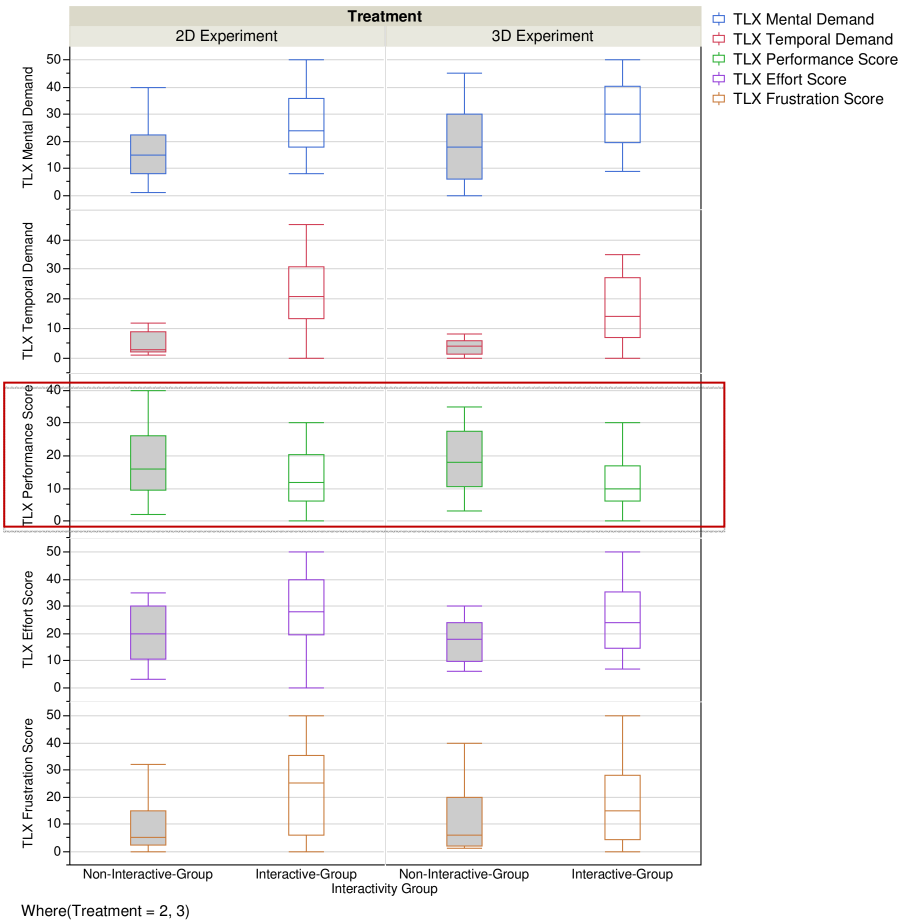


Figure 6.1 – TLX Measures: 2D / 3D Representation by interaction grouping

The interaction treatment effects on the 2D individual workload metrics was significant, $F(5,36) = 7.83$, and on the 3D individual workload metrics, $F(5,36) = 9.91$. Separate univariate ANOVAs on all the 2D TLX variables revealed that there was a significant effect on the mental workload (see Table 6.2), $F(1, 40) = 4.22$, $p < .05$, $r = .31$, and on the temporal workload, $F(1, 40) = 23.89$, $p < .001$, $r = .61$, as well as the effort score $F(1, 40) = 4.68$, $p < .05$, $r = .32$, and frustration score, $F(1, 40) = 7.49$, $p < .05$, $r = .40$. Figure 6.1 shows that interaction significantly increased all the above mentioned workloads, but the performance measure, $F(1, 40) = 3.32$, $p = .076$ was non-significant, thus not affected by interaction.

Separate univariate ANOVAs on all the 3D TLX variables revealed that there was a significant effect on the mental workload (see Table 6.2), $F(1, 40) = 4.45$, $p < .001$, $r = .40$, and on the temporal workload, $F(1, 40) = 27.91$, $p < .001$, $r = .64$, as well as the effort score $F(1, 40) = 7.15$, $p < .05$, $r = .39$. There was no significant differences for the frustration score $F(1, 40) = 2.74$, $p = .106$, nor for the performance measure $F(1, 40) = 3.88$, $p = .056$ (see Figure 6.1). Therefore, frustration and performance workloads were not affected by interaction.

		2D		3D	
TLX-Measure	N	F	p	F	p
TLX Scored Mental Demand	42	4.22	0.047*	7.45	0.009**
TLX Scored Temporal Demand	42	23.89	0.000***	27.91	0.000***
TLX Scored Performance	42	3.32	0.076	3.88	0.056
TLX Scored Effort	42	4.68	0.036*	7.16	0.011*
TLX Scored Frustration	42	7.49	0.009**	2.74	0.106

*$p < .05$, ** $p < .01$,*** $p < .001$
N: Number of participants, ***F*:** F ratio, ***p*:** p-value

Table 6.2 – TLX Measures Univariate ANOVA Results

The results above validate the assumption that interaction affects both the 2D and 3D experimental settings, but the performance workload was not significantly different for the different interaction groups in either visual representations.

6.3 Interaction and Individual Differences Compounding Effects Analysis

This section analyses the effects of VTI and the compounding effects of VTI with the performance-related psychometric measures (LoC, SE, and SA) and the VARK model of learning styles in the context of a problem-solving task. VTI in this experiment is used to explore the problem data-set using a game-based simulation within a 2D and 3D visual representation. This part of the analysis focuses on the combined 2D and the 3D representations outcome to investigate the following research questions.

Research Question 5: Does View Transformation Interaction affect the number and accuracy of insights identified in a problem data-set represented in a game-based simulation, when comparing to an equivalent non-interactive task?

Research Question 8: Independently from the representation, do Locus of Control, Self-efficacy, Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences have a compound effects with View Transformation Interaction, whereby according to the level of the different measures, there will be a significant effect to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?

6.3.1 Overall Percentage Analysis

This section of the analysis focuses on the following overall percentage metrics:

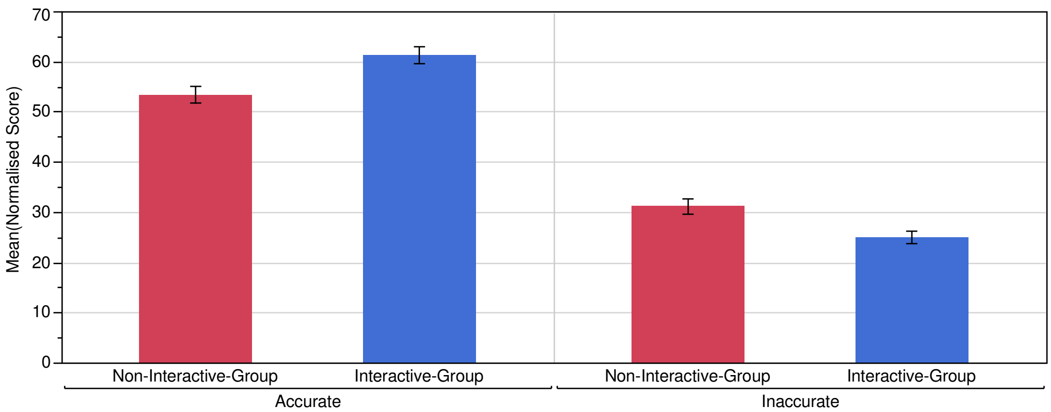
- Overall percentage of accurate insights;
- Overall percentage of inaccurate insights; and,
- Overall percentage of unidentified insights.

Table VTI.20 in Appendix 15, provides the statistical description by interaction treatment of the overall percentages of the accurate, inaccurate and unidentified insights into the data-set and Table VTI.21 and Table VTI.22 in Appendix 15 provide the full statistical description by individual difference ATI-C and IG.

The normality assumptions for the overall percentages variables were violated, as some of the individual differences by interactivity group, had a significant K-S test result, but levene's tests were not significant, thus the homogeneity of variances assumption was met. When testing for the homogeneity of covariance, Box's M test resulted in a value of 11.87 and p = .689 for the psychometric analysis and a value of 12.69 and p = .954 for the learning preferences analysis. Thanks to the MANOVA robustness to normality violation, the MANOVA results could thus be trusted.

Both MANOVAs showed that the interactivity treatment had significant effects on the overall percentages, $F(3,32) = 3.97$ for the psychometric measures and $F(3,30) = 4.76$, in the context of the learning preferences. After the MANOVA the residuals for all variables were tested for normality, the results of the K-S test of normality were non-significant for both variables, validating these results.

Separate univariate ANOVA were performed on the outcome variables, and as shown in the results Table VTI.23 in Appendix 15, interaction significantly increased the accurate insights, $F(1,32) = 10.32$, $r = .49$ and significantly decreased the inaccurate insights, $F(1,32) = 8.13$, $r = .45$ (see Figure 6.2) and had no significant effect on the overall unidentified insights percentage $F(1,32) = 0.84$, $r = .16$.



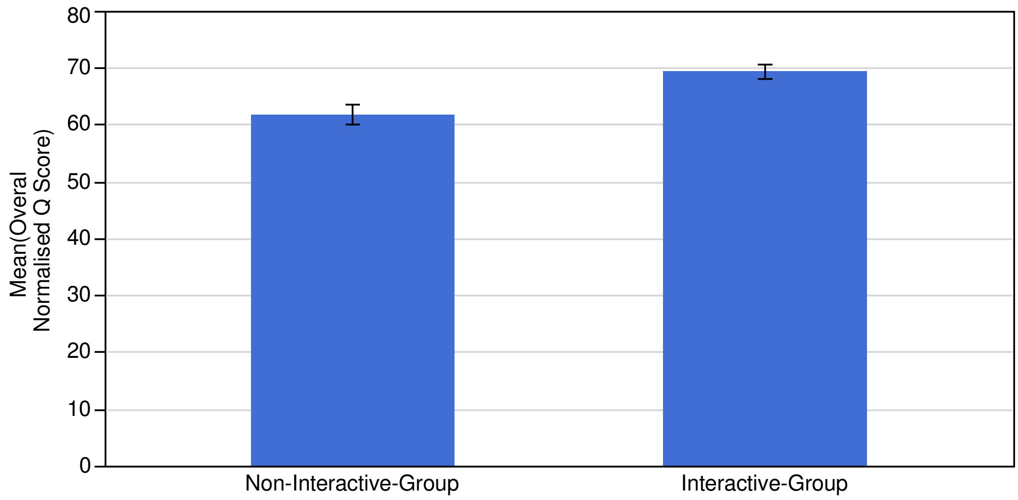
Each error bar is constructed using 1 standard error from the mean.

Figure 6.2 – Accurate and Inaccurate Overall Percentage by interaction grouping

6.3.2 Overall Score

Although the overall score did not meet the normality assumption for all interaction treatment by individual difference grouping, (significant K-S test), levene's test showed no significant outcomes for these groupings, thus fulfilling the heteroscedasticity assumptions and thus validating the use of a univariate factorial ANOVA.

As shown in results Table VTI.24 in Appendix 15, the interaction treatment had a significant effect on the overall score, $F(1,32) = 14.57$, $p = .001$, $r = .56$ (see Figure 6.3), showing that the IG had a significantly higher overall score.



Each error bar is constructed using 1 standard error from the mean.

Figure 6.3 – Overall Score by Interactivity Treatment

Table VTI.25 in Appendix 15 provide a full statistical description of the overall score by interactivity treatment and individual differences ATI-C

6.4 Representation Analysis

This section of the VTI analysis investigates the main effects of the representations – 2D and 3D. Then it covers the analysis of the two-way interaction effects of the interactivity treatment with the view representations. Finally, this section outlines the three-way interactions effects of the interactivity treatment, representation and the individual differences. Overall this analysis aims at investigating the following research questions:

Research Question 6: Does the view representation – 2D and/or 3D, have an effect on the number and accuracy of insights into a problem data-set represented in a game-based simulation?

Research Question 7: Does the representation – 2D and/or 3D and View Transformation Interaction have an interaction effect with regards to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?

Research Question 9: Does the representation – 2D and/or 3D and View Transformation Interaction have compounding effects with Locus of Control, Self-efficacy, Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences, whereby according to the level of the different measures, there will be a significant effect to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?

6.4.1 Normalised Percentage Analysis

This section investigates the following dependent variables:

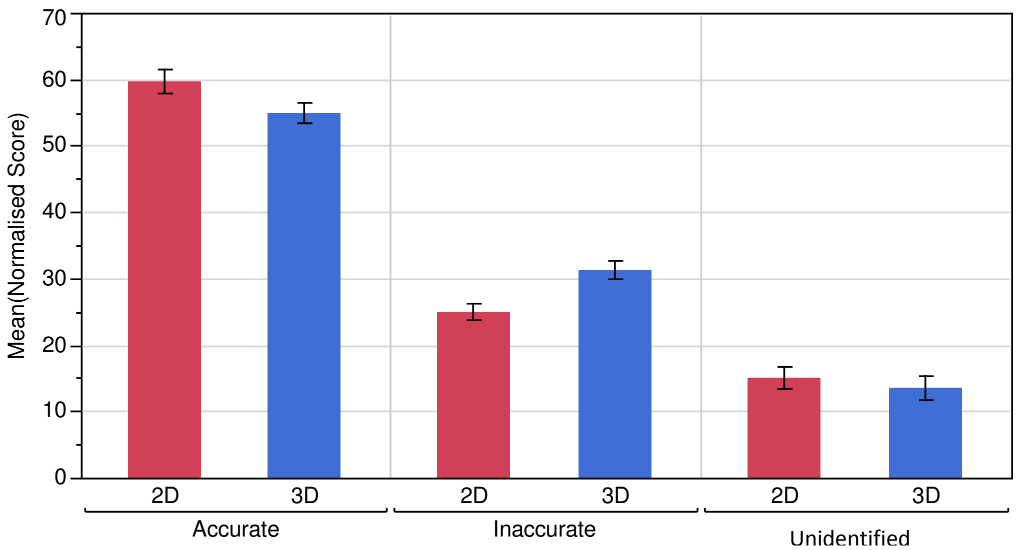
- Total percentage of accurate 2D insights
- Total percentage of inaccurate 2D insights
- Total percentage of unidentified 2D insights
- Total percentage of accurate 3D insights
- Total percentage of inaccurate 3D insights
- Total percentage of unidentified 3D insights.

In preparation of the analysis the parametric assumptions were checked for the repeated measures factorial ANOVA. The normality assumptions were violated, as some of the individual differences by interactivity group had a significant K-S test result, but levene's tests were not significant, thus the homogeneity of variances assumption was met. When testing for the homogeneity of covariance, as shown in Table 6.3 Box's M test results were not significant for both the psychometric and the learning preferences analysis. As the ANOVA is robust to normality violation, the results could be thus trusted.

	Psychometric Measures		Learning Preferences	
	Box's M	p	Box's M	p
Accurate	20.64	0.160	26.74	0.455
Inaccurate	15.06	0.428	21.16	0.690
Unidentified	15.91	0.376	30.23	0.326

Table 6.3 – Box's Test of Equality of Covariance Matrices Results for the Accurate, Inaccurate and Unidentified Percentages

As shown in the results Table VTI.27 in Appendix 15, the mean accurate insights gleaned in the 2D visual representation is significantly higher than the 3D representation, $F(1,32) = 4.81$, $r = .36$ and as was for the inaccurate insights, $F(1,32) = 10.25$, $p < .01$, r = .49 (see Figure 6.4).

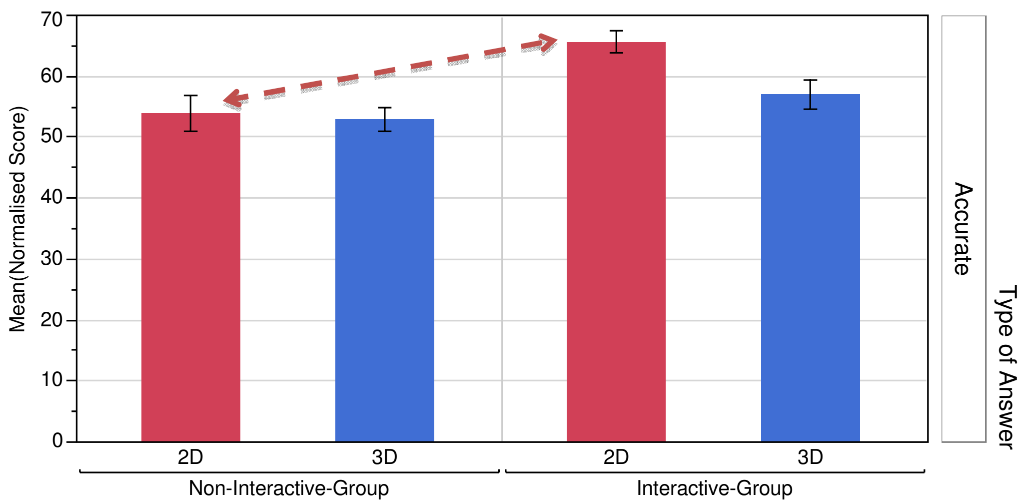


Each error bar is constructed using 1 standard error from the mean.

Figure 6.4 – Accurate and Inaccurate Percentage by Representation

Also, the results confirm the findings of the overall percentage section illustrated by Figure 6.2, with a significant main interaction treatment effect for the accurate percentage, $F(1,32) = 9.81$, $p < .01$, $r = .48$ and the inaccurate percentage, $F(1,32) = 8.14$, $p < .01$, $r = .50$.

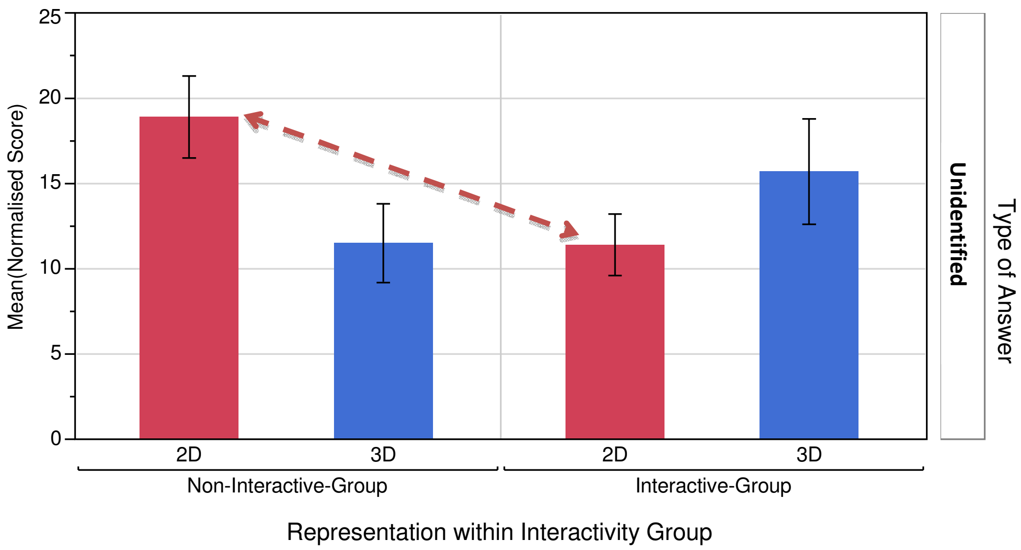
Additionally, there was an interaction effect between the interactivity treatment and visual representation for the accurate insights gathered, $F(1,32) = 10.66$, $p < .01$, $r = .50$ (see Figure 6.5). Follow-up t-tests were performed to analyse the interactivity treatments effect on the different 2D and 3D representations. A Bonferroni correction was used to protect from Type I error, so that all effects were reported at a $p < .025$ level of significance. The result showed that there was a significant difference for the 2D accurate percentage, whereby on average the interactive ($M = 65.66$, *Std. E* = 8.22) was greater than the NIG ($M = 53.92$, *Std. E* = 13.21), $t(33.46) = 3.46$, $p = 0.002$, $r = .51$ this result is highlighted in red in Figure 6.5. Levene's test for the 2D accurate percentage was significant, thus the results did not assume equality of variances and the t-test was adjusted accordingly. The 3D accurate percentage differences between interactivity treatment were not significant, $t(40) = 1.31$, $p = 0.197$, $r = .20$.



Each error bar is constructed using 1 standard error from the mean.

Figure 6.5 – Accurate Percentage by Representation and Interactivity Treatment

Moreover, Table VTI.27 in Appendix 15 shows a significant interaction effect between representation and interactivity treatment for the unidentified percentage, $F(1,32) = 14.82$, $p < .01$, $r = .55$ (see Figure 6.6). Follow-up t-tests were performed to analyse interactivity treatment effect on the 2D and 3D representations. A Bonferroni correction was used to protect from Type I error, so that all effects were reported at a $p < .025$ level of significance. The result showed that there was a significant difference for the 2D unidentified percentage of insights gathered, whereby on average the interactive (M = 11.40, $Std.\ E$ = 8.31) was significantly lower than the NIG (M = 18.91, $Std.\ E$ = 10.96), $t(40) = -2.50$, $p = 0.017$, $r = .38$. This result is highlighted in red in Figure 6.6. Levene's test for the 2D unidentified insight percentage was not significant, thus the results assumed equality of variances and the t-test. The 3D unidentified insight percentage differences between interactivity treatment was not significant, $t(40) = 1.09$, $p = 0.283$, $r = .17$.



Each error bar is constructed using 1 standard error from the mean.

Figure 6.6 – Unidentified Percentage by Representation and Interactivity Treatment

Table VTI.26 in Appendix 15 provides a statistical description of the accurate, inaccurate and unidentified percentages by representation and interactive treatment.

6.4.1 Score Analysis

This section of the analysis focuses on the score metrics as defined in equation (3.2) in section 3.4.1, with following dependent variables as a result:

- 2D Score
- 3D Score.

The normality assumptions were violated for both DVs, as some of the individual differences by interactivity group had a significant K-S test result. For the 3D scores levene's test was not significant, but for the 2D score, levene's test was significant. Several data transforms were attempted on the 2D score to meet the parametric assumptions, in order to permit a repeated measure ANOVA to analyse the effect of representation. Unfortunately, the 2D score could not meet satisfactorily the parametric assumptions. Although, the 3D score met the parametric assumptions, in order to assess the representation effects, a non-parametric approach was chosen in substitution to the repeated measure ANOVA. A Kruskal-Wallis non-parametric test was used with interactivity treatment as the between group factor on the 2D and 3D scores, and Friedman's ANOVAs with the 2D and 3D representation as the within group factor for the different interaction treatments.

In order to analyse the individual differences compound effects with interaction treatment, a Mann-Whitney test was used on the 2D and 3D scores, using the interactivity treatment as the between group factor for the ATI-C individual differences. Note that the Mann-Whitney and Kruskal-Wallis tests are equivalent when the between group factor has two levels, hence in order to calculate the effect sizes the Mann-Whitney test was used for the compound effects analysis.

The results Table VTI.29 in Appendix 15 shows that the 2D representation had a significant higher score than the 3D representation, $\chi^2(1) = 11.52$, $p = .001$. Also, there was an interaction effect between the interactivity treatment and representation, whereby for the interaction-group the 2D representation score was significantly higher than the 3D, $\chi^2(1) = 8.05$, $p = .005$.(see Figure 6.7 red highlight ①). Additionally, in both representations VTI increased the performance score, $H(1) = 6.46$, for the 2D representation, and $H(1) = 7.52$, $p = .006$, for the 3D representation (see Figure 6.7 red highlights ②).

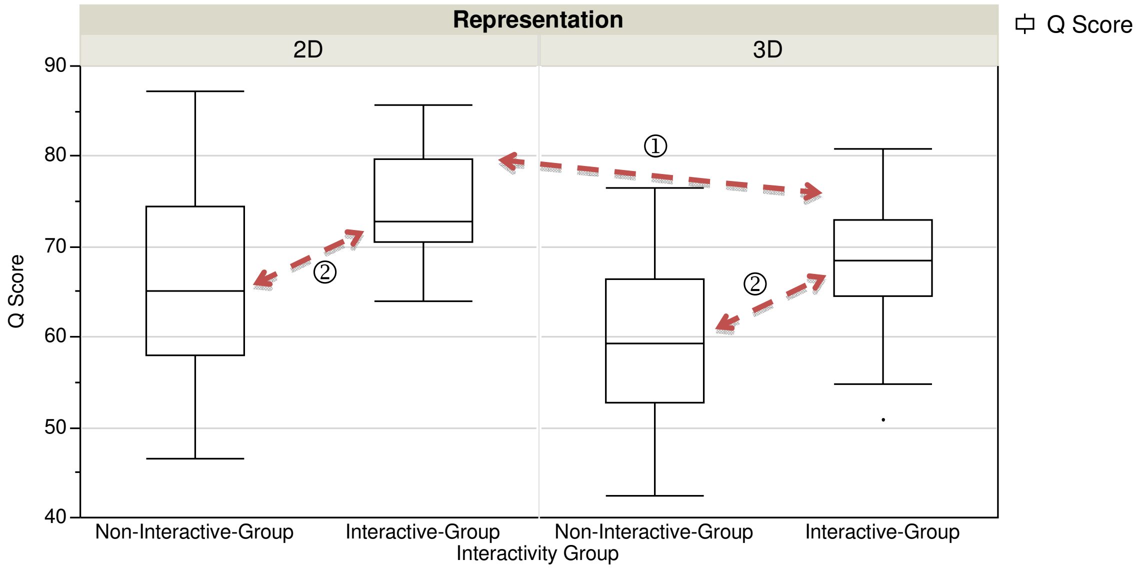


Figure 6.7 – 2D and 3D Scores by Interactivity Treatment

Table VTI.28 in Appendix 15 provides a full statistical description of the 2D and 3D scores by interactivity treatment.

Results Table VTI.31 in Appendix 15 show that the external (below mean) LoC participants in the IG (*Mdn* = 68.42) had a significantly higher 3D score than those in the NIG (*Mdn* = 59.38), $U = 21.50$, $z = -2.131$, $r = -0.48$ (see Figure 6.8), this is highlighted in red in the figure.

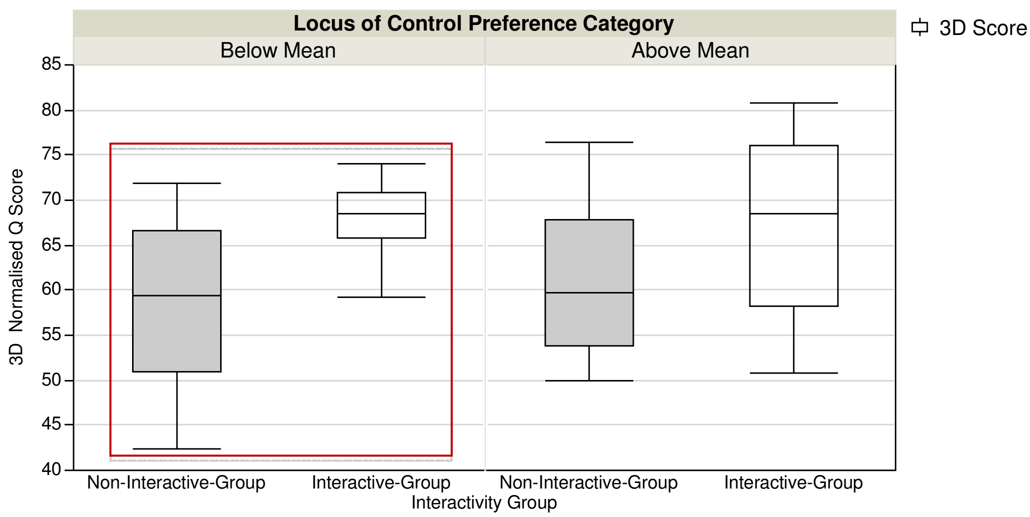


Figure 6.8 – 3D Scores by Locus of control ATI-C and Interactivity Treatment

Further the results Table VTI.31 in Appendix 15 show that both the 2D and 3D scores were significantly higher for the participants with below mean SE measures, in the IG (*Mdn[2D]* = 72.22), (*Mdn[3D]* = 68.89) than the NIG (*Mdn[2D]* = 60.85), (*Mdn[3D]* = 59.16), U = 26.00, z = –2.462, r = –0.51 (see Figure 6.9 red highlight ①) for the 2D score and U = 34.00, z = –1.970, r = –0.41 (see Figure 6.9 red highlight ②) for the 3D score.

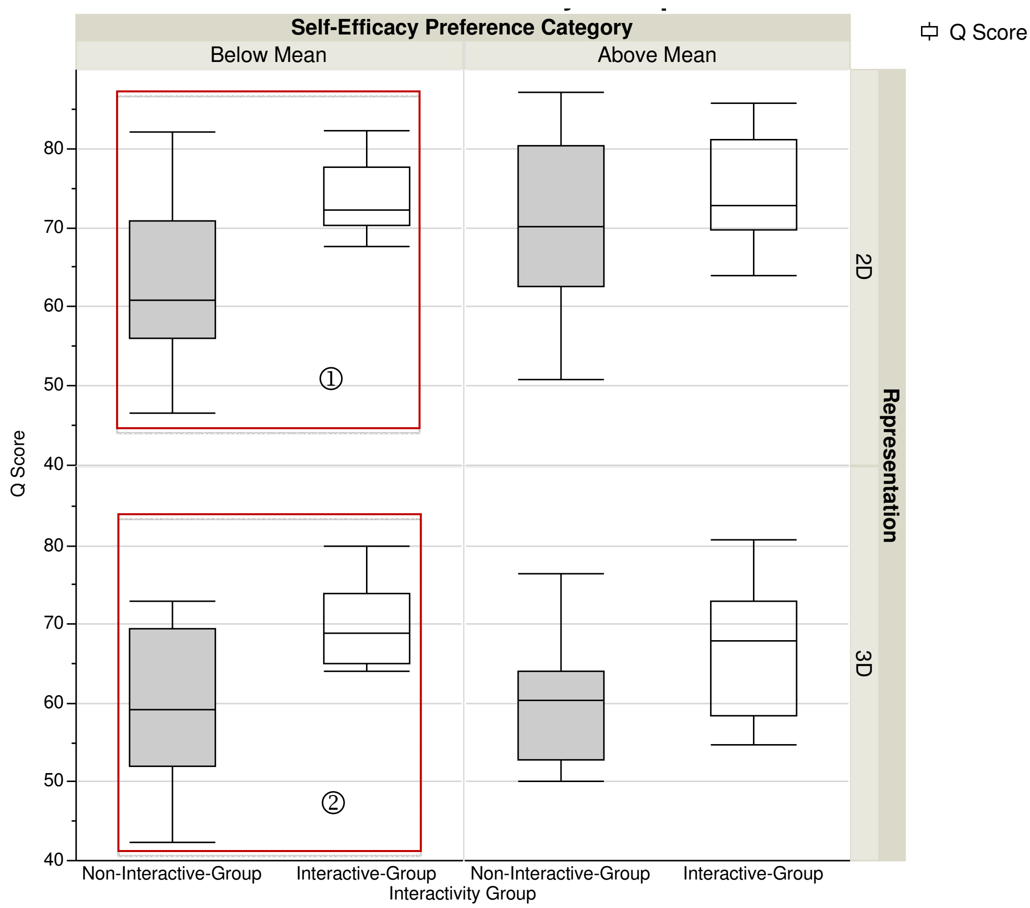


Figure 6.9 – 2D and 3D Scores by Self-Efficacy ATI-C and Interactivity Treatment

As shown in Table VTI.31, the participants with below mean SA measures in the IG (*Mdn* = 70.59) had a significantly higher 2D score than those in the NIG (*Mdn* = 60.47), $U = 16.00$, $z = -2.313$, $r = -0.53$ (see Figure 6.10 red highlight ①). Also, the participants with above mean SA measure in the IG (*Mdn* = 68.42) had a significantly higher 2D score than those in the NIG (*Mdn* = 60.66), $U = 33.00$, $z = -1.985$, $r = -0.41$ (see Figure 6.10 red highlight ②)

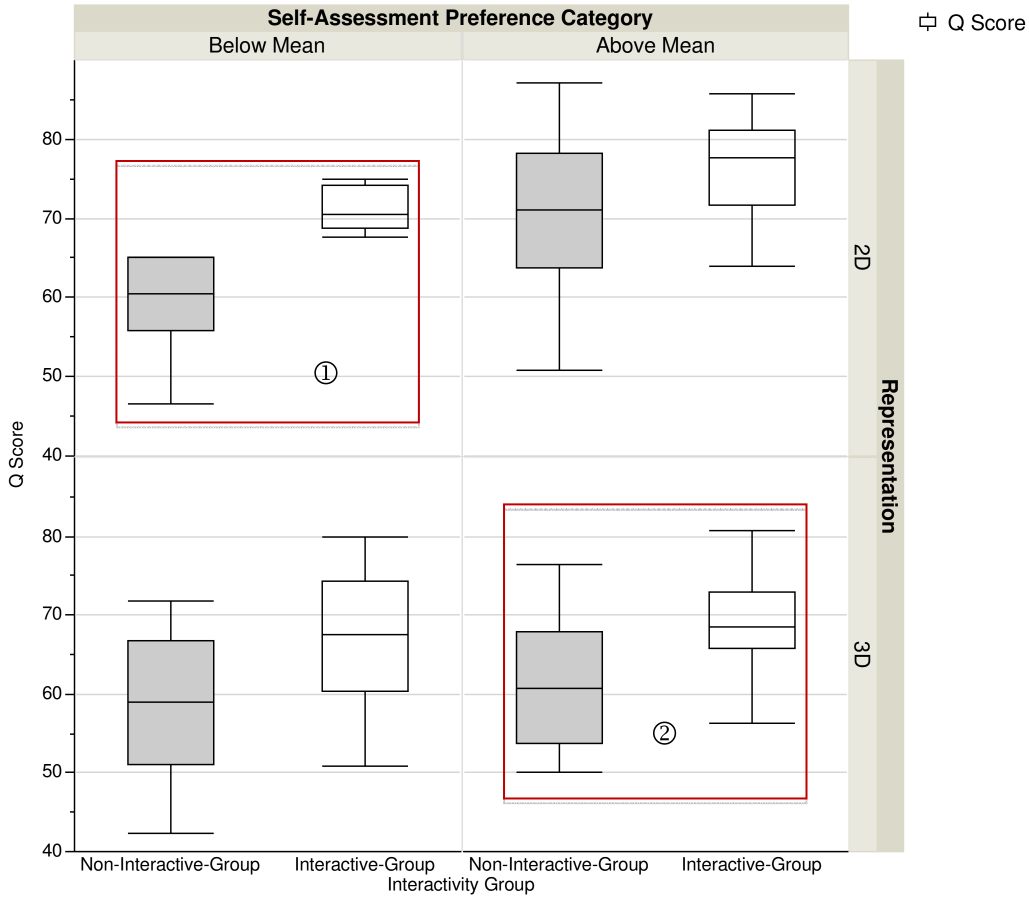


Figure 6.10 – 2D and 3D Scores by Self-Acceptance ATI-C and Interactivity Treatment

Evidence from the results Table VTI.33 in Appendix 15, demonstrated that both the 2D and 3D scores were significantly higher for the participants with above mean visual preference in the IG (*Mdn[2D]* = 72.60), (*Mdn[3D]* = 68.71) than the NIG (*Mdn[2D]* = 64.19), (*Mdn[3D]* = 59.16), U = 22.00, z = –2.006, r = –0.45 (see Figure 6.11 red highlight ①) for the 2D score and U = 21.50, z = –2.045, r = –0.46 (see Figure 6.11 red highlight ②) for the 3D score.

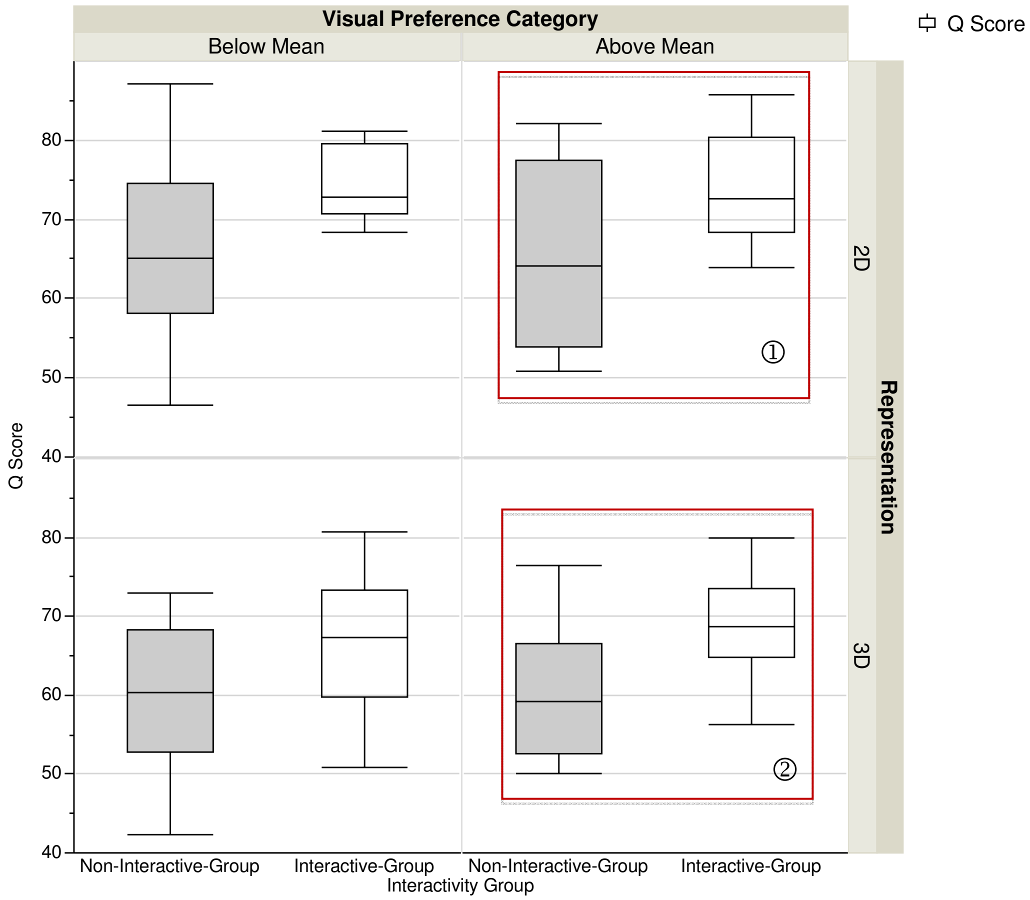


Figure 6.11 – 2D and 3D Scores

by Visual ATI-C and Interactivity Treatment

The results also show that the participants with below mean aural preference in the IG (*Mdn* = 72.22) had a significantly higher 2D score than those in the NIG (*Mdn* = 60.85), U = 10.00, z = –2.502, r = –0.61 (see Figure 6.12 red highlight ①). Also, the participants with above mean SA measures in the IG (*Mdn* = 67.91) had a significantly higher 2D score than those in the NIG (*Mdn* = 51.06), U = 22.00, z = –3.049, $p < .01$, r = –0.61 (see Figure 6.12 red highlight ②)

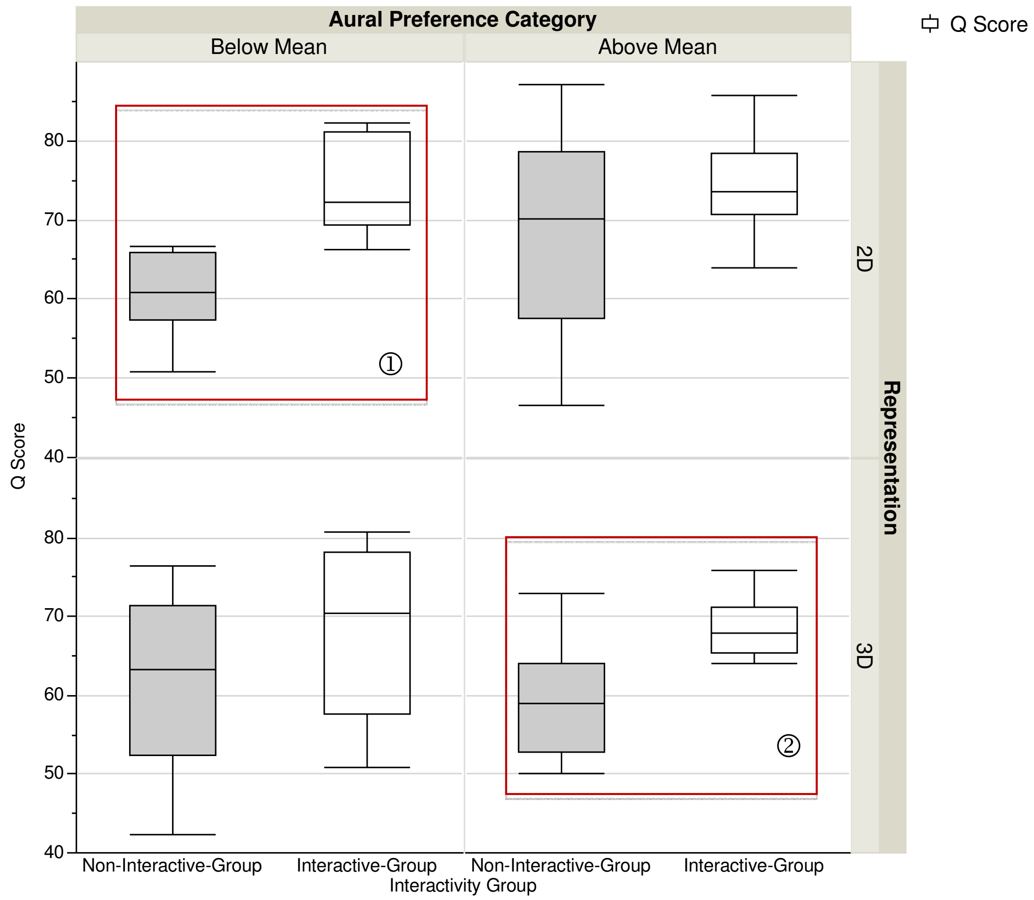


Figure 6.12 – 2D and 3D Scores by Aural ATI-C and Interactivity Treatment

The results Table VTI.34 in Appendix 15 showed that the participants with above mean read-write preference in the IG (*Mdn* = 67.39) had a significantly higher 3D score than those in the NIG (*Mdn* = 51.06), $U = 8.00$, $z = -2.769$, $p < .01$, $r = -0.65$ (see Figure 6.13), this is highlighted in red in the figure.

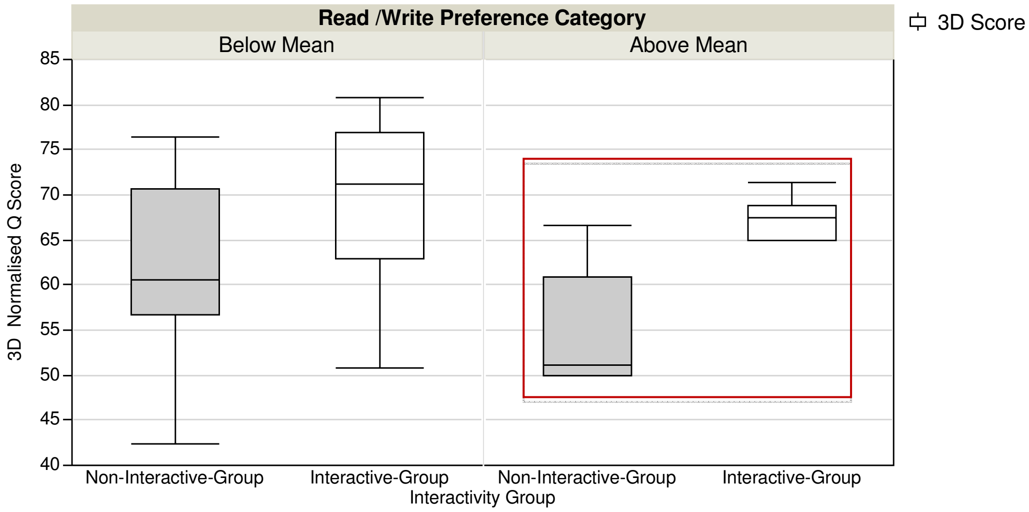


Figure 6.13 – 2D and 3D Scores

by Read-write ATI-C and Interactivity Treatment

The results also showed that the participants with above mean read-write preference in the IG (*Mdn* = 67.86) had a significantly higher 3D score than those in the NIG (*Mdn* = 59.38), $U = 24.00$, $z = -2.939$, $p < .01$, $r = -0.59$ this is highlighted in red in the Figure 6.14.

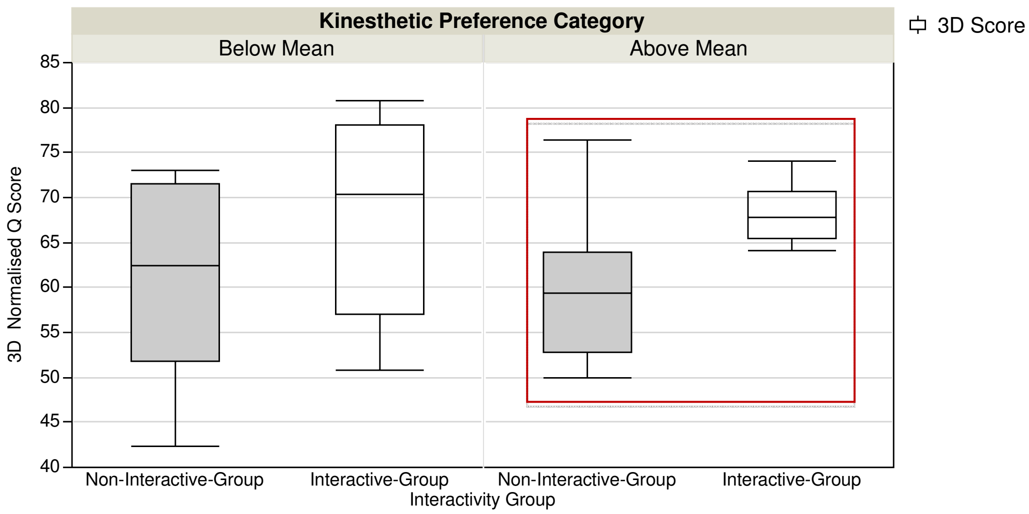


Figure 6.14 – 2D and 3D Scores by Kinaesthetic ATI-C and Interactivity Treatment

6.5 Summary

In summary, VTI in the context of exploring the data-set holistically using information visualisations, had a significant effect by improving the overall accuracy, by increasing the percentage of accurate insights (NIG: *M* = 53.47 *SD* = 10.35 – IG: *M* = 61.59 *SD* = 8.66) and decreasing the percentage of inaccurate insights gathered (NIG: *M* = 31.11 *SD* = 7.70 – IG: *M* = 24.97 *SD* = 5.97). Additionally, the results showed that VTI also improved the overall performance score (NIG: *M* = 63.06 *SD* = 8.46 – IG: *M* = 71.18 *SD* = 5.52). Moreover, there were no significant interaction effects between the individual differences studied and VTI regarding the overall insights gathering percentages.

When comparing the 2D and 3D representations, VTI had a significant effect, by improving the accuracy (NIG: *M* = 53.92 *SD* = 13.21 – IG: *M* = 65.66 *SD* = 8.22) and reducing significantly the unidentified percentage of insights (NIG: *M* = 18.91 *SD* = 10.96 – IG: *M* = 11.40 *SD* = 8.31) for the 2D representation. Also, as a main effect, the 2D representation has a significantly higher accurate percentage (2D: *M* = 59.79 *SD* = 12.39 – 3D: *M* = 55.02 *SD* = 10.18) and significantly lower inaccurate percentage (2D: *M* = 25.05 *SD* = 7.85 – 3D: *M* = 31.36 *SD* = 9.84) than the 3D representation. Otherwise, there were no three-way interactions between the interactivity treatment representation and the individual differences studied with regards to the percentage of accurate, inaccurate and unidentified insights.

When comparing the scores for the 2D and 3D representations, as shown in Table 6.4, VTI improved significantly both the 2D and 3D scores, the results also showed that the 2D score had a statistically significantly higher improvement that the 3D score ($\chi^2(1) = 8.05$, $p = .005$). With regards to the three-way interaction between VTI, visual representation and individual differences the results showed several significant compounding effects.

		2D Score			3D Score		
	N	M	SD	Mdn	M	SD	Mdn
Non-Interactive-Group	21	66.14	11.46	65.00	60.10	8.80	59.38
Interactive-Group	21	74.07	5.81	72.73	67.88	7.81	68.42

N: Number of participants, **M:** Mean, **SD:** Standard deviation, **Mdn**: Median

Table 6.4 – 2D and 3D Score Statistical Description by Interactivity Treatment

In term of psychometric measures, for the LoC measure, external LoC had significantly higher 3D scores when using VTI (NIG: *M* = 59.14 *SD* = 9.21 – IG: *M* = 67.85 *SD* = 4.25), the 2D score was not affected. For the SE measure, VTI equally affected both representations for the participants with below mean SE measures, by significantly increasing the 2D score (NIG: *M* = 63.02 *SD* = 10.88 – IG: *M* = 73.42 *SD* = 4.61) and 3D score (NIG: *M* = 60.13 *SD* = 9.66 – IG: *M* = 68.73 *SD* = 7.62), when comparing to the NIG. VTI affected the SA ATI-C differently for the 2D and 3D scores, whereby participants with below mean SA measures had a significantly higher 2D score (NIG: *M* = 62.13 *SD* = 10.93 – IG: *M* = 71.22 *SD* = 2.70) and the participants with above mean SA measures had a significantly higher 3D score (NIG: *M* = 61.35 *SD* = 8.66 – IG: *M* = 68.54 *SD* = 7.16).

For the VARK learning preferences the results showed several interaction effects between VTI and individual sensory modes for the 2D and 3D scores. The visual preference had a compound VTI effect, whereby participants with above mean visual profiles had higher scores both for the 2D (NIG: *M* = 65.06 *SD* = 11.69 – IG: *M* = 73.77 *SD* = 6.75) and 3D (NIG: *M* = 60.75 *SD* = 8.75 – IG: *M* = 68.75 *SD* = 6.71) representations. VTI affected the participants with aural preference differently for the 2D and 3D scores. The participants with below mean aural profiles had significantly higher 2D scores (NIG: *M* = 62.95 *SD* = 9.93 – IG: *M* = 73.88 *SD* = 6.06), and the participants with above mean aural profiles the 3D scores significantly higher (NIG: *M* = 58.86 *SD* = 6.86 – IG: *M* = 67.67 *SD* = 5.37). For the read-write preference and the kinaesthetic preferences the conjoint effects with VTI were significant only for the 3D scores. Whereby, the participants with above mean read-write profiles (NIG: *M* = 55.72 *SD* = 6.81 – IG: *M* = 66.37 *SD* = 6.15) and participants with above mean kinaesthetic preferences (NIG: *M* = 59.39 *SD* = 7.49 – IG: *M* = 67.39 *SD* = 4.53) the 3D scores was significantly higher using VTI compared to the NIG.

[Intentionally left blank]

Chapter 7. Discussion

This chapter provides a detailed interpretation of the results in the context of existing research and discusses the interpreted results in the context of the research objectives. The chapter is organised according to the objectives and research questions of this thesis.

7.1 Research Summary 142

7.2 Findings 143

7.2.1 Objective One 143

7.2.2 Objective Two 145

7.2.1 Objective Three 149

7.2.1 Objective Four 152

7.3 Summary 158

7.1 Research Summary

The objectives of this research were to understand the effects of interaction and individual differences on a set of visual analytics tasks. Further the research objectives look at interaction in VA using two experimental settings. Firstly, an investigation into the effects of Visual Mapping Interaction and its compounding effects with the performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles on the generation of insights, their accuracy and mental effort required in generating them in the context of performing an analytical task using information. Secondly, the emphasis was on understanding the effect of View Transformation Interaction effects and its compounding effects with the performance-related psychometric measures (LoC, SE, and SA) and the VARK model of learning styles in gathering insights into a problem data-set in the context of a problem-solving task, where View Transformation Interaction is used to explore the problem data-set using a game-based simulation using a 2D and a 3D visual representation.

The research was conducted in a group difference study design (a *post facto* study comparing a variable in two contrasting groups [105]) using an Aptitude-by-Treatment methodology, where interaction was used as an independent factor and the group's differences were the different categories of individual differences (e.g. external versus internal LoC). The studies were conducted using different information visualisation applications using two task settings, an analytical task and a problem-solving task using a video game. In the analytical task, interaction was performed by enabling users to interact with data tables by changing the visual structures in order to perform their exploration using Tableau, a visual analytic tool from Tableau Software [102]. Equally, the non-interactive users were provided with equivalent static visual structures in the form a printed material and PDF electronic documents. With regards to the problem-solving task, the interaction was view transformation centric, enabling users to change the views using different visual representations – 2D, 3D. This was achieved by using 2D and 3D versions of the Portal games from Valve Corporation [104], the IG played the game, and the NIG watched a video walkthrough of the same level as the ones been played.

7.2 Findings

The research in this thesis was constructed to address four objectives and nine resultant research questions. The findings are outlined in this section by objective and associated research question to facilitate the interpretation. (Note that, the use of the term significant refers to statistical significance at a p-value $p < .05$)

7.2.1 Objective One

Investigate the effects of Visual Mapping Interaction in the context of performing an analytical task using information visualisation.

<u>Research Question 1:</u> **Does Visual Mapping Interaction affect the number of insights generated and their accuracy, when compared to an equivalent non-interactive task?**

The results indicate that VMI significantly increases the number of insights generated. The size of the effect was moderate (r = .37), with 59.1% more insights generated when participants interacted with information visualisations. Moreover, VMI also significantly increased the number of accurate insights by 57.8%, with a moderate effect (r = .33). The number of inaccurate insights and the overall performance score were not significantly different between interaction treatments. These results contribute to further the understanding of interaction in the context of visual analytics as expressed by Pike *et al.* [8], where, they call for a better appreciation of the relationship between the elements of interaction. The important aspects to consider in this context are the inquiry process and the capacity to generate knowledge, where knowledge is correlated to insight by its progressive non-linear correlation as defined by Chang *et al.* [118]. Also, Pike *et al.*, stress the need to understand the benefits of interaction and outline that insight generation is the key effect of interaction that should be used as the key metric of investigation.

<u>Research Question 2:</u> When insights are categorised based on mental effort (inferential for high and procedural for low mental effort), does Visual Mapping Interaction have an effect on the number and accuracy of insights generated in each mental effort category?

As shown in Table 7.1, the results suggested that VMI significantly increased the number of procedural insights by 54.6%. However the number of inferential insights and the accuracy distinctions (accurate or inaccurate) for both categories of mental effort were not significantly different between interaction treatments. The mental effort approach used is similar to Green and Fisher [11], and Ziemkiewicz and Crouser [12] analytical tasks categorisation, where the procedural and inferential tasks were used to investigate analytical task visualisation performance. The difference in the research reported in this thesis was to use the output of the analytical task (insights) using the visualisation as the subject of analysis instead of the task itself. The results regarding an increased number of procedural insights are consistent with the above mentioned previous research [11] [12], where procedural tasks were executed faster than inferential tasks. In this thesis the experimental task time was fixed, within an open analytical task, hence the analogous outcome of faster procedural tasks is a higher number of procedural insights. The results showed this effect, thus suggesting that interaction facilitates the procedural insights as these require less mental effort than the inferential ones.

	N	F	p
Total inaccurate insights	42	2.07	0.159
Total inferential Insights	30	1.65	0.210
Total procedural Insights	41	4.42	0.042*
Total accurate inferential insights	30	0.89	0.353
Total inaccurate inferential insights	30	1.44	0.240
Total accurate procedural insights	41	3.17	0.083
Total inaccurate procedural insights	41	1.01	0.321

*p < .05, ** p < .01

N: Number of participants, ***F*:** F ratio, ***p*:** p-value

Table 7.1 – Oneway ANOVA Results for VMI

7.2.2 Objective Two

Investigate the compounding effects of Visual Mapping Interaction with performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles in the context of performing an analytical task using information visualisations.

Research Question 3: **Do Locus of Control, Self-efficacy, and Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences have compound effects with Visual Mapping Interaction, whereby according to the level of the different measures, there will be a significant effect on the generation of insights and their accuracy?**

Psychometric measures

In terms of psychometric measures, there were no significant compounding effects with regards to the overall performance score or the overall accuracy. The surprising result was regarding LoC, where recent research in information visualisation [9–12], has reported significant results whereby LoC influenced the task performance in information visualisation. This incongruence in the result may be due nature of the effects. Previous research [9–12], has linked LoC effects to the adaptability of the users to the visual structures using different visualisations as a distributed cognition (DCog) tool as described by Liu *et al.* [61]. The DCog framework is an internal representation is within bound of the human body and the external representations are the tools used to expand cognition. Within this framework, previous research focused more on the external representational aspects of the cognitive system (human-visualisation). Whereas the research reported in this thesis looks at the effects of the external representations (the visual structures) and the link to the internal representations interpreting these (insights). Thus, this result does not contradict the relationship between LoC and visual structures found in previous research, and suggests that when looking at the interaction process, LoC alone does not have a predictive value.

VARK Learning style

The results indicated that participants with an above mean visual learning preference had compound effects with VMI, increasing significantly the accuracy of insights with accuracy yields of (*M* = 3.38, *SD* = 3.02) for the control group and (*M* = 9.83, *SD* = 7.76) for the IG. Furthermore this increase in accuracy follows a statistically significant increasing trend associated with the increase in the visual preference. The effect size of this trend was moderate (*r* = .45). Moreover, the results indicated that the participants with an above mean read-write learning preference performed significantly better (higher overall score) in the control group (*M* = 91.43, *SD* = 12.15). Additionally, there was a statistically significant trend associated with the increase in the read-write preference, which also had a large effect size (*r* = .56).

These results are perfectly attuned to Fleming's views on VARK [27]. One of VARK's distinguishable differences with other sensory learning style profiling is the separation of the visual channel into two preferences (visual and read-write). This stems from Paivio's [22] dual coding approach, where sensory visual information is separated into verbal and non-verbal processes. With this approach in mind, the results indicate a benefit of VMI for the non-verbal channel, thus participants with that preference, will benefit from VMI in relation to the strength of their Visual preference. However, strong read-write preferences benefit the non-interactive participants. These results, extend the VR research [76–78] into effects of learning styles to the information visualisation field regarding the benefits non-verbal individual differences have on the insight generation performance.

<u>Research Question 4:</u> When categorising insights based on mental effort, do individual differences (LoC, SE, SA, V, A, R, and K) have a compounding effect with VMI with regards to the generation and accuracy of insights?

Psychometric measures

When insights were categorised based on mental effort the results suggested that there was a compound effect between self-efficacy measures and VMI, where the participants with above mean SE measures benefited from VMI by decreasing the mean mental effort yield by 65.6%, generating significantly more procedural insights. Also, participants with half a SD above sample mean SE measures, had a significant decrease in the mental effort yield of the inaccurate insights, explained by an increase in the number of inaccurate procedural insights induced by the compounding effects of VMI and SE. Otherwise, the accurate insight mental efforts yield in the control group follows a significant decreasing trend as the SE level increases.

These results regarding SE are inline with expectations, whereby computer tools would improve the SE [119] and reduce the barrier to effort [120]. Further, SE depicts the individual's confidence in their own capacity to perform a specific task well [121]. Participants with a high level of SE will feel confidence in their capacity to generate insight, but this has no effect on actual performance, thus the results show an increase in the number of inaccurate insights as the number of lower mental effort insights increases.

There were no significant results regarding, Self-acceptance. SA research [26], has associated resiliency to stress and higher effectiveness and a possible interpretation for this outcome is that participants did not perceive this study as stressful. To test this explanation the NASA-TLX could be used as a proxy to evaluate stress. Table VMI.1 shows that the NASA-TLX overall workload score for the IG was 71.05 out of a possible maximum of 150, which equates to 47.37% and 48.47% for the NIG, both percentages are below 50%. Also, when looking at the individual workload measures out of possible maximum of 50, all scores were below 53%. This low overall workload together with a low frustration workload (less than 40%) can be interpreted as an overall low stress in the experiment; explaining the lack of significant results regarding the SA measure.

	NIG		IG	
	M	SD	M	SD
Overall Workload Score	72.71	15.43	71.05	14.84
Mental Demand Score	26.43	12.28	26.00	12.75
Physical Demand Score	3.38	8.86	1.62	3.28
Temporal Demand Score	22.57	15.66	20.48	11.03
Performance Score	12.86	11.17	16.29	11.74
Effort Score	23.52	11.05	26.48	12.24
Frustration Score	20.38	14.54	15.71	14.40

M: Mean, **SD:** Standard deviation

Table 7.2 – Visual Analytics NASA-TLX Scores

Learning style

The results show a significant trend associated with the read-write learning preference. For the IG the number of accurate procedural insights generated decreased as the read-write preference strengthened, whilst the inferential insights remained constant. On the other hand, the kinaesthetic learning preference had also compounding effect with VMI; whereby participants with half a SD below sample mean kinaesthetic preference had a significantly different mental effort yield for accurate and inaccurate insights due to a higher mean number of accurate procedural insights and an equivalent number of inferential and procedural inaccurate insights. Further, for the participants within ± half a SD from the sample mean in the kinaesthetic preference, VMI decreased significantly their inaccurate mental effort yield; this can be attributed to an increase in the number of inaccurate procedural insights generated by this sub-group.

This kinaesthetic related result does not tally with the VARK concepts regarding the kinaesthetic preference. In the VARK framework individuals with kinaesthetic preferences would be expected to ‘learn by doing’ and VMI would be considered a form of action driven learning [27]. Thus, the expectation would be that individuals with half a SD above the sample mean would have a significant conjoint effect with VMI. An interpretation for this mismatching result would be that, the participants with half a SD below sample mean in the kinaesthetic preference, opted for other sensory modes in a multimodal sense of VARK. Table 4.4 – VARK Preference Frequencies in chapter four outlines the fact that there were no single kinaesthetic preference participants in the experiment and 76.2%, 32 participants out of the 42 constituting the sample, had a kinaesthetic component in their multimodal profile.

7.2.1 Objective Three

Objective three investigate the effects of View Transformation Interaction in the context of a problem-solving task, where View Transformation Interaction is used to explore the problem data-set using a game-based simulation using a 2D and 3D visual representation.

Research Question 5: Does View Transformation Interaction affect the number and accuracy of insights identified in a problem data-set represented in a game-based simulation, when comparing to an equivalent non-interactive task?

Combining the 2D and 3D representations, the results indicated that the use of VTI significantly increases the overall percentage of accurate insights gleaned into the data-set by 15.2%. Moreover, VTI also significantly decreased the overall percentage of inaccurate insights by 19.7%, both these effect were moderate. Subsequently, VTI significantly increased the overall score by 12.9%, with a large effect. However VTI did not affect significantly the percentage of unidentified insights (i.e. 'don't know' answers in the post-experiment questionnaire).

In the context of this experiment Portal 2D and 3D representations were used as the view transformation for the problem-solving data-set. These results reinforce the argument claimed by the serious games community [122–124] with regards to performance enhancement in terms of educational and informational benefits, contributing to the debate by providing evidence of the VTI benefits. In the context of the game-based problem set visualisation used in this thesis, VTI yielded mean combined representation accuracy performance scores of 71%, where the control group scored 63%. This increase in performance is significant, when aiming at gleaning insights into problems sets.

<u>Research Question 6:</u> Does the view representation – 2D and/or 3D, have an effect on the number and accuracy of insights into a problem data-set represented in a game-based simulation?

When considering the effect of the view representation without taking in to account VTI, the 2D representation significantly outperformed the 3D representation, in the following measures:

- 8.7% more accurate,
- 20.1% less inaccurate insights, and,
- 7.3% overall better score.

However, the percentage of unidentified insights did not differ significantly between 2D and 3D representations.

There is still a debate regarding 2D versus 3D data visualisation [125] in visual data mining in the context of information visualisation. Research by Tanvanti and Lind [68], suggest that 3D displays are better with regards to cognitive spatial abilities and memory related tasks than 2D. The research in this thesis, suggests that in the context of exploring a problem set and gathering insights into the problem, 2D representation performed better. However, recent research [126], point to difficulties some users may have with 3D environments and the possible disorientations caused by higher degrees of freedom and richer interactions available. Hence the outperformance found in this thesis research could be due to these difficulties.

<u>Research Question 7:</u> Does the representation – 2D and/or 3D and View Transformation Interaction have an interaction effect with regards to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?

When taking VTI into consideration and investigating the conjoint effects with the representations, there were significant interaction effects with the representation, whereby for the 2D representation there was:

- 15.1% more accurate insights
- 39.7% less unidentified insights
- 11.9% overall better score.

However, VTI within the 2D representation did not affect significantly the percentage of inaccurate insights.

For the 3D representation, VTI significantly increased benefit the performance score by 15.2%, and did not affect significantly the accuracy or identification of insights. When comparing 3D representation with the 2D representation; the 2D outperformed 3D by 6.3% in the overall performance score.

As in the previous research question discussion, research by Baumgärtner *et al.* [126] suggest difficulties by some users with richer interactions provided by 3D representations, Further, although the NASA-TLX overall workload showed that the 2D representation (M = 62.21, SD = 19.99) had a marginally higher workload that the 3D representation (M = 59.40, SD = 20.50), the mental workload for the IG was noticeably greater for the 3D (M = 25.43, SD = 11.43) than the 2D representation, (M = 29.90, SD = 12.51), which represents a 9% higher workload. This higher workload experience by the participants in the 3D could account for the lower performance.

7.2.1 *Objective Four*

Investigate the compounding effects of View Transformation Interaction with performance-related psychometric measures (Locus of Control, Self-efficacy, and Self-acceptance) and the VARK model of learning styles; in the context of a problem-solving task, where View Transformation Interaction is used to explore the problem data-set using a game-based simulation.

<u>Research Question 8:</u> **Independently from the representation, do Locus of Control, Self-efficacy, Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences have a compound effects with View Transformation Interaction, whereby according to the level of the different measures, there will be a significant effect to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?**

The results showed that independently from representation there were no compounding effects between VTI and any of the individual differences studied in this thesis. The results of the next research question explore the compounding effects of VTI with individual differences taking representation as a factor.

<u>Research Question 9:</u> **Does the representation – 2D and/or 3D and View Transformation Interaction have compounding effects with Locus of Control, Self-efficacy, Self-acceptance, and Visual, Aural, Read-write, and Kinaesthetic learning preferences, whereby according to the level of the different measures, there will be a significant effect to the number and accuracy of insights identified in a problem data-set represented in a game-based simulation?**

The results showed that there were significant compounding effects between the individual differences and VTI within specific visual representation only for the performance scores[1]. The accuracy and identification metrics did not show any significant compounding effects.

---

[1] $Score = \frac{Accurate}{Accurate + Inaccurate} \cdot 100$

Psychometric measures

The results indicated that VTI had a compounding effect with external locus of control participants, where the 3D performance score was significantly better in the interactivity group with a median 3D score 15.2% better than the NIG. Otherwise the 2D score did not have any significant compounding effect between LoC and VTI.

Individuals with an externally focused LoC are more inclined to trust external representations [12] than the internally focused LoC, which can explain the gain in performance when interacting.

This difference between representation results could be interpreted by the different generalised expectancies for the different situations, as defined by Rotter [23]. The generalised expectancy is linked to previous experience, prior behaviour and responses association. In the case of the experiments in this thesis, the 3D representation was closer to reality, and the user would relate better to the world represented and exploration behaviour needed. This would point to a closer match in the generalised expectancy than with the 2D representation. Hence, explaining the results that VTI helped the participants in the 3D representation understand the problem better.

The self-efficacy measure had a significant compounding effect with VTI, whereby participants with below mean SE measures increased their 2D score by 18.7% and the 3D score by 16.5%. SE and LoC are significantly correlated $r = .43$, $p < .01$ (see Table 4.9). Hence some of the LoC effects translate into these results. In the LoC results the 2D score was not affected, although there was a near significant increase ($p = 0.053$, $r = -.43$) which had a moderate effect. SE can be thus considered as a complementary measure to the LoC. SE, measures the expectancy of performance [127]. In this case, a possible interpretation for a significant score in 2D is that participants had overall a higher expectancy of success in the 2D task, this could be confirmed by the pre-test questionnaire results where, 92.9% had previous experience in 2D gaming. For the participants with above mean SE measures this tendency elevated the 2D score for both interaction treatment groups (*Mdn* = 70.15) for the NIG and (*Mdn* = 72.73) for the IG. Where as the participants with below mean SE measures, the expectancy of success was lower and had lower performance scores, (*Mdn* = 60.85) for the NIG. In contrast in the IG, the participants with below mean measure of SE retained their high expectancy having median performance score (*Mdn* = 72.22) close to the participants with above mean SE measures. This suggests that VTI facilitated the understanding of the problem set for the participants with below mean measures, bringing their median score close to the participants with above mean measures, under the same conditions. A similar argument would apply to the 3D score for the participants with below mean SE measures, with regards to VTI compounding effects. Nevertheless, the results were not significant, although borderline ($p = 0.049$, $r = -.48$) with a moderate size effect. This would suggest that the conjoint VTI effects with the participants with below mean SE measures could be questioned. However, when taking into consideration the correlation of SE with LoC and the prior experience in 3D game of 52.4%, the significant compounding effect of the externally focused LoC with VTI, could indicate that SE actually has little, to no compounding effect with the 3D score and the majority of the effects are attributable to LoC.

Regarding the self-acceptance results, VTI had a different compounding effect for the 2D and 3D scores. VTI for the participants with below mean SA measures increased the 2D score by 16.7%. In contrast VTI for the participants with above mean SA measures increased the 3D score by 14.7%. SA and SE are strongly correlated $r = .62$, $p < .001$, and the SA measure, differentiates itself from SE, by considering aspects of resiliency to stress and acceptance of frustration, as part of life. In the case of the 2D representation, as discussed earlier the overall expectancy of success based on previous experience is high. As with the SE measure, the participants with below mean SA measures had a relatively lower expectancy of success in relation to the participants with above mean SA measures, demonstrated by their median performance scores (*Mdn* = 60.47) for the NIG and (*Mdn* = 70.59) for the IG. Thus, VTI provided the enabling support to gathering insights about the problem set, bringing the 2D score closer to the overall median 2D score (*Mdn* = 70.78). As for the 3D performance score, a possible lower expectancy of success, extrapolated from the low level of prior experience and the 3D representation higher score in the NASA-TLX mental workload, could be interpreted as the 3D representation experiment having a higher stress level. Thus, as seen in the psychometric review in section 3.1.1, leadership assessments using this scale [83] relate high performance, with high levels of SA. Therefore the results indicate that high levels of SA measures have compounding effects with VTI increasing the 3D performance score.

Table 7.3 provides the statistical description for the different visual representation by interactivity treatment.

	2D				3D			
	N	M	SD	Mdn	N	M	SD	Mdn
Non-Interactive-Group	21	66.14	11.46	65.00	21	60.10	8.80	59.38
Interactive-Group	21	74.07	5.81	72.73	21	67.88	7.81	68.42
All	42	70.10	9.83	70.78	42	63.99	9.11	66.09

N: Number of participants, **M:** Mean, **SD:** Standard deviation

Table 7.3 – 2D and 3D Scores Statistical Description by Interaction Treatment

Learning style

Visual preference had a significant compound effect with VTI, whereby participants with above mean visual preference increased their 2D score by 13.1% and their 3D score by 16.1%. These results are expected, as the activity in the experiment is primarily visual, hence the participants with above mean visual preferences, out-perform the participants with below mean visual preferences. Nevertheless, both interaction treatments were equally visual by design, and the interaction was isolated to understand the effects of VTI. In the case of VTI, the interaction occurs with the views, by transforming the viewpoint, and other visual elements, hence it is understandable that the performance would be enhanced by VTI conjointly with higher levels of visual (non-verbal) preference. Yet, it is surprising that the effects are not more noticeable as the effect size is only moderate. Comparatively, the participants with above mean visual preferences, had a median score of (*Mdn* = 72.60) for 2D and (*Mdn* = 68.71) for 3D, which are not the highest of the VARK set of participants' above mean individual preferences, the kinaesthetic preference has a higher median performance score for both the 2D and 3D representation. More research adding aspects of spatial ability [128, 129] may be required to get a fuller picture of the visual channel preference.

Regarding the aural preference results, VTI had a different compounding effect for the 2D and 3D scores. The participants with below mean aural preference in the interactivity group increased the 2D score by 18.7%. In contrast the participants with above mean aural preference increased the 3D score by 15.2%. These results are very surprising, as the aural preference is generally associated with auditory input, verbal collaboration and think out-loud type of behaviours and responses [27], which the participants did not perform. Another observation is that in this sample there are a disproportionally high number of multimodal participants 83.3%, 35 participants out of 42, of which 94.3%, (33 participants out of the 35) have an aural plus other sensory mode(s). In the context of the 2D representation, the audio output of the game was independent from the VTI as all the sounds were omnipresent in the view. The increase in the 2D score for the below mean aural participant, could be due to an increased reliance on other modal preferences which interacted in a more complex compound way with VTI. These multimodal effects require a very large sample to investigate thoroughly, as in Fleming's research [117], where the sample was N = 62094. Further these multimodal differences in performance are consistent with research by Ramirez [130]. For the 3D representation, the game audio is directly related to VTI in surround sound effects, which make the sound part of the interaction. Hence, in the same multimodal rationale the participants with above mean aural preference leveraged their auditory preference with other sensory modes to benefit from VTI in the way that increased their performance as the results showed.

With regards to the read-write preference, the results indicated that VTI had a compounding effect with the participants with above mean read-write preferences. They performed significantly better with in the VTI group with a median 3D score 32% better than the NIG with a statistically large effect. On the other hand, there were no compound effects with VTI for the 2D score. These are interesting results; the non-interactive participants with above mean read-write preference had the lowest median score (*Mdn* = 51.06) of all VARK aptitude-by-treatment categorisation sub-groups, with a median score 16.3% below the overall non-interactive 3D median score (see Table 7.3). For these participants the preference is the verbal process, in the dual coding approach [22]. As the experiments were designed as non-verbal, and visually coded, it would explain this lack of performance for the participants with high levels of read-write preference. Nevertheless, even when the preference is highly verbal, VTI enables the participants with above mean read-write preference to raise their median 3D score (*Mdn* = 67.39) closer to the overall median score for the NIG (*Mdn* = 68.42).

Lastly for the kinaesthetic preference, VTI had a significantly large size compounding effect, whereby; the participants with above mean kinaesthetic preference, increased their 3D score by 14.3% and had no compounding effects on the 2D representation.

Fleming define the kinaesthetic preference as a *“perceptual preference related to the use of experience and practice (simulated or real)”* [27; p.1]. VTI with the 3D representation falls under this description more so than the 2D representation. Hence, an increase in the 3D scores for the participants with above mean kinaesthetic preferences was perfectly explained by the VARK preference model in this instance, also concurring with recent research in VR using the same learning style instrument.

7.3 Summary

These results provide tangible proof of the benefits of VMI in terms of insight generation and improved accuracy in analytical tasks. They also provide that interaction facilitated the lower mental effort procedural tasks. In terms of individual differences the results also showed significant compounding and predictive effects, in particular with the VARK learning style model, where the division of the visual channel into visual and read-write offered good predictability value and opposite in compounding effects. On the other hand when looking at the individual differences interaction effects with mental efforts in the context of VMI, the results showed self-efficacy was the key predictor of success in terms of performance for the lower mental effort procedural tasks in the context of VMI.

The VTI results strengthened the observation that interaction with the problem-set using view transformation improves the understanding and number of insights gleaned into the problem. Further, the results allow a postulation that 2D representation of the problem set is more effective than 3D representation. The results also showed that individual differences alone had no compounding effects with VTI. The individual differences compounding effects with VTI existed in conjunction with the type of visual representation. The results showed that for the 2D representation, SE and SA psychometric measures, as well as visual and aural learning preferences had predictive values. Whereas for the 3D representation, LoC had the strongest predictive value for the psychometric measures and the full VARK set of individual uni-modal learning preferences had a predictive value on the insight generation performance.

These findings contribute to justify the measurement and use of psychometric and learning preferences constructs as performance prediction measures for visual analytics. Further, these results contribute to the advancement of the human characterisation that can be used to profile and select high performing visual analysts, by adding the individual differences measures studied in this thesis to the profiling mix.

Chapter 8. Conclusions

This chapter provides a summary and conclusion to the main research findings. It also outlines the key contributions to the field of visual analytics and suggests future research that would further the knowledge into the interactivity benefits.

8.1 Conclusion 160

8.2 Contributions.. 162

8.3 Future Research 166

8.1 Conclusion

The aims of the research were to understand the effects of interaction and the compound effects individual differences have in generating insights and gather insights using information visualisations.

The first set of research question investigated the VMI aspects in generating insights. The results showed that VMI had a significant impact in the generation of insights moreover it assisted the generation of more accurate insights. In terms of mental effort, VMI facilitated the generation of insights requiring a lower mental effort. Furthermore, VMI interaction effects with individual differences were investigated and SE was the only psychometric measure that was found to have a compounding effect. Participants with an above mean SE measure tended to generate more lower mental effort insights when using VMI, whilst the top end (half a SD above the sample mean) decreased their accuracy.

Regarding the VARK learning preferences sensory model. VMI and the Visual preferences had a conjoint effect increasing the accuracy of insights as a function of the higher levels of the visual preference. At the same time the read-write preference had a compounded trend with VMI decreasing the accurate number of lower mental effort insights, as the Read-write profile measure increased. These results reinforced the separation of the visual channel into verbal (R) and non-verbal (V) modes instigated by Paivio [22]. Further, the results for the kinaesthetic preference in this research highlight the importance of the multimodal aspects, as the kinaesthetic profile measure results are not fully explained by the VARK model.

The second part of the research investigated the VTI in the context of gathering insights from a problem-solving data-set. The results showed that VTI had a significant impact in the performance of users in gleaning an understanding into the problem set. Further, the results showed that the 2D representation lends itself to higher performance gains when using VTI than 3D. Furthermore, VTI interaction effects with individual differences were investigated, showing that that all individual differences had some compound effects with VTI.

For the psychometric measures, the results show that external LoC persons increased their 3D scores with VTI. Participants with below mean SE measures increased both their 2D and 3D scores. Additionally, self-acceptance compounded with VTI in a different manner for 2D and 3D, whereby, the participants with below mean SA measures, performed better in the 2D representation and the participants with above mean SA measures, had higher 3D scores. These psychometric results shows the factorial interaction between these historically performance-related measures.

The VARK profiles also had diverse compounding effects with VTI. The visual preference interacted with VTI increasing the 2D and 3D scores of the participants with above mean visual preference. The aural preference and VTI had different associated effects, where the 2D score was higher for the participants with below means aural preferences and the 3D score increased for the participants with above mean aural preferences. The read-write preference and kinaesthetic preferences compounded with VTI in increasing the 3D score alone for the participants with above mean measures, for both preferences. The interpretation and analysis of the results call for more in-depth research into the effects of multimodality in the interaction between VTI and VARK profiles.

8.2 Contributions

This thesis addresses the challenges posed by the visual analytics field with regards to advancing the 'science of interaction' [5, 8]. The contributions to the visual analytics field are addressed via the objectives of this research. They are the evaluation of the benefits of interaction in terms of insight generation, as well as gaining an understanding of the compounding effects the performance-related psychometric measures and learning styles studied have with interaction.

This research investigated the benefits of visual mapping interaction, studying the effects that visual structure interaction have in a data analysis task. Additionally, this investigation looked at the benefits of view transformation interaction, examining in a problem-solving game environment the effect of view changes in gathering meaningful understanding of the problem set. These studies used an insight-based evaluation approach, where insights are defined as units of discovery, knowledge gained from the use of the interaction in terms of data or relationships in the data. When investigating the compounding effects of individual differences, this thesis presents the usefulness of the ATI methodology to provide individual differences group comparison for insight-based experiments.

1. **Visual mapping interaction effects**

With regards to the visual mapping interaction this thesis contributes by affirming that interaction increases insight generation, increasing the mean by 60%. Additionally, this research shows that interaction improved the accuracy of insights generated, by a mean improvement close to 60%. Further, it proposes that interaction facilitates the generation of insights requiring a lower mental effort, where the increase was 55% when compared to an equivalent non-interactive task.

2. **Compound effects of visual mapping interaction with performance-related psychometric measures (LoC, SE, and SA) and the VARK model of learning styles**

This research contributes, by identifying compounding effects between interaction and Self-efficacy, where SE facilitates the generation of low mental effort insights. Whereby, for the participants with above mean SE measures, interaction decreased their mental effort yield, defined as the difference between high and low mental efforts, by over 65%. Other contributions are that no significant effect were attributable to LoC or self-acceptance measures.

With regards to the VARK model of learning styles, this investigation contributes by demonstrating an overall accuracy trend related to compounding effects with the visual preference, increasing the accuracy of insights proportionally with the increase in the visual score. For above mean visual users, the increase in their accuracy yield (difference between accurate and inaccurate insights) was close to 200%. Additionally, interaction has a negative correlating trend related to the read-write preferences, effectively dropping by 100% the number of accurate insights requiring a lower mental effort procedural tasks between the users with low and high levels of read-write preference.

3. **View transformation interaction effects in the context of 2D and 3D visual representations**

With regards to the view transformation interaction, this thesis confirms that in the context of a 'serious games' a problem-solving interactive exploration increases the number of accurate insights gathered by 15%, reduces the inaccurate insights by 20%, and improves the performance score[2] by 13%.

When looking at the visual representation in isolation, the results show that interaction in a 2D representation provides 15% more accurate insights than watching a video walkthrough of the problem set, reduces by 40% the number on unidentified (unnoticed, 'missed') insights and improves the overall has performance score by 12%. Equally, interaction in a 3D representation increased the performance score by 15%, whilst having no significant effects on the accuracy or identification of insights.

[2] $Performance\ Score = \frac{Accurate}{Accurate + Inaccurate} \cdot 100$

When comparing the performance scores of interaction in 2D versus a 3D representation, interaction in a 2D representation outperformed 3D by 6% overall and specifically by 8% in the number of accurate insight whilst reducing by 20 % the inaccurate insights.

4. **Compound effects of view transformation interaction with performance-related psychometric measures (LoC, SE, and SA) and the VARK model of learning styles, in the context of 2D and 3D visual representations**

The results of this thesis show that view transformation interaction has a conjoint effect with all the performance-related psychometric measures studied namely Locus of Control, Self-Efficacy, Self-Acceptance and the VARK single sensory modes. Low self-efficacy individuals benefited from the view transformation interaction by improving their performance score by 19% and 16.5% within 2D and 3D respectively. The external locus of control individuals benefited from interaction within a 3D representation only improving their performance score by 15%. Whereas, the low self-acceptance users increased their performance by 17% in using the 2D and the high self-acceptance user improved their performance by 15% using the 3D. Regarding the VARK learning preferences single sensory modes. For users with high levels of visual learning preference, interaction increased their performance score by 13% and 16% within 2D and 3D respectively. For the participants with high levels of read-write learning preference, interaction had a very large compensating effect on the 3D representation by increasing their score by 32%. Finally the participants with high kinaesthetic learning preference, benefited from interaction by increasing their performance score using 3D by 14%.

5. Key contribution

The key contribution of this thesis is the tangible proof of the benefits of both the Visual Mapping and View Transformation Interaction in visual analytics. Results obtained strengthens the view held by the visual analytics community that interaction with the problem-set, improves the understanding and amount of insights gleaned into the problem. Further, this thesis indicates that within the context of this study a representation of the problem set using a game-based simulation, is more effective in 2D than 3D. Finally, these findings contribute to justify the measurement and use of performance-related psychometric such as Locus of Control self-efficacy and self-acceptance in addition to the VARK learning preferences constructs as performance predictors for visual analytics. Further, the results of this investigation, provides profiling and selecting measures, that can be used to identify high performing visual analysts.

8.3 Future Research

Further research into the benefits of interaction and the compounding effects of individual differences is still required. The aims of this future research should highlight the interactivity benefits into finer detail, as well as provide confirmatory research to the present thesis. Additionally, investigating the effects of the different visual mappings have would be greatly beneficial.

Future work should be an insight-based evaluation investigating visual mapping interaction in more depth. Using the same categorisation of insight into low and high mental effort for the dependent variable. As for the independent variable, sub-dividing the interaction variable into sub-categories of interaction using an interaction taxonomy such as [62]. Defining the interaction techniques into the following categories.

- Select- mark something as interesting
- Explore- show me something else
- Reconfigure- show me a different arrangement
- Encode- show me a different representation
- Abstract/Elaborate- show me more or less detail
- Filter- show me something conditionally
- Connect- show me related items.

Future research should consider using a large participant pool using a crowd sourced based experiment as in [12], using a platform such as Amazon's Mechanical Turk. This would enable analysis of the full VARK preference spectrum of 23 profiles as described in Table 4.3. This would enable the exploration of the full multimodal aspects of learning style profiling. Additionally, building on the approach taken in [68], the Myers–Briggs Type Indicator (MBTI) could be added, as this would complement the learning style [87] as a valid and reliable additional measure.

In terms of design, the experiment settings should be constructed using the interaction taxonomy mentioned above. The type of interaction could also be considered i.e. data transformation, visual mapping or view transformation according to Card *et al.* [2] framework. Thus, each experiment treatment should isolate an interactivity category component and use a single technique of each type required. For example, the category *Select*, works as a preceding operation to another type of interaction. Therefore a possible experiment setting could be a scatter plot in which user can only selects one or more data point of interest (a cluster for example), and rotate the view. Addition for the same interactive setting provides different visual mapping options. These kind of experimental settings would enable researchers to identify, and catalogue the performance in each interaction and visual mapping category mix according to a learning style and personality profile mix. Additionally, they would permit the comparison of the different interaction and visual mapping categories between each other on the basis of the number of insights generated and the mental efforts required.

The key challenges faced by these studies is the recruitment of a large enough participant sample for a statistically powerful experiment. An option with the crowd sourced experiment is to run the experiments over a long period of time, perhaps several months, similarly to the VARK website [131]. This permitting the accumulation of a sufficiently large sample. Using a crowd sourcing approach has a few challenges also. In particular it is impossible to control the participants environment compared to a laboratory controlled experiment. While this limitation should be considered, the study can be designed to minimise the effects by considering the diversity of computing environments in the design and defining the tasks in such a way that it is not possible to use Internet resources to solve them. Ziemkiewicz *et al.* [12] outline that although there is still some reticence in using crowd sourcing methods in the HCI and visualisation communities, it is gradually becoming more accepted as a user study platform This is thanks to improvements in tools such as Amazon's Mechanical Turk which addresses the perceived limitations such as, the possibility of vote flooding and the lack of incentive for completion.

Other challenges include the development of the tools for the experiment. The isolation of the interactive features is a prerequisite in this design; therefore the interactive visualisation would need to be a web-based bespoke application. A possible avenue would be to develop a mobile application that would run the experiment and upload the data online, capturing the insights via voice recordings. Additionally, with the recent trend in 'gamification' of scientific experiments [132], and its increasing acceptance as a motivational tool [133], this method would be a very promising avenue to explore to access a large pool of participants from which to collect data.

[Intentionally left blank]

References

[1] S. Card, J. Mackinlay, and B. Shneiderman, *Readings in information visualization: using vision to think*. San Francisco, CA: Morgan Kaufmann, 1999, p. 686.

[2] "Data, data everywhere | The Economist," *Feb 25th The Economist*, 2010. [Online]. Available: http://www.economist.com/node/15557443. [Accessed: 15-Sep-2012].

[3] J. Tukey, *Exploratory data analysis*. Reading, MA: Addison-Wesley, 1977, p. 688.

[4] P. C. Wong and J. Thomas, "Visual analytics," *IEEE Computer Graphics and Applications*, vol. 24, no. 5, pp. 20–21, 2004.

[5] J. J. Thomas and K. A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA: IEEE Computer Society, 2005, p. 184.

[6] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler, "Challenges in visual data analysis," in *International Conference on Information Visualization*, 2006, pp. 9–16.

[7] C. Upson, T. Faulhaber, D. Kamins, D. Laidlaw, D. Schlegel, J. Vroom, and R. Gurwitz, "The application visualization system: A computational environment for scientific visualization," *IEEE Comput.Graphics Appl.*, vol. 9, no. 4, pp. 30–42, 1989.

[8] W. A. Pike, J. T. Stasko, R. Chang, and T. A. O�Connell, "The science of interaction," *Information Visualization*, vol. 8, no. 4, pp. 263–274, 2009.

[9] T. M. Green, D. H. Jeong, and B. Fisher, "Using personality factors to predict interface learning performance," in *2010 43rd Hawaii International Conference on System Sciences*, 2010, pp. 1–10.

[10] T. M. Green and B. Fisher, "Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction," in *Visual Analytics Science and Technology (VAST), IEEE Symposium on*, 2010, pp. 203–210.

[11] T. M. Green and B. Fisher, "Impact of personality factors on interface interaction and the development of user profiles: Next steps in the personal equation of interaction," *Information Visualization*, vol. 11, no. 3, pp. 205–221, 2012.

[12] C. Ziemkiewicz, R. J. Crouser, A. R. Yauilla, S. L. Su, W. Ribarsky, and R. Chang, "How locus of control influences compatibility with visualization style," in *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2011, pp. 81–90.

[13] W. Aigner, “Understanding the role and value of interaction: First steps,” in *2011 International Workshop on Visual Analytics*, 2011, pp. 17–20.

[14] C. Ware, *Information visualization: perception for design*, vol. 22. San Francisco, CA: Morgan Kaufmann, 2004, p. 486.

[15] “insight, n.1: Oxford English Dictionary.” [Online]. Available: http://www.oed.com/view/Entry/96831#eid349741. [Accessed: 10-Jul-2012].

[16] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, “Seven Guiding Scenarios for Information Visualization Evaluation,” Unpublished report, University of Calgary, Canada, 2011.

[17] K. H. Lim, I. Benbasat, and P. A. Todd, “An experimental investigation of the interactive effects of interface style, instructions, and task familiarity on user performance,” *ACM Transactions on Computer-Human Interaction*, vol. 3, no. 1, pp. 1–37, 1996.

[18] D. Keim, J. Kohlhammer, and G. Ellis, *Mastering The Information Age-Solving Problems with Visual Analytics*. Goslar: Eurographics Association, 2010, p. 163.

[19] K. Salen and E. Zimmerman, *Rules of play: Game design fundamentals*. MIT Press, 2004, p. 650.

[20] “emergent, adj. and n.: Oxford English Dictionary.” [Online]. Available: http://www.oed.com/view/Entry/61131?redirectedFrom=emergent#eid. [Accessed: 18-Aug-2012].

[21] A. Dillon and C. Watson, “User analysis in HCI: the historical lesson from individual differences research,” *International Journal of Human-Computer Studies*, vol. 45, no. 6, pp. 619–637, 1996.

[22] A. Paivio, *Mental representations: A dual coding approach*. OUP USA, 1990, p. 336.

[23] J. B. Rotter, “Generalized expectancies for internal versus external control of reinforcement,” *Psychological Monographs: General & Applied*, 1966.

[24] J. B. Rotter, “Internal versus external control of reinforcement: A case history of a variable,” *American Psychologist*, vol. 45, no. 4, pp. 489–493, 1990.

[25] A. Bandura, “Self-efficacy,” *Encyclopedia of Human Behavior*, vol. 4, pp. 71–81, 1994.

[26] H. G. Gough, “A leadership index on the California Psychological Inventory,” *Journal of Counseling Psychology*, vol. 16, no. 4, pp. 283–289, 1969.

[27] N. D. Fleming, *VARK teaching and learning styles*. Neil D Fleming, 2012, p. 91.

[28] J. Lehrer, “The eureka hunt,” *New Yorker*, New York, NY, p. 6, 28-Jul-2008.

[29] P. Saraiya, C. North, and K. A. Duca, "An insight-based methodology for evaluating bioinformatics visualizations," *IEEE Transaction on Visualization and Computer Graphics*, vol. 11, no. 4, pp. 443–56, 2005.

[30] J. Rieman, "A field study of exploratory learning strategies," *ACM Transactions on Computer-Human Interaction*, vol. 3, no. 3, pp. 189–218, 1996.

[31] P. Saraiya, C. North, and K. A. Duca, "An evaluation of microarray visualization tools for biological insight," in *IEEE Symposium on Information Visualization*, 2004, pp. 1–8.

[32] P. Saraiya, C. North, V. Lam, and K. A. Duca, "An insight-based longitudinal study of visual analytics," *IEEE Transaction on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1511–22, 2006.

[33] C. North, "Toward measuring visualization insight," *IEEE Computer Graphics and Applications*, vol. 26, no. 3, pp. 6–9, 2006.

[34] G. Klein and A. Jarosz, "A naturalistic study of insight," *Journal of Cognitive Engineering and Decision Making*, vol. 5, no. 4, pp. 335–351, 2011.

[35] G. Klein, J. Phillips, E. Rall, and D. A. Peluso, "A data-frame theory of sensemaking," in *Expertise out of context*, R. R. Hoffman, Ed. Taylor & Francis, 2007, pp. 113–155.

[36] R. Sternberg and J. Davidson, *The nature of insight.* MIT Press, 1995, p. 636.

[37] J. W. Schooler and J. Melcher, "The ineffability of insight.," in *The creative cognition approach.*, S. M. Smith, T. B. Ward, and R. A. Finke, Eds. Cambridge, MA, US: The MIT Press, 1995, pp. 97–133.

[38] R. Weisberg, "Prolegomena to theories of insight in problem solving: A taxonomy of problems," in *The nature of insight*, R. Sternberg and J. Davidson, Eds. MIT Press, 1995.

[39] J. Metcalfe and D. Wiebe, "Intuition in insight and noninsight problem solving," *Memory & Cognition*, vol. 15, no. 3, pp. 238–46, May 1987.

[40] E. M. Bowden, M. Jung-Beeman, J. I. Fleck, and J. Kounios, "New approaches to demystifying insight," *Trends in Cognitive Sciences*, vol. 9, no. 7, pp. 322–8, 2005.

[41] N. Riche, "Beyond system logging: human logging for evaluating information visualization," in *BELIV'10 workshop at ACM SIGCHI*, 2010.

[42] P. Isenberg, A. Tang, and S. Carpendale, "An exploratory study of visual information analysis," in *26th CHI Conference*, 2008, pp. 1217–1226.

[43] V. González and A. Kobsa, "A workplace study of the adoption of information visualization systems," in *I-KNOW'03: 3rd International Conference on Knowledge Management*, 2003, pp. 92–102.

[44] J. Seo and B. Shneiderman, “Knowledge discovery in high-dimensional data: case studies and a user survey for the rank-by-feature framework,” *IEEE Transaction on Visualization and Computer Graphics*, vol. 12, no. 3, pp. 311–22, 2006.

[45] J. Gregory Trafton, S. S. Kirschenbaum, T. L. Tsui, R. T. Miyamoto, J. a. Ballas, and P. D. Raymond, “Turning pictures into numbers: extracting and generating information from complex visualizations,” *International Journal of Human-Computer Studies*, vol. 53, no. 5, pp. 827–850, 2000.

[46] B. Shneiderman and C. Plaisant, “Strategies for evaluating information visualization tools,” in *BELIV ’06, AVI workshop on Beyond Time and Errors Novel Evaluation Methods for Information Visualization*, 2006, pp. 1–7.

[47] W. Willett, J. Heer, and M. Agrawala, “Scented widgets: improving navigation cues with embedded visualizations,” *IEEE Transaction on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1129–36, 2007.

[48] C. North, P. Saraiya, and K. Duca, “A comparison of benchmark task and insight evaluation methods for information visualization,” *Information Visualization*, vol. 10, no. 3, pp. 162–181, Aug. 2011.

[49] P. Saraiya, P. Lee, and C. North, “Visualization of graphs with associated timeseries data,” in *IEEE Symposium on Information Visualization*, 2005, pp. 225–232.

[50] D. Gotz and M. X. Zhou, “Characterizing users’ visual analytic activity for insight provenance,” *Information Visualization*, vol. 8, no. 1, pp. 42–55, Jan. 2009.

[51] R. Amar, J. Eagan, and J. Stasko, “Low-level components of analytic activity in information visualization,” in *IEEE Symposium on Information Visualization*, 2005, pp. 111–117.

[52] S. Wehrend and C. Lewis, “A problem-oriented classification of visualization techniques,” in *Proceedings of the First IEEE Conference on Visualization: Visualization `90*, 1990, pp. 139–143,.

[53] S. F. Roth and J. Mattis, “Data characterization for intelligent graphics presentation,” in *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI ’90*, 1990, pp. 193–200.

[54] A. Leontjev, *Activity. Consciousness. Personality*. Englewood Cliffs, NJ: Prentice-Hall, 1978, p. 62.

[55] J. S. Yi, Y.-A. Kang, J. T. Stasko, and J. A. Jacko, “Understanding and characterizing insights,” in *Proceedings of the BELIV ’08 conference on Beyond Time and Errors Novel Evaluation Methods for Information Visualization*, 2008, pp. 1–6.

[56] P. Pirolli and S. K. Card, “The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis,” in *Proceedings of International Conference on Intelligence Analysis*, 2005, vol. 2005, pp. 2–4.

[57] H. Lam, "A framework of interaction costs in information visualization," *IEEE Transaction on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1149–56, 2008.

[58] N. Elmqvist, a. V. Moere, H.-C. Jetter, D. Cernea, H. Reiterer, and T. Jankun-Kelly, "Fluid interaction for information visualization," *Information Visualization*, vol. 10, no. 4, pp. 327–340, Aug. 2011.

[59] J. Stromer-Galley, "Interactivity-as-Product and Interactivity-as-Process," *The Information Society*, vol. 20, no. 5, pp. 391–394, Nov. 2004.

[60] A. Dix and G. Ellis, "Starting simple," in *Proceedings of the Working Conference on Advanced Visual Interfaces - AVI '98*, 1998, p. 124.

[61] Z. Liu, N. Nersessian, and J. Stasko, "Distributed cognition as a theoretical framework for information visualization," *IEEE Transaction on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1173–80, 2008.

[62] J. S. Yi, Y.-A. Kang, J. T. Stasko, and J. A. Jacko, "Toward a deeper understanding of the role of interaction in information visualization," *IEEE Transaction on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1224–31, 2007.

[63] A. D. Norman, *The design of everyday things*. New York, NY: Basic Books, 2002, p. 288.

[64] Z. Liu and J. T. Stasko, "Mental models, visual reasoning and interaction in information visualization: a top-down perspective," *IEEE Transaction on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 999–1008, 2010.

[65] T. Susi, M. Johannesson, and P. Backlund, "Serious Games – An Overview," Unpublished report, University of Skövde, Sweden, Skövde, 2007.

[66] J. Bown, A. Szymkowiak, K. Fee, A. Sampson, M. Shovman, R. Falconer, A. Goltsov, J. Issacs, P. Robertson, and K. Scott-Brown, "Information visualization and the arts-science-social science interface," in *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia - IITM '10*, 2010, pp. 9–17.

[67] N. Diakopoulos, F. Kivran-Swaine, and M. Naaman, "Playable data," in *CHI '11 Annual Conference on Human Factors in Computing Systems*, 2011, pp. 1717–1726.

[68] M. Tavanti and M. Lind, "2D vs 3D, implications on spatial memory," in *IEEE Symposium on Information Visualization. INFOVIS*, 2001, vol. 2001, pp. 139–145.

[69] M. St. John, M. B. Cowen, H. S. Smallman, and H. M. Oonk, "The Use of 2D and 3D Displays for Shape-Understanding versus Relative-Position Tasks," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 43, no. 1, pp. 79–98, Jan. 2001.

[70] C. Conati and H. Maclaren, "Exploring the role of individual differences in information visualization," in *Proceedings of the Working Conference on Advanced Visual Interfaces - AVI '08*, 2008, pp. 199–206.

[71] B. Allen, "Individual differences and the conundrums of user-centered design: Two experiments," *Journal of the American Society for Information Science*, vol. 51, no. 6, pp. 508–520, 2000.

[72] C. Ziemkiewicz and R. Kosara, "Preconceptions and individual differences in understanding visual metaphors," *Computer Graphics Forum*, vol. 28, no. 3, pp. 911–918, 2009.

[73] M. B. Donnellan, F. L. Oswald, B. M. Baird, and R. E. Lucas, "The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality," *Psychological Assessment*, vol. 18, no. 2, pp. 192–203, 2006.

[74] S. Cassidy and P. Eachus, "Learning Style, Academic Belief Systems, Self-report Student Proficiency and Academic Achievement in Higher Education," *Educational Psychology*, vol. 20, no. 3, pp. 307–322, 2000.

[75] E. R. Tufte, *The visual display of quantitative information.*, 2nd ed. Cheshire, CT: Graphics Press, 2001, p. 197.

[76] C. Chen, "Individual differences in a spatial-semantic virtual environment," *Journal of the American Society for Information Science*, vol. 51, no. 6, pp. 529–542, 2000.

[77] C. Chen, S. Toh, and W. M. F. W. Ismail, "Are Learning Styles Relevant to Virtual Reality?," *Journal of Research on Technology in Education*, vol. 38, no. 2, pp. 123–141, 2005.

[78] H. Hauptman and A. Cohen, "The synergetic effect of learning styles on the interaction between virtual environments and the enhancement of spatial thinking," *Computers & Education*, vol. 57, no. 3, pp. 2106–2117, Nov. 2011.

[79] N. D. Fleming and C. Mills, "Not another inventory, rather a catalyst for reflection," *To Improve the Academy*, vol. 11, no. 1, p. 137, 1992.

[80] L. M. Miller, "Using learning styles to evaluate computer-based instruction," *Computers in Human Behavior*, vol. 21, no. 2, pp. 287–306, Mar. 2005.

[81] M. J. Findley and H. M. Cooper, "Locus of control and academic achievement: A literature review," *Journal of Personality and Social Psychology*, vol. 44, no. 2, pp. 419–427, 1983.

[82] P. Naithani, "Guidelines for developing a robust web survey," *Advances in Information Technology and Management*, vol. 1, no. 1, pp. 20–23, 2012.

[83] H. G. Gough, "A work orientation scale for the California Psychological Inventory.," *Journal of Applied Psychology*, vol. 70, no. 3, pp. 505–513, 1985.

[84] C. Tomlinson-Keasey and C. Smith-Winberry, “Educational strategies and personality outcomes of gifted and nongifted college students,” *Gifted Child Quarterly*, vol. 27, no. 1, pp. 35–41, 1983.

[85] L. R. Goldberg, J. a. Johnson, H. W. Eber, R. Hogan, M. C. Ashton, C. R. Cloninger, and H. G. Gough, “The international personality item pool and the future of public-domain personality measures,” *Journal of Research in Personality*, vol. 40, no. 1, pp. 84–96, 2006.

[86] L. J. Cronbach, “Coefficient alpha and the internal structure of tests,” *Psychometrika*, vol. 16, no. 3, pp. 297–334, Sep. 1951.

[87] D. Lemire, *Technical data for five learning style instruments with instructional applications.* Reno, NV: Creative Therapeutics Publishers., 2001.

[88] D. Kolb, *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall, 1984, p. 288.

[89] A. F. Gregorc and H. B. Ward, “A new definition for individual,” *NASSP Bulletin*, vol. 61, no. 406, pp. 20–26, Feb. 1977.

[90] R. M. Felder and L. K. Silverman, “Learning styles and teaching styles in engineering education,” *Engineering Education*, vol. 78, no. 7, pp. 674–681, 1988.

[91] N. D. Fleming, *Teaching and learning styles: VARK strategies.* Christchurch, New Zealand: Neil D Fleming, 2001, p. 128.

[92] R. Dunn, K. Dunn, and G. Price, *Manual/or the learning styles inventory (LSI)*. Laurence, KS: Price Systems, 1987.

[93] J. Eicher, *Making the message clear: how to master the business communication tools that direct productivity, excellence and power.* Scotts Valley, CA: Grinder, Delozier & Associates, 1993, p. 124.

[94] L. J. Cronbach and R. E. Snow, “Individual Differences in Learning Ability as a Function of Instructional Variables.,” Stanford University, CA, USA, 1969.

[95] R. E. Snow, “Individual differences and the design of educational programs,” *American Psychologist*, vol. 41, no. 10, pp. 1029–1039, 1986.

[96] D. H. Jonassen and B. L. Grabowski, “Handbook of individual differences, learning, and instruction,” *Journal of School Psychology*, vol. 33, no. 3, pp. 265–268, Sep. 1993.

[97] C. Schlick, *Industrial Engineering and Ergonomics*. Berlin: Springer, 2009, p. 710.

[98] G. Reid and F. Eggemeier, “Subjective workload assessment technique,” Air Force Flight Test Center Edwards AFB CA, 1982.

[99] S. Hart, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” NASA, 1988.

[100] S. Rubio, E. Díaz, J. Martín, and J. Puente, “Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods,” *Applied Psychology*, vol. 53, no. 1, pp. 61–86, 2004.

[101] R. Arias-Hernández, J. Dill, B. Fisher, and T. M. Green, “Visual analytics and human-computer interaction,” *Interactions*, vol. 18, no. 1, pp. 51–55, 2011.

[102] Tableau Software, “Tableau Reader.” [Online]. Available: http://www.tableausoftware.com/products/reader. [Accessed: 25-May-2011].

[103] L. R. Peterson and M. J. Peterson, “Short-term retention of individual verbal items,” *Journal of Experimental Psychology*, vol. 58, no. 3, 1959.

[104] “Portal Game - Valve Corporation.” [Online]. Available: http://www.valvesoftware.com/games/portal.html. [Accessed: 20-Sep-2012].

[105] H. Coolican, *Research Methods and Statistics in Psychology.*, 5th ed. London, UK: Hodder Education, 2009, p. 703.

[106] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Analysis.*, 4th Editio. Boston, MA: Allyn and Bacon, 2001.

[107] G. V Glass, P. D. Peckham, and J. R. Sanders, “Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance,” *Review of Educational Research*, vol. 42, no. 3, pp. 237–288, 1972.

[108] N. D. M. Villanueva, A. J. Petenate, and M. A. A. P. Da Silva, “Performance of three affective methods and diagnosis of the ANOVA model,” *Food Quality and Preference*, vol. 11, no. 5, pp. 363–370, 2000.

[109] D. G. Bonett and J. A. Woodward, “Testing residual normality in the ANOVA model,” *Journal of Applied Statistics*, vol. 17, no. 3, pp. 383–387, 1990.

[110] R. Bender and S. Lange, “Adjusting for multiple testing--when and how?,” *Journal of Clinical Epidemiology*, vol. 54, no. 4, pp. 343–349, 2001.

[111] P. C. O’Brien, “The appropriateness of analysis of variance and multiple-comparison procedures,” *Biometrics*, vol. 39, no. 3, pp. 787–94, 1983.

[112] Cohen, *Statistical power analysis for the behavioral sciences.*, 2nd ed. Routledge Academic, 1988, p. 590.

[113] J. Cohen, “A power primer,” *Psychological Bulletin*, vol. 112, no. 1, pp. 155–159, 1992.

[114] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, “Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses,” *Behavior Research Methods*, vol. 41, no. 4, pp. 1149–60, 2009.

[115] J. Nunnally and I. Bernstein, *Psychometric Theory.*, 3rd ed. New York, NY: McGraw-Hill, 1967, p. 736.

[116] N. D. Fleming, “VARK: a guide to learning styles. scoring instruction.” [Online]. Available: http://www.vark-learn.com/documents/scoring.pdf. [Accessed: 23-May-2012].

[117] N. D. Fleming, “VARK: a guide to learning styles. research & statistics.” [Online]. Available: http://www.vark-learn.com/english/page.asp?p=research. [Accessed: 12-May-2012].

[118] R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky, “Defining insight for visual analytics,” *IEEE Computer Graphics and Applications*, vol. 29, no. 2, pp. 14–17, 2009.

[119] D. Lieberman, “The computer’s potential role in health education,” *Health Communication*, vol. 4, no. 3, pp. 211–225, 1992.

[120] B. J. Fogg, G. Cuellar, and D. Danielson, “Motivating, influencing, and persuading users: an introduction to captology,” in *The Human-Computer Interaction Handbook*, Taylor & Francis, 2008, pp. 133–146.

[121] A. Bandura, *Self-efficacy: The exercise of control*. New York: Freeman, 1997.

[122] R. De Lisi and J. Wolford, “Improving children’s mental rotation accuracy with computer game playing,” *The Journal of Genetic Psychology*, vol. 163, no. 3, pp. 272–282, 2002.

[123] J. I. Navarro, E. Marchena, C. Alcalde, G. Ruiz, I. Llorens, and M. Aguilar, “Improving attention behaviour in primary and secondary school children with a Computer Assisted Instruction procedure,” *International Journal of Psychology*, vol. 38, no. 6, pp. 359–365, Dec. 2003.

[124] B. K. Squire and C. Steinkuehler, “Meet the gamers,” *Library Journal*, 2005. [Online]. Available: http://website.education.wisc.edu/kdsquire/tenure-files/27-library-meet-gamers.pdf. [Accessed: 14-Aug-2012].

[125] B. Shneiderman, “Why not make interfaces better than 3D reality?,” *IEEE Computer Graphics and Applications*, vol. 23, no. 6, pp. 12–15, Nov. 2003.

[126] S. Baumgärtner, A. Ebert, M. Deller, and S. Agne, “2D meets 3D: a human-centered interface for visual data exploration,” in *CHI ’07 extended abstracts on Human factors in computing systems - CHI '07*, 2007, pp. 2273–2278.

[127] A. Bandura, “Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems,” *Journal of Personality and Social Psychology*, vol. 45, no. 5, pp. 1017–1028, 1983.

[128] R. Hannafin, “Effects of spatial ability and instructional program on geometry achievement,” *The Journal of Educational Research*, vol. 101, no. 3, pp. 148–156, 2008.

[129] T. Huk, “Who benefits from learning with 3D models? The case of spatial ability,” *Journal of Computer Assisted Learning*, vol. 22, no. 6, pp. 392–404, 2006.

[130] B. U. Ramirez, “The sensory modality used for learning affects grades,” *Advances in Physiology Education*, vol. 35, no. 3, pp. 270–4, 2011.

[131] N. D. Fleming, “VARK - A guide to learning styles.” [Online]. Available: http://www.vark-learn.com/english/index.asp. [Accessed: 27-Aug-2012].

[132] N. Savage, “Gaining wisdom from crowds,” *Communications of the ACM*, vol. 55, no. 3, pp. 13–15, Mar. 2012.

[133] S. Deterding, “Gamification: designing for motivation,” *Interactions*, vol. 19, no. 4, pp. 14–17, 2012.

Appendices

Appendix 1 Consent Form 181
Appendix 2 Participant Information Sheet 182
Appendix 3 Count Down Sheet 183
Appendix 4 Tutorial Insight Grid 184
Appendix 5 Empty Insight Grid.. 185
Appendix 6 Pre-Study Questionnaire.. 186
Appendix 7 Locus of Control Section of Pre-Study Questionnaire 187
Appendix 8 VARK Section of Pre-Study Questionnaire 188
Appendix 9 Self-Efficacy and Self-Acceptance Section of Pre-Study Questionnaire192
Appendix 10 VARK Scoring 193
Appendix 11 NASA-TLX.. 195
Appendix 12 VTI – 2D Representation Post Experiment Questionnaire............... 197
Appendix 13 VTI – 3D Representation Post Experiment Questionnaire............... 203

[Intentionally left blank]

Appendix 1 Consent Form

Loughborough University

Insert Name of Research Proposal

INFORMED CONSENT FORM
(to be completed after Participant Information Sheet has been read)

The purpose and details of this study have been explained to me. I understand that this study is designed to further scientific knowledge and that all procedures have been approved by the Loughborough University Ethical Advisory Committee.

I have read and understood the information sheet and this consent form.

I have had an opportunity to ask questions about my participation.

I understand that I am under no obligation to take part in the study.

I understand that I have the right to withdraw from this study at any stage for any reason, and that I will not be required to explain my reasons for withdrawing.

I understand that all the information I provide will be treated in strict confidence and will be kept anonymous and confidential to the researchers unless (under the statutory obligations of the agencies which the researchers are working with), it is judged that confidentiality will have to be breached for the safety of the participant or others.

I agree to participate in this study.

Your name ________________________________

Your signature ________________________________

Signature of investigator ________________________________

Date ________________________________

Appendix 2 Participant Information Sheet

Loughborough University

Understanding the Value of interaction in insight and Knowledge Generation

Participant Information Sheet

Professor Roy S. Kalawsky, r.s.kalawsky@lboro.ac.uk 01509 635678
Petri Vitiello, p.f.vitiello@lboro.ac.uk 01509 635673
Loughborough University
Leicestershire, UK
LE11 3TU

What is the purpose of the study?

The objective of these experiments is to understand the value of interaction in terms of cognition in particular insight and knowledge building.

Who is doing this research and why?

This study is part of a PhD student research project supported by Loughborough University and EPSRC which aims to understand and quantify the value of interaction in the process on insight & knowledge generation in the visual analytics domain.

Once I take part, can I change my mind?

Yes! After you have read this information and asked any questions you may have we will ask you to complete an Informed Consent Form, however if at any time, before, during or after the sessions you wish to withdraw from the study please just contact the main investigator. You can withdraw at any time, for any reason and you will not be asked to explain your reasons for withdrawing.

How long will it take?

The experiment is estimated to take approximately 90 minutes with a 10 min break in the middle. This includes the expected time to complete the different task and questionnaires.

What will I be asked to do?

There are 3 main experiments; one involves a 3D video first person puzzle game, then a 2D video puzzle game and finally a data visualisation analysis. At the end of each experiment there will be a short questionnaires.

Will my taking part in this study be kept confidential?

The information you provide will be held and used in accordance with the Data Protection Act 1998. All information that is collected about you during the course of the research will be kept strictly confidential and stored securely at Loughborough University. An ID number will identify you and any information about you will have your name and address removed so that you cannot be recognised from it. All video and data recordings will be destroyed six years after the completion of this investigation.

I have some more questions who should I contact?

Petri Vitiello, p.f.vitiello@lboro.ac.uk 01509 635673

What if I am not happy with how the research was conducted?

The University has a policy relating to Research Misconduct and Whistle Blowing which is available online at http://www.lboro.ac.uk/admin/committees/ethical/Whistleblowing(2).htm.

Updated Jan 2011

Appendix 3 Count Down Sheet

COUNT DOWN IN
-3
INCREMENTS

99	
60	
30	
0	

Appendix 4 Tutorial Insight Grid

This table describes some of the insights gained about the data set regarding condos & houses in terms of quantity and ranges, location (North, Mid, South features & clusters), size ranges in square foot (sqft), number of bedrooms & bathrooms, temporal patterns (yearly, monthly, weekly, day of the week), price and price per square foot ($/sqft).

Please analyse the tutorial data set and investigate the findings in the table below where the intersection signifies that the insight refers to the 2 key aspects of the data, although it can contain additional aspects as well.
For examples:

- Grid location – Quantity – Ranges / South: in the south there is 85 properties sold out of 277 which is less than ~30%
- Grid location – North/ Price/Sq Foot: in the north there is no property above $120/sqft

	Quantity - Ranges	North	Mid	South	Size square feet	Beds	Temporal patterns	Price	Price/Sq Foot
Quantity - Ranges	Total of 277 Properties sold	65 sales only 1 condo	127 sales has more than ½ the condos of the dataset.	85 Sales	Range is 640-4,494 The majority (273 out of 277) are less than 2,350 sqf	The total range 0-5 bedrooms The majority in the dataset (270 out of 277) ranges 2-4 bedrooms There is 5, 0 bedroom houses 3 bedroom houses are the most popular (total 190)	Total Weekly Median sale is 14 and the regional median is 4 January ahs the highest proportion of 2 bedroom sales.	Price range $2,500 – $495,000 per property	Range $1-$533 per sqf 271 are less than $120 / sqf
		North	Ratio of 2:1 sales Median sq. feet is 370 sqft lager than mid.	South has 20 more sales of which 6 are condos	Median value 1,576 Max value 4,248 Min Value 862	Range 2-4	In April only 1 sale (week 15)	Wednesday sales have the highest Median Price $129,500 North has the highest median price $97,000 about double of the other regions.	No property above $120/sqf
			Mid	The proportion of House/ condos is the same despite the different sales quantity.	Median value 1,200 Max value 2,216 Min Value 604	Only region with 5 bedroom house	Thursdays has the highest sales by number in themed region (33)	April has the lowest monthly average sales figure $113,300 a third of the other months.	Outlier of $533/sqf property Excluding the outlier the properties above $120/sqf are clustered s on the south east
				South	Median value 1,216 Max value 4,494 Min Value 648	All the 0 bedroom houses are in the south.	Weekly Median sale is 4	The south has the 2nd most expensive house in the dataset $455,500.	More than $120/sqf are only in the south The most expensive house is inline with the market with $101 / sqft
					Size square feet	0 bedroom houses are relatively large Median sqft 1,450	The largest house was sold in march	Over 90% of properties are less than $150k and range from 604-2254 sqf	The largest house in the dataset in $101/sqft
						Beds	3 bedroom March and Feb are	The 5 bedroom house is $64,000	3 bedroom range form $1-$222 / sqf
							Temporal patterns	Most expensive house was sold in January.	For the $50/sqft bin ($40-50/sqft) Wednesday is not as popular (3 sales) compared to a 10 sales median value overall.
								Price	The most expensive houses diverge completely on $/sqft The 1st $495k is $533/sqft whereby the 2nd at 455.5k is only $101/sqft.

Appendix 5 Empty Insight Grid

What can you say about condos, houses and townhouses on the Island of Mercer in terms of quantity and ranges, location (North, Mid, South features & clusters), size ranges in square feet (sqft), number of bedrooms & bathrooms, temporal patterns (yearly, monthly, weekly, day of the week), price and price per square foot ($/sqft).

Please capture your findings in the corresponding grid location in the table below where the intersection signifies that the insight refers to the 2 key aspects of the data, although it can contain additional aspects as well.
For examples:

- Grid location – Quantity – Ranges / South: in the south there is 74 properties sold out of 750 which is less than 10%.
- Grid location - North/Beds: 5 bedrooms in the north are more than $100k

	Quantity - Ranges	North	Mid	South	Size square feet	Beds	Baths	Temporal patterns	Price	Price/Sq Foot
Quantity - Ranges										
		North								
			Mid							
				South						
					Size square feet					
						Beds				
							Baths			
								Temporal patterns		
									Price	

Appendix 6 Pre-Study Questionnaire

Pre-Study Questionnaire

* Required

Name: *

Email address: *

Gender *

- Male
- Female

Age *

Your age in numbers (e.g. 23)

University Department *

Aeronautical and Automotive Engineering

if other, please specify

Status *

- Undergraduate
- Postgraduate
- Staff
- Other

3D first person gaming experience *

First-person shooter video game (e.g. Call of Duty, Quake)

- None
- Some
- A lot

Non-3D Gaming experience *

2D video game (e.g. Simcity, Tetris, Solitaire)

- None
- Some
- A lot

Appendix 7 Locus of Control

Section of Pre-Study Questionnaire

9/24/12 Pre-Study Questionnaire

Pre-Study Questionnaire

* Required

Locus of Control - Questionnaire

Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement whether it is:

1. Very Inaccurate,
2. Moderately Inaccurate,
3. Neither Accurate Nor Inaccurate,
4. Moderately Accurate, or
5. Very Accurate as a description of you.

Locus of Control *

	Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate
Feel comfortable with myself.	○	○	○	○	○
Believe that unfortunate events occur because of bad luck.	○	○	○	○	○
Just know that I will be a success.	○	○	○	○	○
Feel that my life lacks direction.	○	○	○	○	○
Come up with good solutions.	○	○	○	○	○
Love life.	○	○	○	○	○
See difficulties everywhere.	○	○	○	○	○
Believe that my success depends on ability rather than luck.	○	○	○	○	○
Habitually blow my chances.	○	○	○	○	○
Act comfortably with others.	○	○	○	○	○
Feel up to any task.	○	○	○	○	○
Dislike taking responsibility for making decisions.	○	○	○	○	○
Like to take responsibility for making decisions.	○	○	○	○	○
Dislike myself.	○	○	○	○	○
Believe some people are born lucky.	○	○	○	○	○
Am less capable than most people	○	○	○	○	○
Make a decision and move on.	○	○	○	○	○
Believe that the world is controlled by few powerful people	○	○	○	○	○
Feel that I'm unable to deal with things	○	○	○	○	○
Take the initiative	○	○	○	○	○

Appendix 8 VARK Section of Pre-Study Questionnaire

Pre-Study Questionnaire

*Required

Learning Style - Questionnaire

Choose the answer which best explains your preference.
Please choose more than one if a single answer does not match your perception.
Leave blank any question that does not apply

You are helping someone with directions. You would: *

Check all that apply OR N/A if not applicable.

- [ ] Go with the person.
- [ ] Verbally give directions.
- [ ] Write down the directions.
- [ ] Draw, or give a map.
- [ ] N/A

You are not sure whether a word should be spelled `dependent' or `dependant'. You would: *

Check all that apply OR N/A if not applicable.

- [ ] See the words in your mind and choose by the way they look.
- [ ] Think about how each word sounds and choose one.
- [ ] Find it in a dictionary.
- [ ] Write both words on paper and choose one.
- [ ] N/A

You are planning a holiday for a group. You want some feedback from them about the plan. You would: *

Check all that apply OR N/A if not applicable.

- [ ] Describe some of the highlights.
- [ ] Use a map or website to show them the places.
- [ ] Give them a copy of the printed itinerary.
- [ ] Phone, text or email them.
- [ ] N/A

You are going to cook something as a special treat for your family. You would: *

Check all that apply OR N/A if not applicable.

- ☐ Cook something you know without the need for instructions.
- ☐ Ask friends for suggestions.
- ☐ Look through the cookbook for ideas from the pictures.
- ☐ Use a cookbook where you know there is a good recipe.
- ☐ N/A

A group of tourists want to learn about attractions in your area. You would: *

Check all that apply OR N/A if not applicable.

- ☐ Talk about, or arrange a talk for them about tourist attractions in your area.
- ☐ Show them internet pictures, photographs or picture books.
- ☐ Take them to a tourist attraction and walk with them.
- ☐ Give them a book or pamphlets about tourist attractions in your area.
- ☐ N/A

You are about to purchase a digital camera or mobile phone. Other than price, what would most influence your decision? *

Check all that apply OR N/A if not applicable.

- ☐ Trying or testing it.
- ☐ Reading the details about its features.
- ☐ It is a modern design and looks good.
- ☐ The salesperson telling me about its features.
- ☐ N/A

Remember a time when you learned how to do something new. Try to avoid choosing a physical skill, eg. riding a bike. You learned best by: *

Check all that apply OR N/A if not applicable.

- ☐ Watching a demonstration.
- ☐ Listening to somebody explaining it and asking questions.
- ☐ Diagrams and charts - visual clues.
- ☐ Written instructions – e.g. a manual or textbook.
- ☐ N/A

You have a problem physical health issue. You would prefer that the doctor: *

Check all that apply OR N/A if not applicable.

- ☐ Gave you a something to read to explain what was wrong.
- ☐ Used a plastic model to show what was wrong.
- ☐ Described what was wrong.
- ☐ Showed you a diagram of what was wrong.
- ☐ N/A

You want to learn a new program, skill or game on a computer. You would: *

Check all that apply OR N/A if not applicable.

- ☐ Read the written instructions that came with the program.
- ☐ Talk with people who know about the program.
- ☐ Use the controls or keyboard.
- ☐ Follow the diagrams in the book that came with it.
- ☐ N/A

I like websites that have: *

Check all that apply OR N/A if not applicable.

- ☐ Things I can click on, shift or try.
- ☐ Interesting design and visual features.
- ☐ Interesting written descriptions, lists and explanations.
- ☐ Audio channels where I can hear music, radio programs or interviews.
- ☐ N/A

Other than price, what would most influence your decision to buy a new non-fiction book? *

Check all that apply OR N/A if not applicable.

- ☐ The way it looks is appealing.
- ☐ Quickly reading parts of it.
- ☐ A friend talks about it and recommends it.
- ☐ It has real-life stories, experiences and examples.
- ☐ N/A

You are using a book, CD or website to learn how to take photos with your new digital camera. You would like to have: *

Check all that apply OR N/A if not applicable.

- ☐ A chance to ask questions and talk about the camera and its features.
- ☐ Clear written instructions with lists and bullet points about what to do.
- ☐ Diagrams showing the camera and what each part does.
- ☐ Many examples of good and poor photos and how to improve them.
- ☐ N/A

Do you prefer a teacher or a presenter who uses: *

Check all that apply OR N/A if not applicable.

- [ ] Demonstrations, models or practical sessions.
- [ ] Question and answer, talk, group discussion, or guest speakers.
- [ ] Handouts, books, or readings.
- [ ] Diagrams, charts or graphs.
- [ ] N/A

You have finished a competition or test and would like some feedback. You would like to have feedback: *

Check all that apply OR N/A if not applicable.

- [ ] Using examples from what you have done.
- [ ] Using a written description of your results.
- [ ] From somebody who talks it through with you.
- [ ] Using graphs showing what you had achieved.
- [ ] N/A

You are going to choose food at a restaurant or cafe. You would: *

Check all that apply OR N/A if not applicable.

- [ ] Choose something that you have had there before.
- [ ] Listen to the waiter or ask friends to recommend choices.
- [ ] Choose from the descriptions in the menu.
- [ ] Look at what others are eating or look at pictures of each dish.
- [ ] N/A

You have to make an important speech at a conference or special occasion. You would: *

Check all that apply OR N/A if not applicable.

- [ ] Make diagrams or get graphs to help explain things.
- [ ] Write a few key words and practice saying your speech over and over.
- [ ] Write out your speech and learn from reading it over several times.
- [ ] Gather many examples and stories to make the talk real and practical.
- [ ] N/A

Appendix 9 Self-Efficacy and Self-Acceptance Section of Pre-Study Questionnaire

Pre-Study Questionnaire

* Required

Self-Efficacy / Insight - Questionnaire

Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement whether it is:

1. Very Inaccurate,
2. Moderately Inaccurate,
3. Neither Accurate Nor Inaccurate,
4. Moderately Accurate, or
5. Very Accurate as a description of you.

Self-Efficacy / Insight *

	Very Inaccurate	Moderately Inaccurate	Neither Accurate Nor Inaccurate	Moderately Accurate	Very Accurate
Come up with something new.	○	○	○	○	○
Would describe my experiences as somewhat dull.	○	○	○	○	○
Can handle complex problems.	○	○	○	○	○
Never challenge things.	○	○	○	○	○
Throw a new light on the situation.	○	○	○	○	○
Have little to contribute.	○	○	○	○	○
Think quickly.	○	○	○	○	○
Undertake few things on my own.	○	○	○	○	○
Come up with alternatives.	○	○	○	○	○
Consider myself an average person.	○	○	○	○	○
Formulate ideas clearly.	○	○	○	○	○
Let others determine my choices.	○	○	○	○	○
Put a new perspective on things.	○	○	○	○	○
Say nothing new.	○	○	○	○	○
Have excellent ideas.	○	○	○	○	○
Let myself be directed by others.	○	○	○	○	○
Have a vivid imagination.	○	○	○	○	○
Have difficulty imagining things.	○	○	○	○	○
Am quick to understand things.	○	○	○	○	○
Do not have a good imagination.	○	○	○	○	○

Appendix 10 VARK Scoring

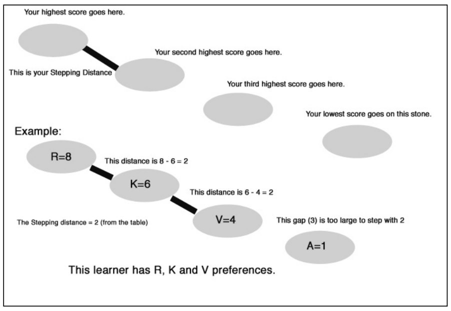
Please fill in these boxes to record your profile of preferences.

	a)	b)	c)	d)
Total for the questionnaire				
	V	**A**	**R**	**K**

You will need the total of your four scores. The total of my four scores is:

SCORING

Because you could choose more than one answer for each question, the scoring is not a simple matter of counting. It is like four stepping-stones across water. Enter your scores **from highest to lowest** on the stones below, with their V, A, R, and K labels.



Your stepping distance comes from this table.

The total of my four VARK scores is -	My stepping distance is
14-21	1
22-27	2
28-32	3
More than 32	4

Follow these steps to establish your preferences.

Step One Your first preference is always your highest score so tick (check) that first stone as one of your preferences.

Step Two Now subtract your second highest score from your highest score. If that figure is larger than your stepping distance go to the paragraph on the next page titled, ***What is the strength of my single preference*?** If not, tick this stone as a second preference and continue with **Step Three** below.

Step Three Subtract your third highest score from your second highest. If that figure is larger than your stepping distance go to the paragraph titled ***Bi-modal Preferences*.** If not, tick this stone as a third preference and continue with **Step Four** below.

Step Four Lastly, subtract your fourth highest score from your third highest. If that number is larger than your stepping distance go to the paragraph headed; ***Tri-Modal Preferences*.** Otherwise, tick your fourth stone as another preference and read the paragraph titled, ***All Four are Preferences!***

Bi-modal Preferences

If you checked two preferences you are bi-modal. You are also part of the large group who are multi-modal – that has more than one preference. Your preferences will be one of the combinations below.

VA VR VK AR AK RK

An example:

Marcelo Total score = 16 Stepping Distance = 1

3	3	5	5
V	A	R	K

Marcelo has a *bi-modal* preference for Read/write and Kinesthetic. Now go to the paragraph titled, ***What is Normal?***

4

Tri-Modal Preferences
If you checked three preferences you are tri-modal. You are also part of the larger group who are multi-modal – that has more than one preference. Your preferences will be one of the combinations below.

VAR VAK VRK ARK

An example:

Adam Total scores = 22 Stepping Distance = 2

8	7	1	6
V	A	R	K

Adam is multimodal with three preferences (V, A and K). His strongest choice (V) is little different from his others (A and K). Now go to the paragraph titled, ***What is Normal?***

All Four are Preferences!
You have checked all four modes (V, A, R and K). They are of similar importance among your preferences for information input and output. You are part of the large group who are *multimodal* – that has more than one preference. Now go to the paragraph titled, ***What is Normal?***

WHAT IS THE STRENGTH OF MY SINGLE PREFERENCE?
This paragraph is for those who have a single preference. Those who have a single preference have their highest score standing out above the others. How much it stands out decides whether it is a *Mild, Strong* or *Very Strong* single preference and the answer depends partly on the total number of responses that you used in the questionnaire. If you have chosen 14 to 21 options in the questionnaire, a score for your highest preference that is six or more ahead of any other score would indicate a *very strong* preference. A difference of only two points between your top two scores would indicate a *mild* preference. If you have chosen 33 or more responses to the 16 questions a *very strong* preference would need to be at least nine (9) ahead of your next highest preference. The table below identifies the strength of your single preference.

Total number of responses?	The difference between my highest score and my next highest score? Ties =0			
Up to 21	6+	4 or 5	2 or 3	0 or 1
22-27	7+	5 or 6	3 or 4	Less than 3
28-32	8+	6 or 7	4 or 5	Less than 4
33+	9+	7 or 8	5 or 6	Less than 5
	Very Strong Preference	**Strong** Preference	**Mild** Preference	**Multimodal.** No single preference

The strength of my single preference is - (check one)

Mild	***Strong***	***Very strong***

Two Examples

Laura Total number of responses = 17 Stepping Distance =1

10	3	2	2
V	A	R	K

Laura's total number of responses (17) can be read in the row of the table above headed "*Up to 21*" and the difference between her highest score (V=10) and her next highest (A=3) is 7. So she has a *Very Strong* Visual preference (V).

Vicki Total number of responses = 27 Stepping Distance =2

5	4	6	12
V	A	R	K

Vicki has a *Strong* Kinesthetic (K) preference because her total score fits the line*"22-27"* in the table and the difference between her two highest scores is 6.

5

Appendix 11 NASA-TLX

NASA - TLX Workload

*Required

What experiment this workload assessment refers to? *

Mercer Island Analysis

How did you perform the experiment? *

Interactively.

Name: *

Mental Demand *

How mentally demanding was the task?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Physical Demand *

How physical demanding was the task?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Temporal Demand *

How hurried or rushed was the pace of the task?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Performance *

How successful were you in accomplishing what you were asked to do?

	1	2	3	4	5	6	7	8	9	10	
Perfect	○	○	○	○	○	○	○	○	○	○	Failure

Effort *

How hard did you have to work to accomplish your level of performance?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Frustration *

How insecure, discouraged, irritated, stressed and annoyed were you?

	1	2	3	4	5	6	7	8	9	10	
Very Low	○	○	○	○	○	○	○	○	○	○	Very High

Rate pairwise the sources of load *
Check ALL true statement that apply

- ☐ Mental Demand is Greater Than Physical Demands.
- ☐ Mental Demand is Greater Than Temporal Demands.
- ☐ Mental Demand is Greater Than Performance.
- ☐ Mental Demand is Greater Than Effort.
- ☐ Mental Demand is Greater Than Frustration Level.
- ☐ Physical Demands is Greater Than Temporal Demands.
- ☐ Physical Demands is Greater Than Performance.
- ☐ Physical Demands is Greater Than Effort.
- ☐ Physical Demands is Greater Than Frustration Level.
- ☐ Temporal Demands is Greater Than Performance.
- ☐ Temporal Demands is Greater Than Effort.
- ☐ Temporal Demands is Greater Than Frustration Level.
- ☐ Performance is Greater Than Effort.
- ☐ Performance is Greater Than Frustration Level.
- ☐ Effort is Greater Than Frustration Level.

Submit

Appendix 12 VTI – 2D Representation

Post Experiment Questionnaire

9/24/12 Portal 2D - Questionnaire

Portal 2D - Questionnaire

*Required

How did you experience Portal 2D? *

Played the game.

Name: *

Have you played Portal Flash version previously? *

- Yes
- No

What properties the blue beam surfaces have? Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Same as any other surface.	◯	◯	◯
Can walk through.	◯	◯	◯
Cannot beam portals through.	◯	◯	◯
Will kill character.	◯	◯	◯

What properties the black surfaces have? Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Same as any other surface.	◯	◯	◯
Are slippery.	◯	◯	◯
Cannot hold portals.	◯	◯	◯

Did you see any cameras? *
Give best guess answers, if not sure.

- Yes
- No

If you saw cameras, what do they do? Answer whether the following statements are true or not.
Give best guess answers, if not sure.

	Yes	No	Don't know
Nothing.	◯	◯	◯
Follow the character	◯	◯	◯
Follow the energy balls	◯	◯	◯

Do the blue and yellow portals have different properties? *
Give best guess answers, if not sure.

- ◯ Yes
- ◯ No

If you answered yes, what are the differences?
Explain briefly in 1-2 sentences.

What happen when you stay between two blue fields when active? Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Nothing.	◯	◯	◯
Character teleports.	◯	◯	◯
Character burns to ashes.	◯	◯	◯

What properties the red beam surfaces have? Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Same as any other surface.	◯	◯	◯
Can walk through.	◯	◯	◯
Cannot beam portals through.	◯	◯	◯
Will kill character.	◯	◯	◯

What do red buttons do? Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Nothing.			
Switch the lights on.			
Activate platforms			
Open gates			

What happens if the character misses the bridging platform on Level 19? Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Nothing.			
Walks onto the blue beam.			
Will kill the character.			

Do you need to jump more than once into a portal to complete a level? *
Give best guess answers, if not sure.

- Yes
- No

If you answered yes, how many times would be the minimum?
Give best guess answers, if not sure.

- 2
- 3
- 4

About energy balls. Answer whether the following statements are true or not. *

Give best guess answers, if not sure.

	Yes	No	Don't know
Can go thought portals.	○	○	○
Can change directions.	○	○	○
Can be switched off.	○	○	○
Create a challenging hazard.	○	○	○
Trigger gate opening.	○	○	○

What happen if an energy ball hits the character? *

Give best guess answers, if not sure.

- ○ Nothing
- ○ Character burns to ashes.
- ○ Energy ball bounces off.
- ○ Don't know.

Do energy balls reset themselves? *

Give best guess answers, if not sure.

- ○ Yes
- ○ No

If you answered yes. What triggers the reset.

Give best guess answers, if not sure.

- ○ Fix number of bounces.
- ○ Fix time.
- ○ Random
- ○ Other: ________

What color is the energy ball target when activated?
Give best guess answers, if not sure.

- ○ Orange
- ○ Green
- ○ Random
- ○ Blue
- ○ Other: ______

About cubes. Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Can go thought portals.	○	○	○
Need to carried by character to go through portals.	○	○	○
Can be held by character whilst creating a portal (with a beam).	○	○	○
Create a challenging hazard.	○	○	○
Can be pushed.	○	○	○
Can be placed on buttons without been carried.	○	○	○
Hold buttons.	○	○	○
Enable character to reach higher.	○	○	○
Enable portals to stay open.	○	○	○
Open gates.	○	○	○
Allow character to speed up.	○	○	○
Protect character from harm.	○	○	○

About solving puzzles. Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Fans have a function.	○	○	○
Need to beam portals through walls.	○	○	○
Need to push more than one button simultaneously.	○	○	○
Need to jump onto moving platform.	○	○	○
Need to climb on cubes.	○	○	○
Need to redirect energy balls.	○	○	○
Need to jump over obstacles.	○	○	○
Need to throw cubes into holes.	○	○	○
Need to cheat the cameras.	○	○	○
Beam moving platform.	○	○	○
Avoid blue surfaces.	○	○	○
Beam through red surface.	○	○	○

In the last level (26), what is the hole in the wall for? *
Give best guess answers, if not sure.

○ Solve the puzzle.
○ Used to beam portal through to exit puzzle impasse.
○ Don't know.
○ Other: ______

Key steps in solving the last puzzle (level 26). Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Put cube onto button to open gate.	○	○	○
Put cube and character simultaneously onto 2 buttons to open gate.	○	○	○
Beam portal once gate open.	○	○	○
Use energy ball into target to open gate.	○	○	○
Accurately aim through holes in the walls.	○	○	○
Beam onto the moving platform.	○	○	○
Avoid red beams.	○	○	○
Beam through blue beam surface.	○	○	○
Change energy ball direction.	○	○	○

Appendix 13 VTI – 3D Representation

Post Experiment Questionnaire

Portal 3D - Questionnaire

*Required

How did you experience Portal 3D? *

Played the game.

Name: *

Have you played Portal from Valve previously? *

- Yes
- No

What properties the blue meshed surfaces have? Answer whether the following statements are true or not. *

Give best guess answers, if not sure.

	Yes	No	Don't know
Same as any other surface.	○	○	○
Can walk through.	○	○	○
Cannot beam portals through.	○	○	○
Will kill character.	○	○	○

What properties the black surfaces have? Answer whether the following statements are true or not. *

Give best guess answers, if not sure.

	Yes	No	Don't know
Same as any other surface.	○	○	○
Cushions the character falls.	○	○	○
Cannot hold portals.	○	○	○

Did you see any cameras? *

Give best guess answers, if not sure.

○ Yes

○ No

If you saw cameras, what do they do? Answer whether the following statements are true or not.

Give best guess answers, if not sure.

	Yes	No	Don't know
Nothing.	○	○	○
Follow the character	○	○	○
Follow the energy balls	○	○	○

Do the blue and yellow portals have different properties? *
Give best guess answers, if not sure.

- ◯ Yes
- ◯ No

If you answered yes, what are the differences?
Explain briefly in 1-2 sentences.

What happen when you stay between two electrified fields when active? Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Nothing.	◯	◯	◯
Character teleports.	◯	◯	◯
Kill the character.	◯	◯	◯

What do the big red buttons do? Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Nothing.	◯	◯	◯
Switch the lights on.	◯	◯	◯
Activate platforms	◯	◯	◯
Open gates	◯	◯	◯

What happen if an energy ball hits the character? *
Give best guess answers, if not sure.

- ◯ Nothing
- ◯ Kills the character.
- ◯ Energy ball bounces off.
- ◯ Don't know.

Do energy balls reset themselves? *
Give best guess answers, if not sure.

- ◯ Yes
- ◯ No

If you answered yes. What triggers the reset.
Give best guess answers, if not sure.

- ◯ Fix number of bounces.
- ◯ Fix time.
- ◯ Random
- ◯ Other: ______

About cubes. Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Can go thought portals.	○	○	○
Need to carried by character to go through portals.	○	○	○
Can be held by character whilst creating a portal (with a beam).	○	○	○
Create a challenging hazard.	○	○	○
Can be pushed.	○	○	○
Can be placed on buttons without been carried.	○	○	○
Hold buttons.	○	○	○
Enable character to reach higher.	○	○	○
Enable portals to stay open.	○	○	○
Open gates.	○	○	○
Allow character to speed up.	○	○	○
Protect character from harm.	○	○	○

About solving puzzles. Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Need to beam portals through small openings in the walls.	○	○	○
Need to push more than one big red button simultaneously.	○	○	○
Need to hold the cube to enable next steps.	○	○	○
Need to climb on cubes.	○	○	○
Need to redirect energy balls.	○	○	○
Need to jump over obstacles.	○	○	○
Need to throw cubes into holes.	○	○	○
Need to cheat the cameras.	○	○	○
Avoid sentry gun fire.	○	○	○
Avoid blue electrified surfaces.	○	○	○
Beam through blue meshed sections.	○	○	○

Key steps in resolving the last level. Answer whether the following statements are true or not. *
Give best guess answers, if not sure.

	Yes	No	Don't know
Use the small red button to recover the cube.	○	○	○
Avoid sentry guns that shoot on sight.	○	○	○
Get to the exit gate, alone.	○	○	○
Knock down the sentry guns.	○	○	○
Avoid falling into the pit.	○	○	○
Jump between some platforms.	○	○	○
Redirect energy ball out of harm.	○	○	○
Throw cube into the pit.	○	○	○

Appendix 14 VMI Statistical Description and Results Tables

		N	M	SD
Total insights	Non-Interactive-Group	21	7.81	4.18
	Interactive-Group	21	12.43	7.00
	Total	42	10.12	6.16
Total accurate insights	Non-Interactive-Group	21	6.26	3.66
	Interactive-Group	21	9.88	6.44
	Total	42	8.07	5.47
Total inaccurate insights	Non-Interactive-Group	21	1.55	1.72
	Interactive-Group	21	2.55	2.69
	Total	42	2.05	2.28
Total inferential Insights	Non-Interactive-Group	14	3.57	1.60
	Interactive-Group	16	4.75	3.09
	Total	30	4.20	2.54
Total procedural Insights	Non-Interactive-Group	20	5.70	3.36
	Interactive-Group	21	8.81	5.74
	Total	41	7.29	4.93
Total accurate inferential insights	Non-Interactive-Group	14	3.18	1.61
	Interactive-Group	16	3.94	2.59
	Total	30	3.58	2.19
Total inaccurate inferential insights	Non-Interactive-Group	14	0.39	0.74
	Interactive-Group	16	0.81	1.11
	Total	30	0.62	0.96
Total accurate procedural insights	Non-Interactive-Group	20	4.35	3.09
	Interactive-Group	21	6.88	5.59
	Total	41	5.65	4.67
Total inaccurate procedural insights	Non-Interactive-Group	20	1.35	1.77
	Interactive-Group	21	1.93	1.91
	Total	41	1.65	1.84

N: Number of participants, **M:** Mean, **SD:** Standard deviation

Table VMI.1 – Number of Insights Statistical Description

	N	F	p
Total insights ‡	42	6.31	0.016*
Total accurate insights	42	5.02	0.031*
Total inaccurate insights	42	2.07	0.159
Total inferential Insights	30	1.65	0.210
Total procedural Insights	41	4.42	0.042*
Total accurate inferential insights	30	0.89	0.353
Total inaccurate inferential insights	30	1.44	0.240
Total accurate procedural insights	41	3.17	0.083
Total inaccurate procedural insights	41	1.01	0.321

‡: Transformed variable, *$p < .05$, ** $p < .01$

N: Number of participants, ***F*:** F ratio, ***p*:** p-value

Table VMI.2 – Omnibus Oneway ANOVA Results

– Number of Insights, Accuracy and Mental Effort

Overall Score	NIG			IG			All		
	N	M	SD	N	M	SD	N	M	SD
ATI-C (LoC)									
Below Mean	11	78.49	23.31	9	77.69	22.62	20	78.13	22.40
Above Mean	10	82.88	11.14	12	73.32	27.54	22	77.67	21.78
ATI-C (SE)									
Below Mean	12	73.20	19.84	11	77.43	16.19	23	75.22	17.90
Above Mean	9	90.42	9.98	10	72.73	33.00	19	81.11	25.90
ATI-C (SA)									
Below Mean	11	73.49	20.78	8	78.09	17.20	19	75.43	18.98
Above Mean	10	88.38	11.41	13	73.41	29.38	23	79.92	24.12
ATI-C (V)									
Below Mean	13	79.80	17.93	9	69.81	31.31	22	75.71	24.13
Above Mean	8	81.85	19.91	12	79.24	19.62	20	80.28	19.25
ATI-C (A)									
Below Mean	8	79.33	21.84	9	87.15	14.28	17	83.47	18.08
Above Mean	13	81.35	16.57	12	66.23	28.05	25	74.09	23.61
ATI-C (R)									
Below Mean	14	75.16	18.67	10	83.10	17.89	24	78.46	18.39
Above Mean	7	91.43	12.15	11	68.01	29.09	18	77.12	26.23
ATI-C (K)									
Below Mean	8	71.75	24.50	9	85.90	15.29	17	79.24	20.80
Above Mean	13	86.02	10.96	12	67.16	28.38	25	76.97	22.84
Overall	21	80.58	18.24	21	75.20	25.04	42	77.89	21.80

N: Number of participants, **M:** Mean, **SD:** Standard deviation
NIG: Non-interactive group, **IG**: interactive group, **LoC:** Locus of control, **SE**: Self-efficacy,
SA: Self-acceptance, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VMI.3 – Overall Score Statistical Description by ATI-C and Interactivity Group

	Overall Score ‡		
	N	F	p
Interactivity Group	42	1.315	0.260
Interactivity Group * ATI-C (LoC)	42	0.49	0.490
Interactivity Group * ATI-C (SE)	42	0.33	0.573
Interactivity Group * ATI-C (SA)	42	0.17	0.680
Interactivity Group * ATI-C (V)	42	0.00	0.999
Interactivity Group * ATI-C (A)	42	0.99	0.328
Interactivity Group * ATI-C (R)	42	5.73	0.023*
Interactivity Group * ATI-C (K)	42	0.00	0.999

*p < .05, ** p < .01
‡: Transformed variable**, N**: Number of participants, ***F*****:** F ratio, ***p*****:** p-value
NIG: Non-interactive group, **IG**: interactive group, **LoC:** Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VMI.4 – Factorial ANOVA Results for the Overall score

Overall Mental Effort yield	NIG			IG			All		
	N	M	SD	N	M	SD	N	M	SD
ATI-C (LoC)									
Below Mean	11	-2.00	4.34	9	-3.89	4.65	20	-2.85	4.46
Above Mean	10	-4.20	3.55	12	-6.17	7.36	22	-5.27	5.90
ATI-C (SE)									
Below Mean	12	-1.67	4.19	11	-2.55	2.94	23	-2.09	3.59
Above Mean	9	-4.89	3.18	10	-8.10	7.80	19	-6.58	6.13
ATI-C (SA)									
Below Mean	11	-2.00	4.22	8	-4.25	3.92	19	-2.95	4.14
Above Mean	10	-4.20	3.71	13	-5.77	7.51	23	-5.09	6.08
ATI-C (V)									
Below Mean	13	-3.46	3.36	9	-3.44	3.13	22	-3.45	3.19
Above Mean	8	-2.38	5.15	12	-6.50	7.80	20	-4.85	7.02
ATI-C (A)									
Below Mean	8	-2.13	3.76	9	-4.44	4.56	17	-3.35	4.24
Above Mean	13	-3.62	4.25	12	-5.75	7.51	25	-4.64	6.01
ATI-C (R)									
Below Mean	14	-3.00	4.40	10	-7.60	7.72	24	-4.92	6.30
Above Mean	7	-3.14	3.53	11	-3.00	3.85	18	-3.06	3.62
ATI-C (K)									
Below Mean	8	-1.88	4.09	9	-4.78	5.38	17	-3.41	4.90
Above Mean	13	-3.77	4.00	12	-5.50	7.14	25	-4.60	5.67
Overall	21	-3.05	4.04	21	-5.19	6.31	42	-4.12	5.34

N: Number of participants, **M:** Mean, **SD:** Standard deviation
NIG: Non-interactive group, **IG**: interactive group, **LoC:** Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VMI.5 – Overall Mental Effort Yield Statistical Description by ATI-C and Interactivity Group

Overall Accuracy Yield	NIG			IG			All		
	N	M	SD	N	M	SD	N	M	SD
ATI-C (LoC)									
Below Mean	11	4.00	4.52	9	6.56	6.39	20	5.15	5.44
Above Mean	10	5.50	3.14	12	7.92	7.56	22	6.82	5.97
ATI-C (SE)									
Below Mean	12	2.83	2.72	11	6.64	5.45	23	4.65	4.58
Above Mean	9	7.22	3.93	10	8.10	8.54	19	7.68	6.60
ATI-C (SA)									
Below Mean	11	2.73	2.83	8	8.13	6.22	19	5.00	5.20
Above Mean	10	6.90	3.84	13	6.85	7.56	23	6.87	6.10
ATI-C (V)									
Below Mean	13	5.54	4.25	9	4.00	4.00	22	4.91	4.13
Above Mean	8	3.38	3.02	12	9.83	7.76	20	7.25	6.98
ATI-C (A)									
Below Mean	8	4.00	3.59	9	7.33	4.00	17	5.76	4.07
Above Mean	13	5.15	4.16	12	7.33	8.72	25	6.20	6.69
ATI-C (R)									
Below Mean	14	3.93	3.71	10	9.20	7.13	24	6.13	5.89
Above Mean	7	6.29	4.07	11	5.64	6.64	18	5.89	5.65
ATI-C (K)									
Below Mean	8	3.38	4.50	9	9.11	6.90	17	6.41	6.43
Above Mean	13	5.54	3.41	12	6.00	6.97	25	5.76	5.30
Overall	21	4.71	3.90	21	7.33	6.95	42	6.02	5.72

N: Number of participants, **M:** Mean, **SD:** Standard deviation
NIG: Non-interactive group, **IG**: interactive group, **LoC:** Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VMI.6 – Overall Accuracy Yield Statistical Description by ATI-C and Interactivity Group

		Overall Accuracy Yield		Overall Mental Effort Yield	
	N	F	p	F	p
Group	42	2.20	0.147	1.63	0.211
Group * ATI-C (LoC)	42	0.13	0.879	0.22	0.801
Group * ATI-C (SE)	42	0.53	0.595	3.69	0.035*
Group * ATI-C (SA)	42	0.49	0.619	0.76	0.475
Group * ATI-C (V)	42	4.61	0.017*	0.72	0.496
Group * ATI-C (A)	42	0.43	0.658	0.38	0.686
Group * ATI-C (R)	42	0.49	0.616	2.15	0.133
Group * ATI-C (K)	42	1.92	0.164	0.58	0.564

*$p < .05$, ** $p < .01$
N: Number of participants, ***F*:** F ratio, ***p*:** p-value
NIG: Non-interactive group, **IG**: interactive group, **LoC:** Locus of control, **SE**: Self-efficacy,
SA: Self-acceptance, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VMI.7 – Factorial ANOVA Results for the Overall Accuracy and Mental Effort Yield

	NIG									IG									All		
	Inferential			Procedural			All			Inferential			Procedural			All					
	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD
ATI-C (LoC)																					
Below Mean	9	0.37	0.24	10	0.30	0.60	11	0.33	0.46	7	0.46	0.48	9	0.57	0.65	9	0.52	0.57	20	0.42	0.51
Above Mean	6	0.42	0.38	10	0.55	0.48	10	0.50	0.44	9	0.47	0.24	12	0.73	0.88	12	0.62	0.68	22	0.57	0.58
ATI-C (SE)																					
Below Mean	9	0.44	0.29	11	0.10	0.41	12	0.25	0.39	9	0.55	0.42	11	0.52	0.42	11	0.53	0.41	23	0.39	0.42
Above Mean	6	0.33	0.31	9	0.82	0.42	9	0.62	0.45	7	0.35	0.21	10	0.81	1.04	10	0.62	0.83	19	0.62	0.67
ATI-C (SA)																					
Below Mean	8	0.38	0.26	10	0.14	0.40	11	0.25	0.36	6	0.61	0.41	8	0.70	0.60	8	0.66	0.51	19	0.43	0.47
Above Mean	7	0.41	0.35	10	0.71	0.54	10	0.58	0.48	10	0.38	0.30	13	0.63	0.89	13	0.52	0.69	23	0.55	0.61
All	15	0.39	0.29	20	0.42	0.55	21	0.41	0.45	16	0.47	0.35	21	0.66	0.77	21	0.57	0.63	42	0.49	0.55

LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **N**: Number of participants, **M:** Mean, **SD:** Standard deviation

Table VMI.8 – Transformed Accuracy Yield Statistical Description by Mental Effort, Interactivity Group and Psychometric Measures ATI-C

	NIG									IG									All		
	Inferential			Procedural			All			Inferential			Procedural			All					
	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD	N	M	SD
ATI-C (V)																					
Below Mean	10	0.33	0.32	13	0.55	0.54	13	0.45	0.46	7	0.35	0.24	9	0.32	0.51	9	0.33	0.40	22	0.40	0.44
Above Mean	5	0.51	0.21	7	0.19	0.52	8	0.33	0.43	9	0.55	0.41	12	0.92	0.85	12	0.76	0.71	20	0.60	0.65
ATI-C (A)																					
Below Mean	5	0.18	0.28	7	0.53	0.52	8	0.39	0.46	7	0.47	0.31	9	0.69	0.53	9	0.60	0.45	17	0.51	0.46
Above Mean	10	0.50	0.25	13	0.36	0.57	13	0.42	0.46	9	0.46	0.40	12	0.63	0.94	12	0.56	0.75	25	0.49	0.61
ATI-C (R)																					
Below Mean	10	0.37	0.36	13	0.32	0.55	14	0.34	0.47	7	0.52	0.22	10	0.89	0.89	10	0.73	0.71	24	0.51	0.60
Above Mean	5	0.43	0.06	7	0.61	0.53	7	0.54	0.40	9	0.42	0.43	11	0.45	0.62	11	0.44	0.53	18	0.48	0.48
ATI-C (K)																					
Below Mean	6	0.49	0.38	8	0.11	0.52	8	0.27	0.49	7	0.57	0.35	9	0.83	0.71	9	0.72	0.58	17	0.51	0.58
Above Mean	9	0.33	0.21	12	0.64	0.47	13	0.50	0.40	9	0.38	0.35	12	0.53	0.83	12	0.47	0.66	25	0.49	0.54
All	15	0.39	0.29	20	0.42	0.55	21	0.41	0.45	16	0.47	0.35	21	0.66	0.77	21	0.57	0.63	42	0.49	0.55

V: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic, **N**: Number of participants, **M:** Mean, **SD:** Standard deviation

Table VMI.9 – Transformed Accuracy Yield Statistical Description by Mental Effort, Interactivity Group and Learning Profiles ATI-C

	Accuracy Yield ‡		
	N	F	p
Group	36	1.39	0.244
Group * ATI-C (LoC)	36	0.03	0.859
Group * ATI-C (SE)	36	0.06	0.806
Group * ATI-C (SA)	36	0.28	0.602
Group * ATI-C (V)	36	5.83	0.019*
Group * ATI-C (A)	36	0.02	0.882
Group * ATI-C (R)	36	1.05	0.310
Group * ATI-C (K)	36	2.79	0.101
Effort * ATI-C (LoC)	36	0.01	0.939
Effort * ATI-C (SE)	36	5.23	0.026*
Effort * ATI-C (SA)	36	1.12	0.295
Effort * ATI-C (V)	36	0.09	0.767
Effort * ATI-C (A)	36	0.43	0.516
Effort * ATI-C (R)	36	0.19	0.662
Effort * ATI-C (K)	36	1.26	0.268
Effort * Group	36	0.73	0.397
Group * Effort * ATI-C (LoC)	36	0.00	0.947
Group * Effort * ATI-C (SE)	36	1.33	0.253
Group * Effort * ATI-C (SA)	36	0.54	0.464
Group * Effort * ATI-C (V)	36	3.43	0.070
Group * Effort * ATI-C (A)	36	0.88	0.351
Group * Effort * ATI-C (R)	36	0.17	0.678
Group * Effort * ATI-C (K)	36	2.72	0.105

*p < .05, ** p < .01
‡: Variable transformed**, N**: Number of participants, ***F*:** F ratio, ***p*:** p-value
NIG: Non-interactive group, **IG**: interactive group,
LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance,
V: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VMI.10 – Univariate Factorial ANOVA Results for the Accuracy Yield

		N	MR	χ^2	p
NIG		42	44.43	0.53	0.465
IG		42	40.57		
Accurate	NIG	21	22.45	0.26	0.613
	IG	21	20.55		
Inaccurate	NIG	21	23.21	0.85	0.357
	IG	21	19.79		

*p < .05, ** p < .01

NIG: Non-interactive group, **IG**: Interactive group, **N**: Number of participants, **MR:** Mean rank, χ^2**:** chi square, ***p*****:** p-value

Table VMI.11 – Kruskal-Wallis Test Results for the Mental Effort Yield by Interactivity Treatment and by Accuracy and Interactivity Group as the Between Group Factor

	Accurate				Inaccurate			All			
	N	M	SD	Mdn	M	SD	Mdn	N	M	SD	Mdn
Non-Interactive-Group											
LoC (Low)	7	-0.86	4.53	-1.50	-1.29	2.80	0.00	14	-1.07	3.63	-0.50
LoC (Mid)	7	-3.07	3.36	-1.50	-0.50	1.73	0.00	14	-1.79	2.89	-1.25
LoC (High)	7	-2.14	3.76	-3.00	-1.29	1.38	-2.00	14	-1.71	2.76	-2.00
SE (Low)	8	0.50	3.75	0.50	-1.75	2.59	-1.00	16	-0.63	3.32	-1.00
SE (Mid)	7	-2.00	0.96	-1.50	-1.14	1.49	-1.50	14	-1.57	1.28	-1.50
SE (High)	6	-5.42	3.72	-6.00	0.08	1.20	0.00	12	-2.67	3.90	-1.00
SA (Low)	8	-0.31	2.14	-1.25	-1.44	2.68	-1.00	16	-0.88	2.41	-1.00
SA (Mid)	6	-3.08	4.27	-3.00	-1.08	1.43	-1.00	12	-2.08	3.21	-2.00
SA (High)	7	-3.07	4.69	-3.00	-0.50	1.61	0.00	14	-1.79	3.63	-1.00
All	21	-2.02	3.83	-1.50	-1.02	1.99	-1.00	42	-1.52	3.06	-1.00
Interactive-Group											
LoC (Low)	5	-2.00	3.81	-3.00	-0.60	1.14	-1.00	10	-1.30	2.75	-1.00
LoC (Mid)	9	-3.61	5.53	-1.00	-1.61	1.83	-1.00	18	-2.61	4.13	-1.00
LoC (High)	7	-5.57	8.18	-3.00	-1.43	1.10	-1.00	14	-3.50	6.00	-2.00
SE (Low)	4	-2.00	2.94	-1.50	-1.25	0.96	-1.50	8	-1.63	2.07	-1.50
SE (Mid)	8	-2.56	4.62	-2.00	-0.81	1.13	-0.75	16	-1.69	3.38	-1.00
SE (High)	9	-5.89	7.92	-3.00	-1.78	1.84	-1.00	18	-3.83	5.97	-1.50
SA (Low)	5	-4.20	5.22	-4.00	-1.40	0.89	-2.00	10	-2.80	3.82	-2.00
SA (Mid)	10	-3.90	7.32	-1.50	-1.20	1.21	-1.00	20	-2.55	5.29	-1.00
SA (High)	6	-3.58	5.48	-2.50	-1.42	2.29	-0.75	12	-2.50	4.16	-1.00
All	21	-3.88	6.09	-2.00	-1.31	1.46	-1.00	42	-2.60	4.57	-1.00
Overall											
LoC (Low)	12	-1.33	4.10	-1.75	-1.00	2.21	-0.50	24	-1.17	3.23	-1.00
LoC (Mid)	16	-3.38	4.57	-1.25	-1.13	1.82	-1.00	32	-2.25	3.61	-1.00
LoC (High)	14	-3.86	6.37	-3.00	-1.36	1.20	-1.50	28	-2.61	4.67	-2.00
SE (Low)	12	-0.33	3.58	-1.00	-1.58	2.14	-1.00	24	-0.96	2.96	-1.00
SE (Mid)	15	-2.30	3.34	-1.50	-0.97	1.27	-1.00	30	-1.63	2.58	-1.50
SE (High)	15	-5.70	6.39	-4.00	-1.03	1.83	-1.00	30	-3.37	5.19	-1.00
SA (Low)	13	-1.81	3.95	-1.50	-1.42	2.11	-1.00	26	-1.62	3.11	-1.25
SA (Mid)	16	-3.59	6.20	-2.50	-1.16	1.25	-1.00	32	-2.38	4.57	-1.50
SA (High)	13	-3.31	4.86	-3.00	-0.92	1.92	-0.50	26	-2.12	3.82	-1.00
All	42	-2.95	5.11	-1.75	-1.17	1.73	-1.00	84	-2.06	3.90	-1.00

LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance
N: Number of participants, **M:** Mean, **SD:** Standard deviation, **Mdn:** Median

Table VMI.12 – Mental Effort Yield Statistical Description

by Accuracy, Interactivity Group and Psychometric Measures Scales

	Accurate				Inaccurate			All			
	N	M	SD	Mdn	M	SD	Mdn	N	M	SD	Mdn
Non-Interactive-Group											
V (Low)	6	-0.75	2.96	-1.25	-1.08	1.63	-1.50	12	-0.92	2.28	-1.25
V (Mid)	9	-2.72	4.81	-2.00	-0.17	1.15	0.00	18	-1.44	3.64	-1.00
V (High)	6	-2.25	3.16	-1.25	-2.25	2.82	-1.50	12	-2.25	2.86	-1.25
A (Low)	8	-2.31	3.87	-1.25	0.19	1.22	0.25	16	-1.06	3.06	-1.00
A (Mid)	7	-1.86	5.05	-3.00	-1.00	1.29	0.00	14	-1.43	3.57	-1.50
A (High)	6	-1.83	2.62	-1.50	-2.67	2.48	-1.75	12	-2.25	2.47	-1.50
R (Low)	7	-0.50	3.67	-1.50	-2.36	2.51	-2.00	14	-1.43	3.17	-1.50
R (Mid)	9	-3.33	4.26	-3.00	-0.11	1.34	0.00	18	-1.72	3.49	-1.00
R (High)	5	-1.80	2.95	-1.00	-0.80	1.30	0.00	10	-1.30	2.21	-1.00
K (Low)	7	-0.71	2.61	-1.50	-2.00	2.84	-2.00	14	-1.36	2.71	-1.50
K (Mid)	7	-0.93	4.25	-1.00	-0.21	0.99	0.00	14	-0.57	2.99	0.00
K (High)	7	-4.43	3.72	-3.00	-0.86	1.49	-1.00	14	-2.64	3.30	-1.50
All	21	-2.02	3.83	-1.50	-1.02	1.99	-1.00	42	-1.52	3.06	-1.00
Interactive-Group											
V (Low)	5	-1.10	2.70	0.00	-1.50	1.22	-2.00	10	-1.30	1.99	-1.25
V (Mid)	7	-2.14	1.95	-2.00	-1.43	2.30	-1.00	14	-1.79	2.08	-1.50
V (High)	9	-6.78	8.33	-4.00	-1.11	0.74	-1.00	18	-3.94	6.43	-1.00
A (Low)	6	-4.83	5.46	-4.00	-0.83	0.75	-1.00	12	-2.83	4.26	-1.50
A (Mid)	10	-3.25	7.10	-0.75	-1.75	1.83	-1.50	20	-2.50	5.11	-1.00
A (High)	5	-4.00	5.70	-2.00	-1.00	1.22	-1.00	10	-2.50	4.20	-1.50
R (Low)	5	-6.30	6.22	-3.00	-1.50	2.55	-0.50	10	-3.90	5.15	-1.50
R (Mid)	8	-4.44	7.67	-2.00	-1.44	1.35	-2.00	16	-2.94	5.54	-2.00
R (High)	8	-1.81	3.96	-1.50	-1.06	0.68	-1.00	16	-1.44	2.77	-1.00
K (Low)	7	-5.21	5.69	-4.00	-0.50	0.76	0.00	14	-2.86	4.60	-1.00
K (Mid)	8	-2.13	3.27	-1.00	-2.25	1.67	-2.00	16	-2.19	2.51	-2.00
K (High)	6	-4.67	9.30	-1.50	-1.00	1.22	-1.00	12	-2.83	6.61	-1.00
All	21	-3.88	6.09	-2.00	-1.31	1.46	-1.00	42	-2.60	4.57	-1.00

V: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic
N: Number of participants, **M:** Mean, **SD:** Standard deviation, **Mdn:** Median

Table VMI.13 – Mental Effort Yield Statistical Description

by Accuracy, Interactivity Group and Learning Profile Scales

	Accurate				Inaccurate			All			
	N	M	SD	Mdn	M	SD	Mdn	N	M	SD	Mdn
Overall											
V (Low)	11	-0.91	2.71	-1.00	-1.27	1.40	-2.00	22	-1.09	2.11	-1.25
V (Mid)	16	-2.47	3.73	-2.00	-0.72	1.80	0.00	32	-1.59	3.02	-1.00
V (High)	15	-4.97	6.96	-3.00	-1.57	1.87	-1.00	30	-3.27	5.30	-1.00
A (Low)	14	-3.39	4.60	-3.00	-0.25	1.14	0.00	28	-1.82	3.66	-1.00
A (Mid)	17	-2.68	6.20	-1.00	-1.44	1.63	-1.00	34	-2.06	4.51	-1.00
A (High)	11	-2.82	4.21	-1.50	-1.91	2.11	-1.50	22	-2.36	3.28	-1.50
R (Low)	12	-2.92	5.51	-1.75	-2.00	2.45	-1.25	24	-2.46	4.20	-1.50
R (Mid)	17	-3.85	5.93	-3.00	-0.74	1.47	-1.00	34	-2.29	4.54	-1.25
R (High)	13	-1.81	3.47	-1.00	-0.96	0.92	-1.00	26	-1.38	2.53	-1.00
K (Low)	14	-2.96	4.85	-1.75	-1.25	2.15	-0.75	28	-2.11	3.78	-1.25
K (Mid)	15	-1.57	3.67	-1.00	-1.30	1.71	-1.00	30	-1.43	2.82	-1.00
K (High)	13	-4.54	6.56	-2.00	-0.92	1.32	-1.00	26	-2.73	4.99	-1.25
All	42	-2.95	5.11	-1.75	-1.17	1.73	-1.00	84	-2.06	3.90	-1.00

V: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic
N: Number of participants, **M:** Mean, **SD:** Standard deviation, **Mdn:** Median

Table VMI.13 (Cont.) – Mental Effort yield Statistical Description

by Accuracy, Interactivity Group and Learning Profile Scales

Accuracy	Scale	Group	N	MR	χ^2	p
Accurate	V (Low)	NIG	6	6	0.00	1.000
		IG	5	6		
	V (Mid)	NIG	9	8.11	0.14	0.709
		IG	7	9		
	V (High)	NIG	6	9.17	0.68	0.409
		IG	9	7.22		
Inaccurate	V (Low)	NIG	6	6.33	0.14	0.707
		IG	5	5.6		
	V (Mid)	NIG	9	9.78	1.54	0.215
		IG	7	6.86		
	V (High)	NIG	6	7.33	0.24	0.625
		IG	9	8.44		
Accurate	A (Low)	NIG	8	8.63	1.37	0.241
		IG	6	6		
	A (Mid)	NIG	7	8.57	0.09	0.769
		IG	10	9.3		
	A (High)	NIG	6	6.33	0.14	0.713
		IG	5	5.6		
Inaccurate	A (Low)	NIG	8	9.13	2.96	0.085
		IG	6	5.33		
	A (Mid)	NIG	7	10.5	1.11	0.291
		IG	10	7.95		
	A (High)	NIG	6	5	1.28	0.258
		IG	5	7.2		

*$p < .05$, ** $p < .01$

NIG: Non-interactive group, **IG**: interactive group, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic, **N**: Number of participants, **MR:** Mean rank, χ^2**:** chi square, ***p*:** p-value

Table VMI.14 (Part 1) – Kruskal-Wallis Test Results for the Mental Effort Yield by Accuracy Factor, Learning Preference, and Interactivity Treatment as the Between Group Factor

Accuracy	Scale	Group	N	MR	χ^2	p
Accurate	R (Low)	NIG	7	7.93	2.66	0.103
		IG	5	4.5		
	R (Mid)	NIG	9	8.5	0.19	0.663
		IG	8	9.56		
	R (High)	NIG	5	7	0.00	1.000
		IG	8	7		
Inaccurate	R (Low)	NIG	7	5.5	1.31	0.253
		IG	5	7.9		
	R (Mid)	NIG	9	11.17	3.67	0.055
		IG	8	6.56		
	R (High)	NIG	5	8.3	0.99	0.320
		IG	8	6.19		
Accurate	K (Low)	NIG	7	9.36	2.80	0.094
		IG	7	5.64		
	K (Mid)	NIG	7	8.64	0.27	0.601
		IG	8	7.44		
	K (High)	NIG	7	6.21	0.63	0.429
		IG	6	7.92		
Inaccurate	K (Low)	NIG	7	6.07	1.73	0.189
		IG	7	8.93		
	K (Mid)	NIG	7	11.14	6.77	0.009**
		IG	8	5.25		
	K (High)	NIG	7	7.21	0.05	0.829
		IG	6	6.75		

*p < .05, ** p < .01
NIG: Non-interactive group, **IG**: interactive group, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic, **N**: Number of participants, **MR:** Mean rank, $\boldsymbol{\chi^2}$**:** chi square, ***p*:** p-value

Table VMI.14 (Part 2) – Kruskal-Wallis Test Results for the Mental Effort Yield by Accuracy Factor, Learning Preference, and Interactivity Treatment as the Between Group Factor

Group	Scale	Accuracy	N	MR	χ^2	p
NIG	V (Low)	Accurate	6	6.33	0.03	0.871
		Inaccurate	6	6.67		
	V (Mid)	Accurate	9	7.5	2.55	0.111
		Inaccurate	9	11.5		
	V (High)	Accurate	6	6.5	0.00	1.000
		Inaccurate	6	6.5		
IG	V (Low)	Accurate	5	5.9	0.18	0.673
		Inaccurate	5	5.1		
	V (Mid)	Accurate	7	6.5	0.84	0.361
		Inaccurate	7	8.5		
	V (High)	Accurate	9	7.78	1.92	0.166
		Inaccurate	9	11.22		
NIG	A (Low)	Accurate	8	6.25	3.64	0.056
		Inaccurate	8	10.75		
	A (Mid)	Accurate	7	6.57	0.71	0.398
		Inaccurate	7	8.43		
	A (High)	Accurate	6	6.92	0.17	0.684
		Inaccurate	6	6.08		
IG	A (Low)	Accurate	6	4.5	3.81	0.051
		Inaccurate	6	8.5		
	A (Mid)	Accurate	10	11.15	0.25	0.620
		Inaccurate	10	9.85		
	A (High)	Accurate	5	4.6	0.93	0.335
		Inaccurate	5	6.4		

*p < .05, ** p < .01
NIG: Non-interactive group, **IG**: interactive group, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic, **N**: Number of participants, **MR:** Mean rank, χ^2**:** chi square, ***p*:** p-value

Table VMI.15 (Part 1) – Kruskal-Wallis Test Results for the Mental Effort Yield by Accuracy Factor, Learning Preference, and Interactivity Treatment as the Between Group Factor

			N	MR	χ^2	p
NIG	R (Low)	Accurate	7	8.64	1.07	0.302
		Inaccurate	7	6.36		
	R (Mid)	Accurate	9	6.78	4.73	0.030*
		Inaccurate	9	12.22		
	R (High)	Accurate	5	4.7	0.74	0.390
		Inaccurate	5	6.3		
IG	R (Low)	Inaccurate	5	3.9	2.83	0.093
		Accurate	5	7.1		
	R (Mid)	Inaccurate	8	7.81	0.34	0.560
		Accurate	8	9.19		
	R (High)	Inaccurate	8	8.06	0.14	0.708
		Accurate	8	8.94		
NIG	K (Low)	Accurate	7	8.07	0.27	0.605
		Inaccurate	7	6.93		
	K (Mid)	Accurate	7	7.21	0.07	0.797
		Inaccurate	7	7.79		
	K (High)	Accurate	7	5.43	3.49	0.062
		Inaccurate	7	9.57		
IG	K (Low)	Inaccurate	7	5.21	4.29	0.038*
		Accurate	7	9.79		
	K (Mid)	Inaccurate	8	9.25	0.41	0.523
		Accurate	8	7.75		
	K (High)	Inaccurate	6	6	0.24	0.627
		Accurate	6	7		

*p < .05, ** p < .01
NIG: Non-interactive group, **IG**: interactive group, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic**,**
N: Number of participants, **MR:** Mean rank, χ^2**:** chi square, **p:** p-value

Table VMI.15 (Part 2) – Kruskal-Wallis Test Results for the Mental Effort Yield by Accuracy Factor, Learning Preference, and Interactivity Treatment as the Between Group Factor

		N	MR	χ^2	p
LoC (Low)	NIG	14	13.14	0.28	0.596
	IG	10	11.60		
LoC (Mid)	NIG	14	17.00	0.07	0.788
	IG	18	16.11		
LoC (High)	NIG	14	14.86	0.05	0.817
	IG	14	14.14		
SE (Low)	NIG	16	13.34	0.70	0.404
	IG	8	10.81		
SE (Mid)	NIG	14	14.46	0.37	0.543
	IG	16	16.41		
SE (High)	NIG	12	16.75	0.41	0.522
	IG	18	14.67		
SA (Low)	NIG	16	15.13	1.92	0.166
	IG	10	10.90		
SA (Mid)	NIG	12	15.92	0.08	0.784
	IG	20	16.85		
SA (High)	NIG	14	13.71	0.02	0.876
	IG	12	13.25		

*$p < .05$, ** $p < .01$

NIG: Non-interactive group, **IG**: interactive group, **LoC**: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **N**: Number of participants, **MR:** Mean rank, χ^2**:** chi square, ***p*:** p-value

Table VMI.16 – Kruskal-Wallis Test Results for the Mental Effort Yield by Psychometric Measure and Interactivity Group as the Between Group Factor

Scale	Group	N	MR	χ^2	p
V (Low)	NIG	12	11.67	0.02	0.894
	IG	10	11.30		
V (Mid)	NIG	18	17.69	0.68	0.410
	IG	14	14.96		
V (High)	NIG	12	15.96	0.06	0.814
	IG	18	15.19		
A (Low)	NIG	16	16.31	1.85	0.174
	IG	12	12.08		
A (Mid)	NIG	14	17.79	0.02	0.888
	IG	20	17.30		
A (High)	NIG	12	10.92	0.22	0.639
	IG	10	12.20		
R (Low)	NIG	14	13.04	0.20	0.659
	IG	10	11.75		
R (Mid)	NIG	18	18.72	0.58	0.445
	IG	16	16.13		
R (High)	NIG	10	14.15	0.12	0.727
	IG	16	13.09		
K (Low)	NIG	14	14.57	0.00	0.963
	IG	14	14.43		
K (Mid)	NIG	14	18.50	3.11	0.078
	IG	16	12.88		
K (High)	NIG	14	12.64	0.39	0.534
	IG	12	14.50		

**p* < .05, ** *p* < .01
NIG: Non-interactive group, **IG**: interactive group, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic**, N**: Number of participants, **MR:** Mean rank, χ^2**:** chi square, ***p*:** p-value

Table VMI.17 – Kruskal-Wallis Test Results for the Mental Effort Yield by Learning Preference and Interactivity Group as the Between Group Factor

Accuracy	Scale	Group	N	MR	χ^2	*p*
Accurate	LoC (Low)	NIG	7	7.00	0.33	0.568
		IG	5	5.80		
	LoC (Mid)	NIG	7	7.64	0.41	0.522
		IG	9	9.17		
	LoC (High)	NIG	7	8.00	0.20	0.653
		IG	7	7.00		
Inaccurate	LoC (Low)	NIG	7	6.64	0.03	0.866
		IG	5	6.30		
	LoC (Mid)	NIG	7	9.93	1.16	0.282
		IG	9	7.39		
	LoC (High)	NIG	7	7.64	0.02	0.897
		IG	7	7.36		
Accurate	SE (Low)	NIG	8	7.31	1.23	0.267
		IG	4	4.88		
	SE (Mid)	NIG	7	7.93	0.00	0.953
		IG	8	8.06		
	SE (High)	NIG	6	7.25	0.28	0.595
		IG	9	8.50		
Inaccurate	SE (Low)	NIG	8	6.63	0.03	0.862
		IG	4	6.25		
	SE (Mid)	NIG	7	7.36	0.28	0.594
		IG	8	8.56		
	SE (High)	NIG	6	11.25	5.48	0.019*
		IG	9	5.83		
Accurate	SA (Low)	NIG	8	8.00	1.39	0.238
		IG	5	5.4		
	SA (Mid)	NIG	6	7.75	0.24	0.623
		IG	10	8.95		
	SA (High)	NIG	7	6.79	0.05	0.830
		IG	6	7.25		
Inaccurate	SA (Low)	NIG	8	7.69	0.68	0.409
		IG	5	5.90		
	SA (Mid)	NIG	6	8.92	0.08	0.783
		IG	10	8.25		
	SA (High)	NIG	7	7.71	0.53	0.466
		IG	6	6.17		

*$p < .05$, ** $p < .01$
NIG: Non-interactive group, **IG**: interactive group, **LoC**: Locus of control, **SE**: Self-efficacy,
SA: Self-acceptance, **N**: Number of participants, **MR:** Mean rank, χ^2**:** chi square, ***p*:** p-value

Table VMI.18 – Kruskal-Wallis Test results for the Mental Effort Yield

by Accuracy Factor,Psychometric Measure,

and Interactivity Treatment as the Between Group Factor

Group	Scale	Accuracy	N	MR	χ^2	*p*
NIG	LoC (Low)	Accurate	7	7.43	0.00	0.949
		Inaccurate	7	7.57		
	LoC (Mid)	Accurate	7	6.00	1.84	0.175
		Inaccurate	7	9.00		
	LoC (High)	Accurate	7	6.43	0.95	0.331
		Inaccurate	7	8.57		
IG	LoC (Low)	Accurate	5	4.40	1.35	0.245
		Inaccurate	5	6.60		
	LoC (Mid)	Accurate	9	9.44	0.00	0.964
		Inaccurate	9	9.56		
	LoC (High)	Accurate	7	6.36	1.05	0.305
		Inaccurate	7	8.64		
NIG	SE (Low)	Accurate	8	9.81	1.23	0.268
		Inaccurate	8	7.19		
	SE (Mid)	Accurate	7	6.50	0.84	0.361
		Inaccurate	7	8.50		
	SE (High)	Accurate	6	3.83	6.71	0.010**
		Inaccurate	6	9.17		
IG	SE (Low)	Accurate	4	4.38	0.02	0.882
		Inaccurate	4	4.63		
	SE (Mid)	Accurate	8	7.25	1.12	0.290
		Inaccurate	8	9.75		
	SE (High)	Accurate	9	8.50	0.64	0.423
		Inaccurate	9	10.5		
NIG	SA (Low)	Accurate	8	8.63	0.01	0.915
		Inaccurate	8	8.38		
	SA (Mid)	Accurate	6	5.17	1.68	0.195
		Inaccurate	6	7.83		
	SA (High)	Accurate	7	5.79	2.39	0.122
		Inaccurate	7	9.21		
IG	SA (Low)	Accurate	5	4.90	0.41	0.525
		Inaccurate	5	6.10		
	SA (Mid)	Accurate	10	9.55	0.53	0.468
		Inaccurate	10	11.45		
	SA (High)	Accurate	6	5.83	0.42	0.518
		Inaccurate	6	7.17		

*$p < .05$, ** $p < .01$
NIG: Non-interactive group, **IG**: interactive group, **LoC**: Locus of control, **SE**: Self-efficacy,
SA: Self-acceptance, **N**: Number of participants, **MR:** Mean rank, χ^2**:** chi square, ***p*:** p-value

Table VMI.19 – Kruskal-Wallis Test Results for the Mental Effort Yield by Interactivity Treatment, Psychometric Measure, and Accuracy as the Between Group Factor

Appendix 15 VTI Statistical Description and Results Tables

	Non- Interactive			Interactive			All		
	N	M	SD	N	M	SD	N	M	SD
Accurate	21	53.47	10.35	21	61.59	8.66	42	57.53	10.28
Inaccurate	21	31.11	7.70	21	24.97	5.97	42	28.04	7.48
Unidentified	21	15.42	9.67	21	13.44	10.39	42	14.43	9.97

N: Number of participants, **M:** Mean, **SD:** Standard deviation

Table VTI.20 – Overall Percentage of (Accurate, Inaccurate and Unidentified) Insights Statistical Description by Interactivity Treatment

	NIG							IG							All						
		Accurate		Inaccurate		Unidentified			Accurate		Inaccurate		Unidentified			Accurate		Inaccurate		Unidentified	
	N	M	SD	M	SD	M	SD	N	M	SD	M	SD	M	SD	N	M	SD	M	SD	M	SD
ATI-C (LoC)																					
Below M	11	51.29	10	31.79	10.64	16.92	11.43	9	57.33	11.3	23.13	5.31	19.54	12.95	20	54.01	10.9	27.9	9.6	18.1	12.05
Above M	10	55.83	12.14	30.82	9.8	13.36	10.98	12	64.39	9.09	26.55	8.61	9.06	8.27	22	60.5	11.31	28.49	9.31	11.02	9.72
ATI-C (SE)																					
Below M	12	50.91	10.55	31.87	9.74	17.22	11.06	11	61.17	10.36	25.04	7.5	13.79	13.09	23	55.81	11.57	28.61	9.31	15.58	12.06
Above M	9	56.84	11.35	30.6	10.88	12.56	11.2	10	61.58	11.07	25.13	7.68	13.29	10.13	19	59.34	11.31	27.72	9.61	12.94	10.51
ATI-C (SA)																					
Below M	11	49.51	9.72	32.54	9.9	17.95	11.03	8	61.04	8.42	27.48	6.94	11.48	9.13	19	54.37	10.75	30.41	9.03	15.23	10.65
Above M	10	57.78	11.27	30	10.48	12.22	10.93	13	61.57	11.86	23.61	7.57	14.83	12.94	23	59.92	11.64	26.39	9.4	13.69	12.05

LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **N**: Number of participants, **M:** Mean, **SD:** Standard deviation

Table VTI.21 – Transformed Accuracy Statistical Description by Interactivity Group and Psychometric Measures ATI-C

	NIG							IG							All						
		Accurate		Inaccurate		Unidentified			Accurate		Inaccurate		Unidentified			Accurate		Inaccurate		Unidentified	
	N	M	SD	M	SD	M	SD	N	M	SD	M	SD	M	SD	N	M	SD	M	SD	M	SD
ATI-C (V)																					
Below M	13	53.32	12.36	31.04	10.64	15.64	12.53	9	60.07	11.58	25.08	8.25	14.85	13.14	22	56.08	12.37	28.6	10.08	15.32	12.64
Above M	8	53.66	9.28	31.8	9.58	14.54	9.06	12	62.34	9.9	25.09	7.06	12.58	10.56	20	58.87	10.46	27.77	8.7	13.36	9.91
ATI-C (A)																					
Below M	8	51.31	9.84	31.51	11.61	17.19	13.86	9	59.7	11.13	24.63	9.44	15.67	13.29	17	55.75	11.22	27.87	10.92	16.38	13.37
Above M	13	54.77	11.9	31.22	9.36	14.02	9.35	12	62.61	10.2	25.43	5.83	11.96	10.23	25	58.53	11.69	28.44	8.32	13.03	9.74
ATI-C (R)																					
Below M	14	55.77	10.75	30.13	9.74	14.1	11.03	10	61.23	11.55	24.22	8.22	14.55	13.41	24	58.05	11.3	27.67	9.51	14.29	11.94
Above M	7	48.81	10.89	33.72	10.85	17.47	11.69	11	61.49	9.88	25.87	6.87	12.65	9.99	18	56.56	11.91	28.92	9.33	14.52	10.79
ATI-C (K)																					
Below M	8	51.6	10.75	31.42	10.65	16.98	11.19	9	59.04	11.98	24.48	8.03	16.49	13.36	17	55.54	11.86	27.75	9.85	16.72	12.2
Above M	13	54.59	11.47	31.27	10.02	14.14	11.34	12	63.11	9.26	25.54	7.21	11.35	9.88	25	58.68	11.22	28.52	9.17	12.8	10.65

V: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic **N**: Number of participants, **M:** Mean, **SD:** Standard deviation

Table VTI.22 – Transformed Accuracy Statistical Description by Interactivity Group and Learning Profiles ATI-C

		Accurate		Inaccurate		Unidentified	
	N	F	p	F	p	F	p
Group	42	10.32	0.003**	8.13	0.008**	0.84	0.368
Group * ATI-C (LoC)	42	2.03	0.164	0.04	0.845	2.20	0.147
Group * ATI-C (SE)	42	0.29	0.592	0.25	0.621	0.02	0.88
Group * ATI-C (SA)	42	1.53	0.225	0.23	0.632	0.67	0.418
Group * ATI-C (V)	42	0.08	0.784	0.04	0.845	0.01	0.911
Group * ATI-C (A)	42	0.29	0.591	0.10	0.76	0.08	0.784
Group * ATI-C (R)	42	1.91	0.176	0.30	0.585	0.76	0.391
Group * ATI-C (K)	42	0.21	0.652	0.16	0.688	0.02	0.899

*p < .05, ** p < .01
N: Number of participants, ***F*:** F ratio, ***p*:** p-value **LoC:** Locus of control, **SE**: Self-efficacy,
SA: Self-acceptance, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VTI.23 – Univariate ANOVA Results

for the Overall Percentage of (Accurate, Inaccurate and Unidentified) Insights

	N	F	p
Group	42	14.57	0.001**
Group * HiLo_LoC	42	0.19	0.828
Group * HiLo_SE	42	0.88	0.424
Group * HiLo_SA	42	2.32	0.113
Group * HiLo_V	42	0.02	0.984
Group * HiLo_A	42	0.25	0.779
Group * HiLo_R	42	1.94	0.160
Group * HiLo_K	42	0.58	0.567

*p < .05, ** p < .01
N: Number of participants, ***F*:** F ratio, ***p*:** p-value
LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance,
V: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VTI.24 – Univariate ANOVA Results for the Overall Score

by Interaction Treatment and ATI-C

	NIG			IG			All		
	N	M	SD	N	M	SD	N	M	SD
ATI-C (LoC)									
Below Mean	11	61.95	8.15	9	71.32	3.76	20	66.16	8.03
Above Mean	10	64.29	8.67	12	71.08	6.56	22	67.99	8.29
ATI-C (SE)									
Below Mean	12	61.49	8.73	11	71.24	4.73	23	66.15	8.6
Above Mean	9	65.16	7.65	10	71.12	6.3	19	68.3	7.56
ATI-C (SA)									
Below Mean	11	60.45	8.37	8	69.08	4.62	19	64.08	8.22
Above Mean	10	65.94	7.62	13	72.48	5.65	23	69.63	7.32
ATI-C (V)									
Below Mean	13	63.16	8.21	9	70.88	5.71	22	66.32	8.21
Above Mean	8	62.9	8.92	12	71.41	5.39	20	68.01	8.14
ATI-C (A)									
Below Mean	8	62.18	8.10	9	71.13	6.86	17	66.92	8.69
Above Mean	13	63.61	8.66	12	71.22	4.28	25	67.26	7.88
ATI-C (R)									
Below Mean	14	64.95	7.43	10	71.78	6.22	24	67.8	7.71
Above Mean	7	59.28	9.17	11	70.64	4.76	18	66.22	8.77
ATI-C (K)									
Below Mean	8	62.29	8.79	9	70.82	5.98	17	66.8	8.55
Above Mean	13	63.54	8.25	12	71.46	5.16	25	67.34	7.98
All	21	63.06	8.46	21	71.18	5.52	42	67.12	8.21

*$p < .05$, ** $p < .01$

N: Number of participants, **M:** Mean, **SD:** Standard deviation

Table VTI.25 – Overall Score Statistical Description by Interactivity Treatment and Individual Differences ATI-C

		NIG			IG			All		
		N	M	SD	N	M	SD	N	M	SD
2D	Accurate	21	53.92	13.21	21	65.66	8.22	42	59.79	12.39
	Inaccurate	21	27.16	9.30	21	22.94	5.51	42	25.05	7.85
	Unidentified	21	18.91	10.96	21	11.40	8.31	42	15.16	10.33
3D	Accurate	21	52.98	8.98	21	57.07	11.09	42	55.02	10.18
	Inaccurate	21	35.49	9.36	21	27.23	8.66	42	31.36	9.84
	Unidentified	21	11.53	10.48	21	15.70	14.10	42	13.62	12.45

N: Number of participants, **M:** Mean, **SD:** Standard deviation

Table VTI.26 – Percentage of Accurate, Inaccurate and Unidentified Insights Statistical Description by Representation and Interactivity Treatment

	Accurate		Inaccurate		Unidentified	
	F	*p*	*F*	*p*	*F*	*p*
Group	9.81	0.004**	8.14	0.008**	0.67	0.418
Group * ATI-C (LoC)	2.04	0.162	0.05	0.829	2.22	0.145
Group * ATI-C (SE)	0.35	0.560	0.25	0.618	0.04	0.852
Group * ATI-C (SA)	1.62	0.211	0.22	0.639	0.72	0.402
Group * ATI-C (V)	0.08	0.775	0.03	0.861	0.02	0.893
Group * ATI-C (A)	0.26	0.614	0.07	0.793	0.08	0.786
Group * ATI-C (R)	1.81	0.188	0.30	0.589	0.69	0.414
Group * ATI-C (K)	0.19	0.669	0.16	0.692	0.01	0.914
Representation	4.81	0.036*	10.25	0.003**	0.99	0.327
Representation * Group	10.66	0.003**	0.34	0.564	14.08	0.001**
Representation * Group * ATI-C (LoC)	0.00	0.995	0.42	0.519	0.44	0.510
Representation * Group * ATI-C (SE)	3.23	0.081	0.06	0.810	2.22	0.146
Representation * Group * ATI-C (SA)	1.59	0.216	0.02	0.885	1.86	0.181
Representation * ATI-C (V)	0.06	0.803	2.76	0.107	1.91	0.176
Representation * ATI-C (A)	2.39	0.132	0.47	0.499	0.70	0.409
Representation * ATI-C (R)	0.18	0.673	0.06	0.816	0.42	0.523
Representation * ATI-C (K)	0.41	0.528	4.05	0.053	1.82	0.187
Representation * Group * ATI-C (V)	0.22	0.642	0.34	0.565	1.06	0.311
Representation * Group * ATI-C (A)	1.86	0.182	1.71	0.200	0.00	0.961
Representation * Group * ATI-C (R)	2.39	0.132	0.00	0.972	2.17	0.151
Representation * Group * ATI-C (K)	0.94	0.339	0.01	0.913	0.70	0.409

*$p < .05$, ** $p < .01$, **N**: Number of participants, ***F*:** F ratio, ***p*:** p-value
LoC: Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance,
V: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic

Table VTI.27 – Repeated Measures Factorial ANOVA Results for the Accurate, Inaccurate and Unidentified Percentages with Representation as the Within Subject Factor and Interactivity Group and Individual Differences ATI-C as Between Subject Factor

		2D Score			3D Score		
	N	M	SD	Mdn	M	SD	Mdn
Non-Interactive-Group	21	66.14	11.46	65.00	60.10	8.80	59.38
Interactive-Group	21	74.07	5.81	72.73	67.88	7.81	68.42
All	42	70.10	9.83	70.78	63.99	9.11	66.09

$^*p < .05$, $^{**}\, p < .01$,**N**: Number of participants, **M:** Mean, **SD:** Standard deviation, **Mdn**: Median

Table VTI.28 – 2D and 3D Score Statistical Description by Interactivity Treatment

		N	MR	χ^2	p
2D		42	1.76 [a]	11.52	0.001**
3D		42	1.24 [a]		
NIG	2D	21	1.71 [a]	3.86	0.050
	3D	21	1.29 [a]		
IG	2D	21	1.81 [a]	8.05	0.005**
	3D	21	1.19 [a]		
2D	NIG	21	16.69 [b]	6.46	0.011*
	IG	21	26.31 [b]		
3D	NIG	21	16.31 [b]	7.52	0.006**
	IG	21	26.69 [b]		

$^*p < .05$, $^{**}\, p < .01$
[a] Friedman's ANOVA
[b] Kruskal-Wallis test
N: Number of participants, **MR:** Mean rank, $\boldsymbol{\chi^2}$**:** chi-squared, ***p***: p-value

Table VTI.29 – Kruskal-Wallis Test and Friedman's ANOVA Results for Post-experiment Questionnaire Scores by Representation and Interactivity Treatment

Psychometric Measures

Table VTI.30 provide a full statistical description of the 2D and 3D scores by interactivity treatment and psychometric measures scale.

	2D Score								3D Score							
	Non-Interactive-Group				Interactive-Group				Non-Interactive-Group				Interactive-Group			
	N	M	SD	Mdn	N	M	SD	Mdn	N	M	SD	Mdn	N	M	SD	Mdn
ATI-C (LoC)																
Below Mean	11	64.96	11.37	61.22	9	74.22	5.21	74.60	11	59.14	9.21	59.38	9	67.85	4.25	68.42
Above Mean	10	67.44	12.02	65.88	12	73.96	6.46	71.53	10	61.15	8.69	59.66	12	67.90	9.89	68.43
ATI-C (SE)																
Below Mean	12	63.02	10.88	60.85	11	73.42	4.61	72.22	12	60.13	9.66	59.16	11	68.73	7.62	68.89
Above Mean	9	70.30	11.47	70.15	10	74.78	7.10	72.73	9	60.05	8.09	60.38	10	66.94	8.32	67.86
ATI-C (SA)																
Below Mean	11	62.13	10.93	60.47	8	71.22	2.70	70.59	11	58.96	9.19	58.93	8	66.80	9.18	67.60
Above Mean	10	70.56	10.84	71.08	13	75.82	6.59	77.59	10	61.35	8.66	60.66	13	68.54	7.16	68.42

ATI-C: ATI Category, **LoC:** Locus of control, **SE**: Self-efficacy, **SA**: Self-acceptance, **N**: Number of participants, **M:** Mean, **SD:** Standard deviation, **Mdn**: Median

Table VTI.30 – 2D and 3D Scores Statistical Description by Interactivity Treatment and Psychometric Measures ATI-C

		2D Score						3D Score					
		N	MR	SR	U	*z*	*p*	N	MR	SR	U	*z*	*p*
ATI-C (LoC)													
Below Mean	NIG	11	8.18	90.00	24.00	-1.937	0.053	11	7.95	87.50	21.50	-2.131	0.033*
	IG	9	13.33	120.00				9	13.61	122.50			
Above Mean	NIG	10	9.00	90.00	35.00	-1.649	0.099	10	9.10	91.00	36.00	-1.583	0.113
	IG	12	13.58	163.00				12	13.50	162.00			
ATI-C (SE)													
Below Mean	NIG	12	8.67	104.00	26.00	-2.462	0.014*	12	9.33	112.00	34.00	-1.970	0.049*
	IG	11	15.64	172.00				11	14.91	164.00			
Above Mean	NIG	9	8.67	78.00	33.00	-0.980	0.327	9	7.56	68.00	23.00	-1.796	0.072
	IG	10	11.20	112.00				10	12.20	122.00			
ATI-C (SA)													
Below Mean	NIG	11	7.45	82.00	16.00	-2.313	0.021*	11	8.23	90.50	24.50	-1.611	0.107
	IG	8	13.50	108.00				8	12.44	99.50			
Above Mean	NIG	10	10.00	100.00	45.00	-1.240	0.215	10	8.80	88.00	33.00	-1.985	0.047*
	IG	13	13.54	176.00				13	14.46	188.00			

Table VTI.31 – Mann-Whitney Results for the 2D and 3D Scores

by Psychometric Measures ATI-C and Interaction Treatment as Between Group Factor

Learning Preferences

Table VTI.32 provide a full statistical description of the 2D and 3D scores by interactivity treatment and psychometric measures scale.

	2D Score								3D Score							
	Non-Interactive-Group				Interactive-Group				Non-Interactive-Group				Interactive-Group			
	N	M	SD	Mdn	N	M	SD	Mdn	N	M	SD	Mdn	N	M	SD	Mdn
ATI-C (V)																
Below Mean	13	66.81	11.74	65.08	9	74.46	4.64	72.73	13	59.70	9.16	60.38	9	66.71	9.37	67.39
Above Mean	8	65.06	11.69	64.19	12	73.77	6.75	72.60	8	60.75	8.75	59.16	12	68.75	6.71	68.71
ATI-C (A)																
Below Mean	8	62.95	9.93	60.85	9	73.88	6.06	72.22	8	62.11	11.53	63.26	9	68.15	10.61	70.37
Above Mean	13	68.10	12.26	70.15	12	74.21	5.89	73.67	13	58.86	6.86	58.93	12	67.67	5.37	67.91
ATI-C (R)																
Below Mean	14	67.80	10.16	65.04	10	73.90	5.80	72.34	14	62.28	9.07	60.64	10	69.53	9.36	71.13
Above Mean	7	62.83	13.95	60.00	11	74.22	6.11	74.60	7	55.72	6.81	51.06	11	66.37	6.15	67.39
ATI-C (K)																
Below Mean	8	63.51	10.75	60.85	9	72.78	3.58	72.46	8	61.24	11.08	62.49	9	68.53	11.11	70.37
Above Mean	13	67.76	12.00	65.08	12	75.04	7.05	76.30	13	59.39	7.49	59.38	12	67.39	4.53	67.86

ATI-C: ATI Category, **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic**, N**: Number of participants, **M:** Mean, **SD:** Standard deviation, **Mdn**: Median

Table VTI.32 – 2D and 3D Scores Statistical Description by Interactivity Treatment and Learning Preference ATI-C

		2D Score						3D Score					
		N	MR	SR	U	z	p	N	MR	SR	U	z	p
ATI-C (V)													
Below Mean	NIG	13	9.54	124.00	33.00	-1.703	0.089	13	9.65	125.50	34.50	-1.603	0.109
	IG	9	14.33	129.00				9	14.17	127.50			
Above Mean	NIG	8	7.25	58.00	22.00	-2.006	0.045*	8	7.19	57.50	21.50	-2.045	0.041*
	IG	12	12.67	152.00				12	12.71	152.50			
ATI-C (A)													
Below Mean	NIG	8	5.75	46.00	10.00	-2.502	0.012*	8	7.81	62.50	26.50	-0.915	0.360
	IG	9	11.89	107.00				9	10.06	90.50			
Above Mean	NIG	13	11.31	147.00	56.00	-1.197	0.231	13	8.69	113.00	22.00	-3.049	0.002**
	IG	12	14.83	178.00				12	17.67	212.00			

*$p < .05$, ** $p < .01$ **NIG**: Non-interactive group, **IG**: Interactive group, **ATI-C**: ATI Category **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic, **N**: Number of participants, **MR:** Mean rank, **SR**: Sum of ranks, ***z***: z-value, ***p***: p-value, ***r**:* effect size

Table VTI.33 – Mann-Whitney Results for the 2D and 3D Scores by Learning Preferences ATI-C (V) and ATI-C (A) and Interaction Treatment as Between Group Factor

		2D Score						3D Score					
		N	MR	SR	U	z	p	N	MR	SR	U	z	p
ATI-C (R)													
Below Mean	NIG	14	10.11	141.50	36.50	-1.962	0.050	14	10.25	143.50	38.50	-1.845	0.065
	IG	10	15.85	158.50				10	15.65	156.50			
Above Mean	NIG	7	7.00	49.00	21.00	-1.586	0.113	7	5.14	36.00	8.00	-2.769	0.006**
	IG	11	11.09	122.00				11	12.27	135.00			
ATI-C (K)													
Below Mean	NIG	8	6.56	52.50	16.50	-1.879	0.060	8	7.19	57.50	21.50	-1.396	0.163
	IG	9	11.17	100.50				9	10.61	95.50			
Above Mean	NIG	13	10.69	139.00	48.00	-1.632	0.103	13	8.85	115.00	24.00	-2.939	0.003**
	IG	12	15.50	186.00				12	17.50	210.00			

*p < .05, ** p < .01 **NIG**: Non-interactive group, **IG**: Interactive group, **ATI-C**: ATI Category **V**: Visual, **A**: Aural, **R**: Read-write, **K**: Kinaesthetic, **N**: Number of participants, **MR:** Mean rank, **SR**: Sum of ranks, ***z***: z-value, ***p***: p-value, ***r:*** effect size

Table VTI.34 – Mann-Whitney Results for the 2D and 3D Scores by Learning Preferences ATI-C (R) and ATI-C (K) and Interaction Treatment as Between Group Factor