Investigating
User Experience and User Engagement
for Design

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy
In the Faculty of Humanities

2014

Jennefer Hart

Manchester Business School
# Table of Contents

1 **CHAPTER 1 Introduction** .............................................................................................................. 19
   1.1 Introduction .......................................................................................................................... 20
      1.1.1 Conflicting Perspectives to UX ....................................................................................... 20
      1.1.2 UX Frameworks and Models .......................................................................................... 21
      1.1.3 User Quality Judgement Models ..................................................................................... 22
      1.1.4 Interactivity ................................................................................................................... 23
      1.1.5 Individual Differences .................................................................................................... 24
      1.1.6 Dynamics of UX ............................................................................................................ 25
   1.2 Aims and Objectives .............................................................................................................. 25
   1.3 Research Questions ............................................................................................................... 27
   1.4 Outline of Thesis ................................................................................................................... 28

2 **CHAPTER 2 Literature Review** ............................................................................................. 31
   2.1 Overview ............................................................................................................................ 32
   2.2 From Usability to Experience ............................................................................................. 33
      2.2.1 What is Usability? ........................................................................................................... 33
      2.2.2 The Problem with Usability .......................................................................................... 34
   2.3 Transcending Usability ........................................................................................................ 35
      2.3.1 Emotional Experience ................................................................................................... 36
      2.3.2 Affective Responses ...................................................................................................... 36
   2.4 Aesthetics and Usability ......................................................................................................... 38
   2.5 Hedonics and Usability .......................................................................................................... 40
   2.6 Defining User Experience ...................................................................................................... 41
   2.7 UX Theories, Frameworks and Models .................................................................................... 42
      2.7.1 Reductionist versus Holistic ............................................................................................ 42
      2.7.2 Early UX Frameworks ...................................................................................................... 43
      2.7.3 Modelling UX ................................................................................................................ 44
         2.7.3.1 Construct Models ...................................................................................................... 44
         2.7.3.2 Inference Models ....................................................................................................... 45
         2.7.3.3 Process Models .......................................................................................................... 46
         2.7.3.4 Technology Acceptance Models (TAM) .................................................................. 49
   2.8 Clarifying User Engagement ................................................................................................... 49
      2.8.1 Flow and Presence .......................................................................................................... 50
      2.8.2 Interactivity ................................................................................................................... 51
         2.8.2.1 Interactivity and HCI ............................................................................................... 52
   2.9 Expanding UX ....................................................................................................................... 53
      2.9.1 Individual Difference ....................................................................................................... 53
      2.9.2 UX Over Time ................................................................................................................ 55
   2.10 Methodology and Rationale ................................................................................................ 56
      2.10.1 The Studies .................................................................................................................. 57
Table of Contents

2.10.2. Measures ............................................................................................................. 59
2.11. Chapter Summary ................................................................................................. 60

3  CHAPTER 3 Study One: Interactivity and Affect ................................................. 61

3.1. Introduction .............................................................................................................. 62
3.2. Materials ................................................................................................................ 63
3.3. Method ..................................................................................................................... 66
  3.3.1. Experimental Design ......................................................................................... 66
        3.3.1.1. Procedure ................................................................................................. 66
        3.3.1.2. Participants .............................................................................................. 68
  3.3.2. Measurement Scales ......................................................................................... 68
        3.3.2.1. Demographics and Disposition ................................................................. 68
        3.3.2.2. Affect Scale ............................................................................................. 68
        3.3.2.3. Website Quality ....................................................................................... 69
        3.3.2.4. Immersion and Presence ....................................................................... 70
        3.3.2.5. Usability Scale ......................................................................................... 70
        3.3.2.6. Overall Experience ............................................................................... 71
        3.3.2.7. Criteria Importance ................................................................................. 71
  3.4. Quantitative Analysis and Results ................................................................. 71
     3.4.1. Reliability ........................................................................................................ 72
     3.4.2. Affect Scale ................................................................................................... 72
            3.4.2.1. Aggregated Affect Scale ................................................................. 73
            3.4.2.2. Individual Affect Scale Items .............................................................. 73
            3.4.2.3. Summary of Affect Scale Results ................................................. 76
     3.4.3. Participant Disposition .............................................................................. 77
            3.4.3.1. Prior Knowledge and Brand Awareness ........................................... 77
            3.4.3.2. Artistic and Aesthetic Appreciation ............................................... 78
            3.4.3.3. Participant Disposition ..................................................................... 78
     3.4.4. Website Quality Scale ............................................................................. 78
            3.4.4.1. Subscale Website Quality Analysis ................................................... 78
            3.4.4.2. Analysis of Individual Items within Website Quality .................... 80
            3.4.4.3. Summary of Website Quality Results ......................................... 83
     3.4.5. Immersion and Presence ....................................................................... 84
     3.4.6. Usability Scale .............................................................................................. 84
     3.4.7. Overall Preference .................................................................................... 84
            3.4.7.1. Experience Scale ............................................................................... 84
            3.4.7.2. Forced Choice Preference ............................................................... 84
     3.4.8. Inter-variable Relationships (Regressions) ............................................ 85
     3.4.9. Criteria Importance ................................................................................. 88
     3.4.10. Exploration Time ................................................................................... 89
     3.4.11. Summary of Quantitative Results ........................................................ 90
  3.5. Qualitative Analysis and Results ................................................................. 91
     3.5.1. Qualitative Analysis ............................................................................... 91
     3.5.2. Memory Recall ....................................................................................... 92
Table of Contents

3.5.3. Interview Comments........................................................................................................... 93
  3.5.3.1. Website Features........................................................................................................... 93
  3.5.3.2. Engagement.................................................................................................................. 96
  3.5.3.3. General Impressions..................................................................................................... 99
3.5.4. Qualitative Results Summary .......................................................................................... 100
3.6. Individual User Experience ................................................................................................. 102
3.7. Chapter Summary ................................................................................................................ 104

4  CHAPTER 4 Study Two: Interactivity – Avatars and Media................. 106
  4.1. Introduction....................................................................................................................... 107
  4.2. Materials ........................................................................................................................ 108
    4.2.1. Avatar and Videos....................................................................................................... 109
    4.2.2. Interactive Links and Map ....................................................................................... 109
  4.3. Method ............................................................................................................................ 110
    4.3.1. Experimental Design ............................................................................................... 110
      4.3.1.1. Procedure ........................................................................................................... 111
      4.3.1.2. Participants ....................................................................................................... 112
    4.3.2. Measurement Scale ................................................................................................ 112
  4.4. Quantitative Analysis and Results .................................................................................. 115
    4.4.1. Reliability ................................................................................................................ 115
    4.4.2. Website Design and Task Differences ..................................................................... 115
      4.4.2.1. Affect Scale ....................................................................................................... 116
      4.4.2.2. Website Quality Scale ...................................................................................... 117
    4.4.3. Analysis of Post-Interaction Experience ................................................................ 118
      4.4.3.1. PANAS Scale .................................................................................................... 118
      4.4.3.2. Immersion, Usability, Quality and Content ....................................................... 119
      4.4.3.3. Overall Experience .......................................................................................... 119
    4.4.4. Individual User Experience ..................................................................................... 119
      4.4.4.1. The Effect of Predisposition on Affect and Website Quality .............................. 120
    4.4.5. Criteria Importance ................................................................................................ 123
    4.4.6. Overall Site Preference ............................................................................................ 123
    4.4.7. Inter-variable Relationships (Regressions) ............................................................... 123
    4.4.8. Summary of Quantitative Results ............................................................................ 125
  4.5. Qualitative Analysis and Results ...................................................................................... 127
    4.5.1. Memory Recall ........................................................................................................ 127
    4.5.2. Interview Comments ............................................................................................... 127
    4.5.3. Qualitative Results Summary .................................................................................. 131
  4.6. Chapter Summary ............................................................................................................ 132

5  CHAPTER 5 Chapter Three: Longitudinal Study (Qualitative)............. 133
  5.1. Introduction..................................................................................................................... 134
  5.2. Study Design .................................................................................................................. 135
    5.2.1. The iPad .................................................................................................................. 135
    5.2.2. Participants ............................................................................................................. 135
5.2.3. Method...........................................................................................................136
  5.2.3.1. Questionnaires..........................................................................................136
  5.2.3.2. Semi-Structured Interviews.......................................................................136
  5.2.3.3. Diary-Study...............................................................................................137
5.2.4. Measures.......................................................................................................137
  5.2.4.1. Questionnaire Scales ................................................................................138
  5.2.4.2. Diary Study Mini-Questionnaires ...............................................................139
  5.2.4.3. Verbal Interview Ratings .........................................................................139
  5.2.4.4. User Predispositions ...............................................................................139
5.3. General Results.................................................................................................139
  5.3.1. Device Usage ..............................................................................................139
5.4. Activity Changes Over Time ............................................................................140
5.5. Quality Variables Analysis Over Time .............................................................141
  5.5.1. Questionnaire Results ................................................................................141
      5.5.1.1. Affect Scale ............................................................................................141
      5.5.1.2. Design Quality Scale ...........................................................................142
      5.5.1.3. Usefulness and Overall Experience Scales .............................................143
      5.5.1.4. Summary of Main Questionnaire Results .............................................144
  5.5.2. Diary Mini-Questionnaires ...........................................................................144
      5.5.2.1. Activities ................................................................................................144
      5.5.2.2. Affective Experience .............................................................................145
      5.5.2.3. Gender, Location and Pre-Owned Checks .............................................146
      5.5.2.4. Summary of Diary Mini-Prompts ..........................................................146
  5.5.3. Verbal Interview Ratings ............................................................................147
      5.5.3.1. Ease-of-Use, Attractive and Satisfaction ..............................................147
      5.5.3.2. Summary of Verbal Interview Items ....................................................148
  5.5.4. Inter-variable Relationships (Regressions)....................................................148
  5.5.5. Overall Summary .......................................................................................151
5.6. User Characteristics ..........................................................................................152
  5.6.1. Cluster Analysis ..........................................................................................152
5.7. Between Group Differences ............................................................................153
  5.7.1. Main Questionnaire Results (Between Groups) ..........................................153
      5.7.1.1. Affect .....................................................................................................153
      5.7.1.2. Design Quality .....................................................................................154
      5.7.1.3. Usefulness and Overall Satisfaction .....................................................155
  5.7.2. Diary Mini-Questionnaires ..........................................................................155
  5.7.3. Interview Verbal Scale Items .......................................................................155
  5.7.4. Between Group Summary .........................................................................156
5.7.5. Application Analysis ....................................................................................157
      5.7.5.1. Application Audit ................................................................................157
      5.7.5.2. Application Usage .................................................................................158
      5.7.5.3. Frequency of Downloads ......................................................................159
      5.7.5.4. Application Ratings ............................................................................160
6 CHAPTER 6 Chapter Three: Longitudinal Study (Qualitative) .......... 170

6.1. Introduction .................................................................................. 171
6.2. Qualitative Analysis ..................................................................... 171
  6.2.1. Interviews ................................................................................ 171
  6.2.2. Diary Prompts ......................................................................... 173
  6.2.3. Method ..................................................................................... 173
6.3. Qualitative Findings ..................................................................... 174
  6.3.1. Quality Ratings (Device and Applications) ............................ 175
     6.3.1.1. Utility and Content ............................................................ 176
     6.3.1.2. Functionality .................................................................... 176
     6.3.1.3. Usability ........................................................................... 177
     6.3.1.4. Usefulness ....................................................................... 177
     6.3.1.5. Ease-of-Use ................................................................. 178
     6.3.1.6. Aesthetics ................................................................. 178
  6.3.2. Motivation and Barriers ...................................................... 178
     6.3.2.1. Motivators ................................................................. 179
     6.3.2.2. Barriers ................................................................. 180
  6.3.3. Affect ....................................................................................... 182
     6.3.3.1. Positive Affect ............................................................ 182
     6.3.3.2. Negative Affect .......................................................... 183
  6.3.4. Interactivity ............................................................................ 184
  6.3.5. Activities ............................................................................... 186
     6.3.5.1. Work Activity ............................................................... 186
     6.3.5.2. Leisure Activity ............................................................ 187
  6.3.6. Application Type ............................................................... 188
     6.3.6.1. Individual App Analysis .............................................. 188
  6.3.7. Summary of Main Study Qualitative Findings ..................... 192
6.4. Non-Adopter Study .................................................................... 194
  6.4.1. Method ............................................................................... 194
  6.4.2. Participants .......................................................................... 195
  6.4.3. Quantitative Results ............................................................. 195
  6.4.4. Qualitative Results ............................................................... 196
     6.4.4.1. Quality (Device and Applications) ................................ 198
     6.4.4.2. Motivation and Barriers .............................................. 200
     6.4.4.3. Positive and Negative Affect ...................................... 202
     6.4.4.4. Interactivity ............................................................... 203
     6.4.4.5. Learning and Leisure Activity ................................... 203
  6.4.5. Non-Adopter Verbal Ratings Results ................................. 204
  6.4.6. Brand and UserDisposition ............................................... 206
  6.4.7. Summary of Non-Adopter Results ..................................... 207
# Table of Contents

6.5. Chapter Summary ........................................................................................................209

7  **CHAPTER 7 Main Discussion** .................................................................................... 212

7.1. Introduction .................................................................................................................. 213

7.2. Summary of Main Findings .......................................................................................... 213

7.2.1. Main Contributions ................................................................................................... 213

7.2.2. Research Constraints ............................................................................................... 215

7.3. Expanded Model ......................................................................................................... 217

7.3.1. Product Qualities – Interactivity ............................................................................ 219

7.3.2. User Profile – User Groups .................................................................................... 220

7.3.3. Long Term Experience – UX Over Time ................................................................. 220

7.4. Main Discussion ......................................................................................................... 222

7.4.1. Interactivity ............................................................................................................ 222

7.4.2. User Groups .......................................................................................................... 227

7.4.3. UX Over Time & Product Adoption ....................................................................... 230

7.4.4. Methodological Insights ....................................................................................... 235

7.5. Heuristics .................................................................................................................... 238

7.6. Conclusion and Future Research ................................................................................ 241

8  **BIBLIOGRAPHY** ....................................................................................................... 244

9  **APPENDICES** ............................................................................................................. 262

Total Word Count: 80,471
List of Tables

Table 1: Summary of the three website features selected for the study ........................................... 63
Table 2: Summary of the three website features selected for the study ........................................... 66
Table 3: The Affect Scale individual items and origins ........................................................................ 69
Table 4: The Website Quality sub-scale items and origins ............................................................... 69
Table 5: The Immersion/Presence Scale questions and origins ......................................................... 70
Table 6: The Usability Scale questions and origins ........................................................................... 71
Table 7: Internal reliability scores for questionnaire scales using Cronbach coefficient ........... 72
Table 8: Component matrix table for Affect Scale factor analysis across the three activities... 72
Table 9: Post Hoc Affect Scale means for site and task ................................................................... 73
Table 10: The sig. results, mean (M) & standard deviation (SD) for Brand Awareness ........... 77
Table 11: Website Quality Scales showing significant main effects and interactions .............. 79
Table 12: Website Quality rank order means & post hoc sig. results (PQ & HQS) .................... 80
Table 13: Post Hoc results show sig. differences between sites for 6 items within the Website Quality Scale, where 4 mean patterns are revealed ................................................................. 81
Table 14: Mean rank order scores for Overall Experience Scale for the 3 sites ....................... 84
Table 15: The regression analysis results for the three sites (pre & post interaction) ............... 86
Table 16: The general to specific (inference) model regression analysis results for 3 sites .... 87
Table 17: Mean scores and ranking order of websites for Criteria Importance ................................. 89
Table 18: Summary of rank order across quantitative measures ...................................................... 90
Table 19: The 8 key main themes and 17 sub-themes that formed the coding framework .... 94
Table 20: NV's for the main Website Feature themes by site ............................................................... 95
Table 21: NV's for the Immersion/Presence themes by site ............................................................... 95
Table 22: NV's for the Positive and Negative Affect comments ...................................................... 96
Table 23: NV's for the General Impression themes by site ............................................................... 98
Table 24: NV's for the General Impression themes by site ............................................................... 99
Table 25: Cluster group means for three sites: National, Google and Louvre .......................... 101
Table 26: Summary of mean ranking & frequency of quantitative and qualitative results .... 104
Table 27: Repeated measures design (2x2) and (1x2) ................................................................. 111
Table 28: Summary of all measurement scales used within Study Two ....................................... 114
Table 29: The Power User Scale from Marathe & Sundar, (2008) ........................................114
Table 30: Cronbach Alpha (α) results for all the main scales ........................................115
Table 31: Three cluster groups from cluster analysis of Power User groups ....................120
Table 32: Repeated measures (2x2x3) mixed design for Affect and Website Quality ........121
Table 33: Table 33. Mean ratings for Criteria Importance items for both sites .................123
Table 34: Summary of the regression analysis results for the three sites (pre & post interaction) ........................................................................................................................124
Table 35: Summary of the general to specific (inference) model regression analysis results for 3 sites ................................................................................................................................................125
Table 36: Summary of the significant effects of the quantitative results .........................126
Table 37: NV's for the Main Website Feature themes by site ............................................127
Table 38: Frequency and net valency of comments for the Interactive Features for the interactive site ........................................................................................................................................128
Table 39: Summary of all measurement scales used within the longitudinal study three ....138
Table 40: Reliability results for all scales using Cronbach Alpha coefficient ....................141
Table 41: Results for the significant scale items from the main questionnaire across each 3 time frames ........................................................................................................................................143
Table 42: Multiple Regression results for 5 Dependant Variables analysis across three time frames ........................................................................................................................................149
Table 43: Summary of the general to specific (inference) model regression analysis results for 3 time frames ........................................................................................................................................150
Table 44: Summary of the sig. results across all scales showing change over time ..........151
Table 45: K-Means for the three Power Users clusters ......................................................153
Table 46: Summary of the between groups (PU) differences and changes over time ..........156
Table 47: Summary of the key applications used prior to conducting the study ...............157
Table 48: Summary of the most prominent Clinical/Medical applications ........................158
Table 49: Means for 6 Quality Ratings for 11 key applications ........................................160
Table 50: Mean ratings for 7 most used Medical Apps after 5 months of usage ...............163
Table 51: Summary of the rank order of apps across the 4 measurement scales ..............165
Table 52: Summary of the main questionnaire quantitative findings ................................166
Table 53: Total frequency of comments comparing Interview and Diary-Prompts ............174
Table 54: The NV and % frequency of Quality (Device & Apps) comments for interview and diary-prompts .......................................................................................................................................175
List of Tables

Table 55: The NV and % frequency of the 6 Quality themes split between the Device & Prompts ...........................................................................................................................................................................176

Table 56: The Motivator and Barrier sub-themes results split between Device and Apps......179

Table 57: The Positive and Negative Affect sub-themes results split between Device and Apps ...........................................................................................................................................................................182

Table 58: Interactivity sub-themes results split between Interview/Prompts and Device/App ...........................................................................................................................................................................184

Table 59: Work & Leisure Activity themes results split between interview & prompts..........185

Table 60: Summary of App results grouped into App Types.................................................188

Table 61: The individual App Ranking scores according to the aggregated (Interview & Diary) results...........................................................................................................................................................................189

Table 62: Summary of the main qualitative findings across all themes ................................192

Table 63: Independent T-Test and Mann-Whitney U Test results comparing Adopters with Non-Adopter...........................................................................................................................................................................195

Table 64: Summary of all Themes comparing the Actual and Normalised results ............197

Table 65: The 6 Quality Rating themes comparing Actual & Normalised results, for Device and Apps...........................................................................................................................................................................198

Table 66: The four Motivator sub-theme comparing Adopter and Non-Adopter results.....200

Table 67: The four Barrier sub-theme comparing Adopter and Non-Adopter results........200

Table 68: The Affective (Positive/Negative) themes comparing Non-Adopters and Adopters results...........................................................................................................................................................................202

Table 69: The Interactivity sub-themes comparing Adopters and Non-Adopters results.....203

Table 70: The Activity (Work and Leisure) themes comparing Adopters and Non-Adopters results...........................................................................................................................................................................203

Table 71: The Verbal Mean ratings comparing Non-Adopter with Adopter, showing % difference...........................................................................................................................................................................205

Table 72: Means and Independent T-Test and Mann-Whitney U-Test Sig Results.........206

Table 73: Summary of the Main Qualitative themes comparing Non-Adopters with Adopters results...........................................................................................................................................................................207

Table 74: A summary of the main contributions from this thesis....................................214
List of Figures

Figure 1: The expanded 3-stage cognitive process model, adapted from Hartman et al., (2008) ................................................................................................................................. 25
Figure 2: Outline structure of this thesis report .......................................................................................................................................................................................... 28
Figure 3: An overview of the main sections of this literature review ........................................................................................................................................ 33
Figure 4: One of the first definitions of usability (Eason, 1984) ........................................................................................................................................... 34
Figure 5: The two dimensions of core affect (Russell, 2003) ................................................................................................................................................ 37
Figure 6: Hassenzahl’s, (2004) summary of the relationships between attribute groups .... 41
Figure 7: UX Model proposed by Porat et al.,(2007); Porat & Tractinsky, (2012) ....... 45
Figure 8: Hassenzahl’s (2004) Inference Model (van Schaik, Hassenzahl, & Ling, 2012) .... 45
Figure 9: The CUE-Model (Components of User Experience) from Mahlke & Thüring, (2007) .......................................................... 47
Figure 10: The initial affective processing model by Lindgaard et al., (2011) .............. 48
Figure 11: The initial Adaptive Decision Making Framework by Hartmann et al., (2008) ..... 48
Figure 12: The exploratory sequential mixed methods approach used for the first two studies .................................................................................................................................................. 57
Figure 13: The multiphase mixed design used for the longitudinal third study .......... 58
Figure 14: The three interactive websites used within the study .................................................................................................................................................. 65
Figure 15: Overview of the study method and scales used ........................................... 66
Figure 16: Aggregated Affect Scale for Site (3) & Task (3), showing significance differences. 73
Figure 17: Aggregated Mean for the Individual Affect Scale items ................................ 74
Figure 18: Mean pattern for individual items Mood, Fun and Exciting (from Affect Scale).... 75
Figure 19: Mean patterns for individual items Absorbing & Curiosity (from Affect Scale) ..... 75
Figure 20: Means for single items Attractive, Interesting, Pleasurable & Engaging (from Affect Scale) ............................................................................................................... 76
Figure 21: Mean patterns for Hedonic Quality Stimulation (HQS) & Pragmatic Quality (PQ). 79
Figure 22: Mean Pattern One: shows all ratings increased to the same level post-interaction .................................................................................................................... 81
Figure 23: Mean Pattern Two: Shows Google and Louvre ratings increase far greater than National for Creative and Innovative items ....................................................................................... 81
Figure 24: Mean Pattern Three: Shows Google receiving higher spaciousness ratings than National & Louvre ........................................................................................................................................ 82
Figure 25: Mean Pattern Four: Shows National being rated higher for Clear Structure and Predictable, and Google being rated more unpredictable post-interaction ................................................................................................................. 83
Figure 26: Overall Percentage score from Forced Rank choice .................................................................................................................................................. 85
List of Figures

Figure 27: Summary of the regression results from the specific to general, and general to specific (inference) model perspectives ................................................................. 87
Figure 28: Mean Patterns for Engaging Interaction & Aesthetic Importance .................................. 88
Figure 29: Mean Patterns of Time Exploration ............................................................................. 89
Figure 30: The qualitative data analysis method........................................................................... 91
Figure 31: Frequency of the first two most recalled Items............................................................ 92
Figure 32: NV scores for the two most remembered aspects across the three sites ................. 93
Figure 33: Website Feature frequency of Positive & Negative comments across the 3 sites... 94
Figure 34: Immersion/Presence frequency of comments (pos/neg) across the 3 sites............ 97
Figure 35: Frequency of Positive & Negative Affective comments over the three sites......... 99
Figure 36: The structure of the two bespoke websites (interactive & standard).......................108
Figure 37: The two bespoke website homepages.........................................................................108
Figure 38: The two different sizes and positions of the avatar .................................................109
Figure 39: The interactive map showing hotlinks when clicked it reveals a picture.............. 110
Figure 40: Procedure method for Study Two showing methods and measures .................... 111
Figure 41: Affect Mean ratings for website (x2) and task (x2)................................................. 115
Figure 42: Aggregated means for single item affect responses comparing the 2 sites ......... 116
Figure 43: Hedonic Scale (HQI & HQS) mean ratings for website (x2) and task (x2)......... 117
Figure 44: Pragmatic Scale (PQ) mean ratings for website (x2) and task (x2)...................... 117
Figure 45: Single item Attention ratings for website (x2) and task (x2)................................. 118
Figure 46: Single item Beauty mean ratings for website (x2) and task (x2)......................... 118
Figure 47: Scree plot showing the coefficient change against the number of clusters ........ 120
Figure 48: User group analysis for Attractive (from Affect scale)............................................. 121
Figure 49: User group analysis for PQ (Pragmatic Scale)....................................................... 122
Figure 50: User group analysis for Enthusiasm & Happy (+PANAS)........................................ 122
Figure 51: Summary of the regression results from the specific to general, and general to specific (inference) model perspectives ................................................................. 125
Figure 52: Plan of the study methods over six months.............................................................. 136
Figure 53: An overview of the structure for the three interviews (Early, Mid, Final)........... 137
Figure 54: Previous device ownership (Computer and Mobile Phone)................................. 140
Figure 55: Mean scores for Positive Affect, and the single item Frustrating (from Negative Affect)..................................................................................................................... 142
Figure 56: Mean ratings for Pragmatic Quality (PQ), Hedonics (HQ) and Beauty.................. 142
Figure 57: Mean ratings for Usefulness, and single item Motivated...................................... 143
Figure 58: The Mean Ratings for Learning & Leisure Activity over 24 weeks...................... 145
Figure 59: The mean ratings for Negative Affect and Overall Satisfaction............................ 146
List of Figures

Figure 60: Mean ratings for Ease-of-Use, Attractive & Overall Satisfaction (Verbal Scale) .....147
Figure 61: Summary of the regression models: Specific to General & General to Specific (inference) model perspectives, showing changes over time ..................................................150
Figure 62: Mean ratings for single item Frustration (Neg. Affect) and Hedonics (HQ) ..........154
Figure 63: Mean ratings for Pragmatic Quality (PQ) and Usefulness ..................................155
Figure 64: Mean ratings for Interview Verbal Scale Item Ease-Of-Use ................................156
Figure 65: Key Applications (%) that were downloaded and used ......................................159
Figure 66: How often the 8 main Apps were used (comparing Mid to Final-Study results) ....159
Figure 67: Overall mean ratings for all 11 Apps (comparing Mid to Final-Study results) .....161
Figure 68: The 15 most Popular Medical Apps (out of 44) after 3 months (Mid-Study) ....162
Figure 69: How often the 6 most popular medical apps were used (Final-Study) ...............163
Figure 70: The 11 most recalled apps during the Mid & Final-Study Interviews .................164
Figure 71: An overview of the three longitudinal interviews .............................................171
Figure 72: An individual participant’s timeline graph used as a prompt in the final-interview ........................................................................................................................................172
Figure 73: Two common patterns in the individual participants’ time-line graphs ................173
Figure 74: The qualitative data analysis method for both interview and diary-prompt data ...173
Figure 75: Mean scores comparing Non-Adopter with Adopter questionnaire results .......196
Figure 76: Graphs comparing Non-Adopters with Adopters Verbal Ratings Expected Use with Actual Use ........................................................................................................196
Figure 77: Summary Model of Adopters and Non-Adopters qualitative results ...............206
Figure 78: The expanded model of users’ decision-making process during UX ...............218
Figure 79: The transitional effect of UX judgement criteria over time ................................221
Figure 80: An overview of the five main contributions of this thesis .................................222
Figure 81: Summary model of main reasons for Adoption or Non-Adoption of the iPad ....234
Understanding the interactive experience of using digital technologies is a complex process. Traditional methods of evaluating interactive technologies originate from usability, which focuses on ease of use, ease of learning and performance. User Experience (UX) emerged from the recognition that usability alone does not account for the more subjective emotional responses experienced when interacting with a product. Although the term UX has become widely accepted within the area of Human Computer Interaction (HCI), its definition still remains unclear, making it difficult to evaluate and design for. This thesis adopts a hybrid perspective by bridging the division between the reductionist and holistic approaches to UX research. Using a multi-methods approach that combine the strengths of both quantitative (objective) and qualitative (subjective) methods, will provide deeper insights into the users’ judgement process of interactive products.

Various theories have been proposed to understand UX, yet no consensual UX theory or model has emerged. The importance of aesthetics in influencing decisions about a products quality gained much attention in early UX research with conflicting results, sparking a surge of research into understanding the complexities of user quality judgement. Past UX research has focused on the multi-constructs of pragmatics, hedonics and aesthetics, and how these may influence user judgement, which can vary depending on the context, task and user background. However, little attention has been given to the impact of interactive design features upon UX. Findings from this thesis clearly show that interactivity is an important element within UX in both short and long-term usage.

This thesis expands the existing process model of user quality judgement, through a series of three studies to reveal the importance of interactivity, and how initial perception and judgement of a products quality can change over time. The first two studies identify the importance of interactivity in positive influencing UX. Both studies revealed that affective and hedonic ratings increased as a result of interaction, demonstrating the powerful effect of interaction, and showed clear differences for websites that contained enhanced interactive features, despite the presence of usability problems. Further exploration using cluster analysis revealed three sub-groups that categorised users not only by their interactive style preferences, but also by their predispositions towards technology. This perspective of user sub-group analysis is a contribution to the field which bridges population-level quantitative analysis with qualitative findings that focuses on the individual ethnographic interpretations of experience.

Considerable UX research has focused on short-term evaluations, based on users first impressions pre and post-interaction, with few studies capturing long-term usage. The third study reports on an ecological longitudinal investigation into how UX changes over time and long-term product use. A group of novice iPad users were tracked over six months to reveal that despite poor usability, hedonic ratings remained high, yet over time usefulness and utility were dominating factors affecting UX and product adoption. The influence of both device and app revealed that although users found the device more pleasurable, it was the variety of apps contained on the device that facilitated positive UX.

The overall findings from this research provided some valuable methodological insights and aided the creation of set of practical UX heuristics that can be used to inform both future research and design practice.
Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree of qualification of this or any other university or other institute of learning.
Copyright Statement

The following four notes on copyright and the ownership of intellectual property rights must be included as written below:

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) University’s policy on Presentation of Thesis.
Statement of Contribution

This disclaimer is to state that the research reported in this thesis is primarily the work of the author and was undertaken as part of her doctoral research. Referenced papers of which the student is not the sole author represent the role of the supervisors in the research, to direct the work and enhance the written style of these papers.

The work reported in Chapters 3 and 4 has been published as follows. The content of these papers have been re-interpreted and rewritten in the thesis.


Acknowledgements

I would like to express my gratitude to Professor Alistair Sutcliffe, who has been an excellent supervisor, providing me with continual support and invaluable guidance throughout the PhD process. I also thank Professor Antonella De Angeli for her valuable advice, especially in the early stages of the thesis.

I am also very grateful to the Manchester Medical School for giving me the opportunity to conduct my third longitudinal study. In particular I thank Tim Cappelli, Hilary Dexter and Lucie Byrne-Davis for their advice, support and interest in my research.

This thesis would not be possible without the participants who took part in the three studies, so I would like to thank all of them for their time and valuable contribution.

For financing this research, I am grateful to the Engineering and Physical Science Research Council (EPSRC) studentship.

Finally, I wish to thank my somewhat neglected family and friends for being so understanding. But the biggest thanks goes to my patient companion, Steve Hooper, for his constant support and encouragement from beginning to end.
Chapter 1  Introduction

This chapter first presents a synopsis of the research area in which this thesis is grounded. It then outlines the rationale to the research undertaken and then frames the main research aims and objectives. Three broad research questions are defined that direct the research, and the chapter concludes with a brief overview of the thesis structure.
1.1. Introduction

The computer technology landscape has changed considerably over the last 30 years and has permeated into almost every aspect of our lives. As computers have become more powerful and ubiquitous, consequently the way in which people experience technology has dramatically changed. With the advent of new digital technology, both academic researchers and industry practitioners alike are continually seeking for a richer understanding of how best to approach the design of new technology in order to improve our experiences during interaction.

User experience (UX) is not new, it was over two decades ago that Carroll and Thomas (1988) explored the relationship between ‘ease of use’ and ‘fun’, and challenged the HCI community calling for a scientific study of fun. There was a blurring of boundaries as technology moved out of the workplace to pervade everyday life, creating a dichotomy between technology used as ‘tools’ for work, and those used as ‘toys’ for leisure (Igbaria, Schiffman, & Wieckowski, 1994; Monk, Hassenzahl, Blythe, & Reed, 2002). As a consequence, the concept of UX evolved out of concerns that traditional usability used to evaluate the more instrumental task-orientated aspects of work experience, did not account for the more emotional hedonic qualities of use. This brought a shift in attention, away from the objective aspects of product (e.g., function, content, presentation), to encompass the more subjective view of users’ feelings, emotions, attitudes and motivations. Due to this paradigm shift, UX was slow to develop as a research area, as subjective opinions of users appeared too fuzzy to be considered credible (Monk et al., 2002), raising doubts about the measurability of UX.

However, the seminal paper entitled ‘What is beautiful is usable’ by Tractinsky, Katz, & Ikar, (2000), raised the importance of aesthetics in influencing decisions about a products quality, and sparked a surge in UX research that explored a whole range of aspects including pleasure (Jordan, 2000), joy of use (Hassenzahl, Beu, & Burmester, 2001), hedonic quality (Hassenzahl, 2001), attractiveness (Norman, 2002), and aesthetic beauty (Hassenzahl & Monk, 2010; Lavie & Tractinsky, 2004).

1.1.1. Conflicting Perspectives to UX

Since the 1990s there have been a wealth of frameworks, models and theoretical approaches produced under the umbrella term of UX, but still no one consensual theory has emerged (Law, van Schaik, & Roto, 2014; Law & van Schaik, 2010). This is partly due to the deep-rooted controversy regarding the theoretical stance of UX manifest in a tension between the two differing perspectives; the model-based reductionist approach, with roots in cognitive psychology, and the design-based holistic approach that is grounded in pragmatist philosophy (Blythe, Hassenzahl, Law, & Vermeeren, 2007; Law, 2011). The reductionists
typically use quantitative psychometric measures such as Likert or semantic differential scales (Likert, 1932; Osgood & Tannenbaum, 1957) to measure various quality constructs. They consider that some experiences are “far less unique and far less variable”, so evaluative judgements of users can provide generic insights into UX (Hassenzahl, 2008b, P. 14).

Advocates of the holistic paradigm propose situated qualitative techniques such as narrative storytelling (Battarbee, 2003), experience prototypes (Buchenau & Suri, 2000) or cultural probes (Gaver, Dunne, & Pacenti, 1999), that capture a user’s experiences while interacting with a product. These opposing approaches serve two purposes, holism provides a greater understanding of the context that enhances design inspiration, while the reductionists enables product evaluation by assessing the value of the design to provide models and guidelines in order to improve design (Hassenzahl, 2008b). There have been recent calls to reconcile the two conflicting approaches, and to develop a hybrid approach to UX (Bargas-avila & Hornbæk, 2011). This research aims to bridge both these approaches by using quantitative measures to evaluate various constructs, and qualitative methods to provide rich situated insights into UX.

1.1.2. UX Frameworks and Models

The field of UX research is entangled with many diverse theoretical models with different approaches, which lack an in-depth discourse within the UX community to define the foundations of UX theories (Obrist et al., 2011). Some of the first theoretical frameworks originated from the design-based holistic approach, with conceptualisations that aimed to assist designers in understanding the nature of experience (Battarbee, 2003; Forlizzi & Ford, 2000). Mostly based on Dewey’s pragmatist theory of experience, which posits experience is created by the individual and object, situated in the environment. This was exemplified by McCarthy & Wright, (2004) notion of ‘felt experience’, which placed emphasis on the contextual and uniqueness of experience. In contrast, the reductionist measurement models share a similar paradigm with usability evaluation and Technology Acceptance Models (TAM) (Venkatesh, Morris, Davis, & Davis, 2003), by manipulating various quality constructs in order to test causal relationships. Various well cited studies have found that generic constructs such as classical or expressive aesthetics (Tractinsky, 1997), hedonics and pragmatics (Hassenzahl, 2004; Tractinsky, 1997), can influence user judgements of the overall quality of technology products.

One of the challenges is how to select the appropriate UX measures to evaluate and design for interactive experiences, especially when there is still no unified definition of UX (Law & van Schaik, 2010). Despite many attempts at defining UX, (for a review see: Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009), including a formal definition, ISO 9241-110, (2010),
which appears somewhat abstract (see Section 2.6), the concept of UX is still somewhat elusive. A vague description gained from a survey of 275 UX practitioners and researchers conducted by Law, Roto, Hassenzahl, Vermeeren, & Kort, (2009, p.727), suggest that UX is “dynamic, context-dependant and subjective”. Although this is a start, it is far from providing a solid base to work from, and researchers agree that the ingredients of UX require further definition, and the diversity of UX measures require greater clarification (Bargas-avila & Hornbæk, 2011; Law et al., 2009, 2014).

1.1.3. User Quality Judgement Models

UX models are useful for investigating the relationships of multiple constructs, with early models focusing on the associations between perceived aesthetics, usability, affect and pleasure (Porat & Tractinsky, 2012; Tractinsky, 1997). Hassenzahl’s, (2004) inference model of UX was influential in drawing attention to the relative contribution that the general constructs of hedonics and pragmatics make to users’ overall judgement of an interactive product. The inference premise is that users infer unavailable attributes (actual usability experience) from available attributes (perceived aesthetics) to infer the overall quality of a product (goodness). Hassenzahl & Monk, (2010) demonstrated that evaluations of these constructs remained stable across four different product types, so avoiding the fixed fallacy problem (Monk, 2004), thus suggesting that user judgement of UX remains consistent across product type.

Alternative models have focused on the cognitive process and are used to evaluate changes in perception and user judgement over time. Initially, these models focused mainly on the early stages of the process, when users form initial impressions based on the aesthetics and design qualities of the product (Lindgaard, Fernandes, Dudek, & Brown, 2006). Response time for aesthetic judgements was found to be much faster than for usability judgements, as the latter involved ‘cognitive reflection’. So the visual appeal of a product (website) can influence a users’ judgement, and enhance their motivation to use the product or not (Lindgaard, Dudek, Sen, Sumegi, & Noonan, 2011). This corresponds to the inference models, where initial perceptions of aesthetics and anticipated usability could influence other unaccessible qualities (actual usability or goodness) (Hassenzahl & Monk, 2010; Hassenzahl, 2004); although, the directional relationship between usability and aesthetics is still unclear, due to the various studies using differing construct manipulations and measurement scales (Tuch, Roth, Hornbæk, Opwis, & Bargas-Avila, 2012). Hartmann, Sutcliffe, & De Angeli, (2007) showed that users’ overall preferences for websites that had similar content and design, can be swayed by framing effects, in which user judgement of aesthetics and overall preference varied according to the task. While trade-off effects were found between beauty and usability.
by Diefenbach & Hassenzahl, (2009), where beauty is valued, people discount it in situations that do not allow easy justification (cost vs beauty). Similar trade-off effects were also found between hedonic and pragmatic attributes in a further study by (Diefenbach & Hassenzahl, 2011).

Theoretical models are important tools for communicating ideas at how best to evaluate and design for UX, and there is a need to improve the current models to gain a deeper understanding of UX. This research uses a process-model of user judgement (Hartmann, Sutcliffe, & De Angeli, 2008) as a starting point, and aims to expand the existing model by investigating the influence of interactivity upon UX, and how this may impact on the various quality ratings of an interactive product.

1.1.4. Interactivity

Much less research has focused on summative evaluations of interactive products, with less attention to the interactive features contained within the product. Similar to UX, interactivity is a widely used term, but the concept is somewhat vague. There are many differing models of interactivity (Hoffman & Novak, 1996; Kristof & Satran, 1995; Lee, 2005; Wu, 2006), but what benefits are gained from the various interactive features, and how they may influence UX remains unclear. HCI research has predominantly focused on the use of interactive features to enhance UX within the games and entertainment field, with emphasis on immersion (Jennett et al., 2008; Korhonen, Montola, & Arrasvuori, 2009; Sanders & Cairns, 2010; Takatalo, et al., 2007), flow and presence (Qiu & Benbasat, 2005; Trevino & Webster, 1992), and playfulness (Korhonen et al., 2009). There is also a growing area of research in the use of interactive multi-media technology within cultural spaces and museums to enhance user engagement (Haywood & Cairns, 2006; Othman, Petrie, & Power, 2011), and an increasing interest in serious games and gamification, that harnesses the principles of playful design with interactive technology (Clark, 1987; Deterding, Dixon, Khaled, & Nacke, 2011). The few experimental studies on the effect of interactivity within e-commerce websites have shown that it can positively influence user satisfaction, enjoyment and motivation (Cyr, Head, & Ivanov, 2009; O’Brien, 2010; Teo, Oh, Liu, & Wei, 2003). However, these studies have explored only a limited range of constructs, so the influence of UX constructs such as aesthetics, emotion and usability in relation to specific interactive features (such as 3D environments, or avatars) remains unclear.

Interactivity doesn’t just happen; it has to be designed for. Designers need to know how best to use interactive features, and how they can be used to influence UX. Nielsen, (1999) argued that different multimedia technologies provide numerous design options, which required
Chapter 1  Introduction

constrained use to avoid confusion. Despite the existence of formal guidelines aimed to avoid these problems (ISO 14915, 2002) they provided little advice when designing for user experience and aesthetics (Sutcliffe, Kurniawan, & Shin, 2006). Few studies have considered how different design features may influence user judgement of their experience. Kim, Lee, & Choi, (2003) identified elements of aesthetic design features within a website (e.g., hue, brightness, shape, texture etc.) and were able to link them to 13 generic dimensions of emotion, although these are more akin to aesthetic perceptions (e.g., strong and powerful, calm and balanced, classical and conventional), rather than emotional responses. Some formal design guidelines can be found in the design community (Galitz, 2007; Kristof & Satran, 1995), but they focus mainly on general layout, style and colour suggestions, which is only a small aspect in understanding the influence of interactive design features (Lidwell, Holden, & Butler, 2010). Frameworks of multimedia design have also been provided (Heller, Martin, Haneef, & Gievska-Krliu, 2001), along with methods to support the appropriate selection of interactive media in order to direct attention and enhance user engagement (Sutcliffe et al., 2006), however very little attention has been given to the influence of interactive design features on UX.

In addition to interactivity, this thesis aims to expand its view beyond the additional factors of context and task which have been shown to influence UX (Hartmann et al., 2008; Hassenzahl & Tractinsky, 2006). There is limited research in understanding how individuals differ in their perceptions of product quality, and how these judgements may change over time.

1.1.5. Individual Difference

Although user diversity is recognised in HCI research, the idiosyncratic nature of users’ subjective judgement of interactive products is very much underestimated by researchers and developers. Current approaches distinguish users in terms of generic groups, such as age, gender, education etc., and do not reflect users’ individual cognitive differences (Karapanos & Martens, 2007). Different people respond differently to a given product, as experience is partly dependant on the characteristics (motives, preferences, skills, goals etc.) the user brings into the interaction (Desmet & Hekkert, 2007). A few studies have found that culture differences can influence mobile phone preference of icon recognition (Kim & Lee, 2005), and individual differences in perception of design quality was shown to be influenced by user background (Bloch, Brunel, & Arnold, 2003; Hartmann et al., 2008). New methods are needed to investigate the finer granularity of individual differences and avoid generic findings at the population level.
1.1.6. Dynamics of UX

Past research has focused mainly on short-term evaluations, when a user first perceives an interactive product, rather than long-term usage. The individual perceptions of a product’s quality can change over time, with a few studies showing changes in usability, hedonics and aesthetics perceptions (Karapanos, Hassenzahl, & Martens, 2008; von Wilamowitz-Moellendorff, Hassenzahl, & Platz, 2006). The dynamics of UX has been repeatedly noted as being an important area of research, but there are few longitudinal studies that have focused on long-term usage, beyond one month (Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, & Sinnelä, 2011; Vermeeren, Law, & Roto, 2010), and most of these studies used very few participants. Although there is a large body of longitudinal research within TAM, the technology acceptance field (Karahanna, Straug, & Chervany, 2013; Venkatesh & Brown, 2001), there are few similar studies within UX. Some frameworks have been proposed that describe various experiential phases of experience from anticipation to refection (Karapanos, Zimmerman, Forlizzi, & Martens, 2009; Kujala, Vogel, Pohlmeyer, & Obrist, 2013), which give interesting conceptualisations of the dynamics of UX, but do not provide insights into the dynamic changes of user judgement and how various interactive features may contribute or hinder user satisfaction and acceptance.

1.2. Aims and Objectives

The starting point for this research is the three-stage cognitive process model of user judgement for quality proposed by Hartmann et al., (2008), and Sutcliffe, (2009), as shown in Figure 1. It is based on the Adaptive Decision Making Theory (ADM) from Payne, Bettman, & Johnson, (1993) who assert that “people’s decision making is adaptive and contingent upon task, context and their background-experience” (Hartmann et al., 2008, P. 388).

![Figure 1. The expanded 3-stage cognitive process model, adapted from Hartmann et al., (2008)](image-url)
Chapter 1  Introduction

The model proposes that peoples’ decision making is adapted by the task-in-hand (user goals), which depends on the user’s profile or background (experience), and the context (serious vs. not serious). User judgement will also depend on the prioritisation of the product qualities (usability, utility, aesthetics etc.) by which the product is judged.

This research aims to expand the cognitive model in two ways, first by exploring additional constructs that are drawn from psychology, and then to investigate how these constructs may change over prolonged use and time. The principles of affect (mood and emotions), flow (immersion) and presence (involvement) are well researched in other fields (Berlyne, 1960; Csikszentmihalyi, 1990; Witmer & Singer, Michael, 1998), yet how these influence UX is unknown. Key design features that may promote engaging interaction will be evaluated in relation to these constructs. It is expected that designs that promote high affect, flow and presence (such as interactive features, avatars and virtual environments) will generate positive experiences. The influence of task, context and user characteristics on perceptions of UX are also considered, along with investigating the factors that influence long-term user satisfaction and technology acceptance.

This thesis adopts a hybrid approach that aims to bridge the division between the reductionist and holistic approaches to UX evaluation. It uses a mixture of methods that combine quantitative measures (reductionist) with qualitative narratives (holistic) to evaluate both the objective and subjective interpretations on UX. It has been argued that traditional HCI research methods are too abstract and theoretical, as they do not support UX design practice (Kujala et al., 2011). The second aim of this research is the challenge to develop new investigative research methods that can capture UX in the field and then track how these may change over time. These research methods will aid the creation of practical UX evaluation principles. The final aim of this research is to produce simple UX design heuristics that can assist both the research and design community.

To summarise the main objectives of this research are:

- To extend the cognitive process model (Hartmann et al., 2008) by investigating various constructs (e.g., affect, flow and presence) during user interaction and how these change over a prolonged time frame.

- To develop appropriate UX research methods that can be used to access interactive experiences.

- To produce design heuristics informed by experiments and a longitudinal study on how interactive design features elicit affective responses to achieve positive UX.
1.3. Research Questions

Three main research questions are identified, which were used to guide the research:

- **What are the key constructs that contribute to user engagement and user experience?**
  
  Previous studies have demonstrated that aesthetics, usability, content and brand influence user judgement of interactive technologies, which is dependant on context, task framing and user background (De Angeli, Hartmann, & Sutcliffe, 2009; Hartmann et al., 2008). The importance of hedonics (emotion), pragmatics (usability) and beauty are found to influence users perception of the products goodness (Hassenzahl, 2004), however the relationship between aesthetics and usability still remains unclear (Tuch et al., 2012). To understand what attracts people to use, and continue to use interactive technologies, these constructs will be explored in relation to their affective responses prior and during interaction. The additional measures of affect, flow and presence will be used to explore user engagement with different types of interactive design features (see below).

- **Does enhanced interactivity (design features) lead to more positive user experience?**
  
  The positive effects of interactivity has been investigated in other fields, such as technology acceptance (TAM), (Venkatesh et al., 2003) and e-commerce (Cyr, Head, Larios, & Pan, 2009; O’Brien, 2010); however, there are few studies that have explored its influence on UX quality judgement measures (e.g., usability, aesthetics etc.). The influence of various interactive design features (e.g., 3D environments, avatars) that may promote positive affect, immersion, flow or presence will also be explored in relation to other UX measures.

- **What are the key constructs that contribute to user satisfaction and product adoption over time?**
  
  There have been recent calls for more longitudinal studies that investigate UX over time and to propose new practices and methods in order to improve design for long-term product use (Kujala et al., 2011, 2013). Most UX research has focused on user judgement within session, few studies have captured long-term usage, in relation to the changing perceptions of a products quality constructs (aesthetics, usability, utility etc.). Users background (inclination for exploration) is known to influence users tendency towards product adoption (Magni, Susan Taylor, & Venkatesh, 2010), however little is known about what other aspects may impact on the adoption decision-making process over time.
1.4. Outline of Thesis

An outline of the seven chapters contained in this thesis is shown in Figure 2. Chapter 1 introduces the research area, presents the aims and objectives of the research, which is guided by the three research questions. Chapter 2 provides a detailed discussion of the main literature surrounding UX drawing from the HCI field and beyond. Chapters 3 and 4 report on two experimental studies, while Chapters 5 and 6 report on a six-month longitudinal study which is divided between the quantitative finding (Chapter 5) and the qualitative findings (Chapter 6). Chapter 7 provides a detailed discussion of the main contributions of this thesis in relation to the relevant literature in the field. It summarises the key findings generated from the three studies (Chapter 3 to 6) in relation to the original research question, and outlines how these findings impact on the original theoretical framework. The chapter concludes with a set of twelve heuristics and some recommendations for future research.

![Figure 2. Outline structure of this thesis report](image)

**Chapter Summaries**

**Chapter 2:** Begins with a review of the literature that lead to the evolution of UX, as traditional usability did not account for the subjective experiences interacting with technology (Carroll & Thomas, 1988). The surrounding literature of emotion, affect, aesthetics and hedonics is also reviewed, and the suggested definitions of UX are presented. An overview of the various conceptual frameworks is provided, which is grounded in the long-standing debate between reductionism and holism (Law, 2011). Three types of UX measurement models are reviewed, the multiple construct approach (Lavie & Tractinsky, 2004; Tractinsky, 1997), the inference models (Hassenzahl & Monk, 2010; Hassenzahl, 2004).
Chapter 1  Introduction

2004), and the cognitive process models (Lindgaard et al., 2011, 2006; Mahlke & Thüring, 2007), which introduces the cognitive model of quality judgement (Hartmann et al., 2008). The technology acceptance (TAM) (Davis, 1989; Venkatesh et al., 2003), process model is discussed in relation to product adoption and its potential use within UX research. Drawing from literature beyond HCI, a review of the psychological concepts of flow and presence is discussed in relation to user engagement and immersion (Jennett et al., 2008; Witmer & Singer, Michael, 1998). The literature on interactivity and relative frameworks are reviewed (Cyr, Head, & Ivanov, 2009; Teo et al., 2003), along with an overview of the limited research on individual difference (De Angeli, Sutcliffe, & Hartmann, 2006; Bloch et al., 2003), and longitudinal studies within UX (Karapanos et al., 2008; Kujala et al., 2011). The chapter concludes with a summary of the research methodology adopted within this research, the rationale to using a mixed methods approach and how this was applied across the three studies, along with short discussion of the current debate around measurement scales.

Chapter 3: The first empirical study investigates the experiences and preferences across three real-world art gallery websites. Varying interactive styles are offered within these three websites; with two sites employing enhanced interactive features (3D navigation and avatar guide), while the third site offered limited interactivity, so acted as a control to the study. The study employed a mixed-methods approach that used a combination of questionnaires and interviews, which compared pre and post-interaction experiences. The study tested the hypotheses that highly interactive websites provide positive user engagement, as measured by affect, flow and presence, and their influence on other quality measures (e.g., usability, aesthetics, hedonics) was also explored.

Chapter 4: The same research method from study one (Chapter 3) informed the design for this controlled experiment, that focuses on testing the key constructs of presence (avatar) and immersion (engagement), using a 'with and without' interactivity comparison. Two bespoke websites were developed that were aesthetically identical, yet one site contained an avatar guide and interactive media (video and interactive links), while the other did not. This second study tested the hypotheses that interactive features (avatar and interactive media) will show higher affect and immersion (engagement) ratings.

Chapter 5: This chapter reports on the quantitative findings from a six-month longitudinal study that investigated how UX changes over prolonged use and time. The study follows the changing experiences of a group (N=51) of third year medical students from when they received their first tablet computer (iPad), from their initial expectations, through to product adoption. The study aims to capture reasons for user satisfaction/dissatisfaction from (1) the user perspective (motivations, attitudes and experiences), (2) the device (usability, utility,
functionality), (3) the applications used (leisure/learning), (4) the education context (how it supports learning), (5) and how individual differences might influence user satisfaction and product adoption. A triangulation of methods is used within this longitudinal study, with this chapter reporting on the quantitative methods, analysis and findings.

**Chapter 6:** The qualitative methods used within the six-month longitudinal study are reported, along with the analysis and findings. It further reports on a small study that investigated reasons for non-adoption of the iPad, where a small participant group of non-adopters (who had owned the iPad for more than a year, but had stopped using it), were interviewed to elicit their reasons and motivations for non-adoption. The chapter concludes with an overall discussion comparing the adopter and non-adopter findings.

**Chapter 7:** This chapter presents an overview of the four main research contributions generated from the three studies reported in Chapters 3-6. These findings are reviewed in relation to the original research questions, and a number of research constraints are presented. The expanded cognitive model of quality judgment is presented, and then all four key contributions (interactivity, user groups, UX over time and methodology) are discussed alongside the relevant literature in the field. The chapter concludes with a set of twelve heuristics that can be used as a guide when evaluating, researching or developing new interactive technologies.
Chapter 2   Literature Review

This chapter reviews the literature surrounding UX, drawing from a wider area of research beyond HCI. It will outline the significant aspects that have shaped the development of UX, and will provide an overview of the main theoretical models used to aid the evaluation and design of interactive technologies. The wider influences affecting UX will also be discussed in relation to how they may affect user quality judgement.
2.1. Introduction

The term ‘user experience’ (UX) is well known by the Human Computer Interaction (HCI) academic community and with industry practitioners alike, and has been related to a wide variety of meanings, ranging from usability, aesthetics, emotions, affect and experiential value. Yet it has been regularly criticised for being too vague, obscure and ambiguous. The landscape of UX is fragmented with complicated models and theoretical approaches developed to help understand experience, with contributions from design, business, philosophy, psychology, cognitive and computer science (Forlizzi & Battarbee, 2004). However, there is still no cohesive theory by which to evaluate and design for the experiential qualities of technology. Although, still yet to be settled, many agree that UX is a dynamic, highly contextual-dependent and subjective account of human-technology interaction, which is heavily shaped by individual emotions (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009).

This chapter reviews the main literature surrounding User Experience (UX) by first outlining the main influences that prompted the advent of UX, drawing from a wider context beyond HCI. An overview of the main sections of the literature review is shown in Figure 3. This review begins by identifying the emerging literature from usability to user experience (2.2), motivated by the need to look beyond traditional usability, as it did not account for the subjective nature when interacting with technology. It discusses the emotional, affective and hedonic literature (2.3) that bought about huge change within HCI research, drawing from the wider areas of psychology, ergonomics, industrial design and marketing and consumer research. The rising importance of aesthetics (2.4) within product design is discussed, along with the conflicting views about its relative influence on usability, which resulted in the claim “what is beautiful is usable” (Tractinsky, Katz, & Ikar, 2000). The influential literature on Hedonics (Hassenzahl, 2004) outlines the importance of the subjective nature of pleasure, along with its relationship to pragmatics (usability), beauty and goodness, and how these all can influence user judgement (2.5). Due to the multidimensional aspect of UX, the problem with gaining a clear definition of UX is outlined (2.6), along with the most recent suggestions (Law et al., 2009). A short review of the UX theories, frameworks and models is given (2.7), which first outline the two contrasting approaches (holistic and reductionist), then follows a summary of three varying model types (construct, inference, process), which are compared to similar models within wider fields, (e.g., TAM). The nature of User Engagement (UE) is discussed in relation to UX, where additional constructs of flow and presence and interactivity are identified (2.8). In order to expand existing UX research, the area of individual difference is discussed, along with the growing area of research that is investigating UX over time (2.9).
2.2. From Usability to Experience

Early UX research argued that usability concepts were too focused on task efficiency, productivity and learnability, and a more boarder notion of quality was needed. The problem with usability was its focus on the instrumental aspects, which was highlighted by Gaver & Martin, (2000) who argued for a wider range of alternative list of needs in which to evaluate technology.

2.2.1. What is Usability?

“Usability of a system or equipment is the capability in human functional terms to be used easily and effectively by the specified range of users, given specified training and user support, to fulfill the specified range of tasks, within the specified range of environmental scenarios” (Shackel & Richardson, 1991, P.24)

Usability is a fundamental concept within HCI that has been debated for decades, both in terms of its definition and its use as a measurement tool (Dillon & Watson, 1996). It was first introduced in the 1980s to replace the term ‘user friendly’ by compelling designers to take a more user-centred approach and consider the users all the way through the design stages (Stevens, 1983). This was a turning point in computer system design, which traditionally focused on a top-down design approach, where all requirements were outlined in the planning phase. Usability evolved through various stages of definition, where initial
expressions of ‘ease of use’ (Miller, 1971), were characterised by Eason, (1984) using four components, the user characteristic’s, task characteristic, system function and the environment (see Figure 4).

![Diagram of usability components](image)

**Figure 4. One of the first definitions of usability (Eason, 1984)**

This was later expanded by Shackel & Richardson, (1991) who added four operational components (effectiveness, learnability, flexibility and attitude) to enable usability evaluation during use, which paved the way for Nielsen’s systematic evaluation methods (1993), consisting of various measurable attributes (e.g., learnability, efficiency, memorability, errors and satisfaction). The first formal definition arose in 1998 by the International Organisation for Standardization, (ISO 9241-11, 1998), defining usability as, “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” (P.2). This definition places emphasis on the measurable aspects (effectiveness, efficiency and satisfaction) that should be assessed in terms of context of use (task, user and situation).

### 2.2.2. The Problem with Usability

Both effectiveness and efficiency are the objective measures of usability, so more easily measured, while satisfaction is a more subjective measure. Effectiveness refers to the accuracy and completeness with which users achieve their specified goals, typically measured through learning and completion time, while efficiency is measured by error rates, task completion and the quality of the outcome of a given task. However, satisfaction relates to the users’ freedom from discomfort and their positive attitudes towards using the product, as measured through psychometric scales on overall preference, and general perceptions of product quality (Hornbæk & Law, 2007; Hornbæk, 2006). It was assumed that these three measures of usability strongly correlate, as outlined in a study by Nielson & Levy, (1994) who conducted a meta-analysis of 57 studies and found a strong association with users average task performance and average subjective satisfaction. However these findings have since
been questioned by a number of subsequent studies (Hornbæk & Law, 2007), that suggest subjective perceptions of usability differ from objective performance measures, and can depend on the users past computer experience and preferences (Bailey, 1993; Kissel, 1995). Despite the usability standards being ‘enhanced’ in (2001, ISO 9126) to include a further two characteristics (learnability and security), both ISO standards lacked harmony, as they were developed from two different perspectives (e.g., ISO 9241-11 originates from HCI experts, while ISO 9126 came from the Software Engineering field), making them difficult to interpret and therefore apply (Abran, Khelifi, & Sury, 2003).

Due to the lack of a consistent standard for usability and the inconsistencies between the objective measures of performance and the subjective assessments of user judgement, evaluating usability proved a challenge. However, a variety of practical approaches to evaluating usability were proposed, such as Heuristic Evaluation (Nielsen, 1994), cognitive walkthroughs (C. Lewis, Polson, Wharton, & Rieman, 1990), GOMs (Card, Newell, & Moran, 1983) and user testing methods (Agarwal & Venkatesh, 2002). These usability evaluation approaches assume that it is possible to identify usability to varying levels of granularity. Nielsen, (1993) offers a wide range of design principles or “heuristics” that are conducted by small groups of evaluators to determine the usability of websites (Nielsen & Molich, 1990; Nielsen, 1994). According to Nielsen (1993), usability comprised of five attributes (learnability, efficiency, memorability, errors and satisfaction), which can provide valuable diagnostic aid to practitioners to improve usability. Although these methods are useful in ensuring products are more usable, they were found limited when assessing subjective aspects such the aesthetics and attractiveness (Sutcliffe, 2002) of a product, and in evaluating how these may influence the emotional experiences of users’ while using a system (Hassenzahl & Tractinsky, 2006). Therefore, a broader prospective was needed, that extended the traditional usability concept to allow for a user’s experience.

2.3. Transcending Usability

Criticism of the traditional approach to usability that emphasised the objective aspects of effectiveness, efficiency, and learnability, (which focus on the measurable aspects of task performance and productivity), was that laboratory based experiments missed out the cognitive context of HCI (Whiteside & Wixon, 1987; Winograd & Flores, 1986). This contextual view that “artefacts truly cannot be understood apart from the situation in which they are used” (Carroll & Campbell, 1989, P. 251), prompted research to consider the wider context, beyond usability.
Chapter 2  Literature Review

2.3.1. Emotional Experience

Carroll & Thomas (1988), outlined the challenge of separating ‘ease’ of use and the motivational aspect of ‘fun’, arguing that ease of use implies simplicity, which is not necessarily conducive to making something fun, hence there is a chance it could become boring. As they recognised that “experiences that are fun are more attractive to people” and “we do not necessarily want to make things as simple as possible”, (P. 22), they called for a scientific study of fun (Carroll & Thomas, 1988; Hassenzahl, Platz, Burmester, & Lehner, 2000; see also, Monk, Hassenzahl, Blythe, & Reed, 2002). Similarly, Malone (1981) first raised the difference between users playing education games to those using a computer as a tool for practical tasks, and produced the first heuristics for intrinsic motivation.

Other aspects of human behaviour derived from consumer research identified the emotional and hedonic experiences as powerful motivators of use (Desmet & Overbeeke, 2001; Hirschman & Holbrook, 1982; Holbrook & Hirschman, 1982). Ergonomic and design research identified that usability did not include positive feelings of “pleasure in product use” that takes into account “the emotional and hedonic benefits”, which goes beyond usability and product satisfaction (Jordan, 1998, P.26). Jordan (2000) proposing a hierarchy of needs that placed functionality as the base level, followed by usability then pleasure at the top. Vilnai-Yavetz & Rafaeli, (2005) argued that the appreciation of any item is related to three dimensions, instrumentality (usability and usefulness), aesthetics (sensual experience), and symbolism (a products’ meaning). Norman (1994) questioned traditional usability in his book ‘Emotional Design’, outlining the importance of the emotional and affective responses to a product, indicating three levels, (visceral, behaviour and reflective), which reflect different types of pleasure (D. A. Norman, 2002). These studies brought a change within HCI, to expand from the narrow product-based view of usability to encompass subjective quality concepts such as “enjoyment”, “joy of use” and “pleasure” (Hassenzahl, Beu, & Burmester, 2001; Monk et al., 2002).

2.3.2. Affective Responses

At the same time as the emergence of the motivational aspects of fun, pleasure and enjoyment, an interest in affect was becoming popular within the industrial design field. Although the concept of affect embraces a variety of psychological states including emotions, feelings, moods and passions etc., the terms affect, emotion and mood are often interchangeably, where affect may often be referred to as both mood and emotion. An emotion is usually a reaction, that has a specific stimulus (person, object or event) preceding it, and usually is brief in duration (seconds or minutes). In contrast a mood is more subtle, non-specific, longer lasting and less intense, giving a person a positive or negative tendency
Affect can be viewed as 'feelings' that are evident in both moods and emotions. Defined by Russell, (2003) 'Core Affect' is considered the neurophysiological state that is an integral blend of two dimensions, hedonic valence (pleasure-displeasure), and activation value (calm-excited), see Figure 5. Core affects are typically experienced in relation to no known stimulus, and are free flowing. However the affective quality of a stimulus (object, places, event) has the capacity to change a person’s core affect, which is known as ‘affect infusion’ or affective judgement; where "the process whereby affectively loaded information exerts and influence on and becomes incorporated into the judgemental process” (Forgas, 1995, P39).

From an HCI perspective, the affective responses to a product were the main focus (Desmet & Hekkert, 2002; Desmet & Overbeeke, 2001; Mcdonagh, Hekkert, VanErp, & Gyi, 2003). This was reflected in the ‘Affective Computing’ paradigm that sought to enable computers to interact naturally with human needs, with the ability to recognise and express emotions (Picard, 1997). A further stream of research argued that humans react to ‘computers as social actors’ (CASA), suggesting that people respond to computers in the way they do with other people, and used interactive technology, such as embodied agents or avatars to elicit such behaviour (Khan & Sutcliffe, 2014; Reeves & Nass, 1996). These new avenues of research were due to the advancement of computer graphics and multi-media computer systems that used various visual design factors (colour, light, music etc.) to enhance emotional experiences of its users (Scheirer & Picard, 1999). Affective computing provided insight into understanding phenomena such as attention, memory and aesthetics, and elevated the importance of emotions in relation to human thinking and decision-making (Isen, 1995, 2001; Zhang & Li, 2004).
Affect within UX represents a person's primary subjective response to a product, which can vary between different users, whereas perceived usefulness or perceived ease-of-use is a person's cognitive response to a product (Brave & Nass, 2003;Forgas, 1995; Zhang & Li, 2004). This is in accordance with Norman (2002), who suggests that affect, emotion and cognition interact with each other and influence value judgements; “You cannot escape affect; it is always there. More important, the affective state, whether positive or negative affect, changes how we think” (Norman, 2002, P6). According to Norman (2002), the causal chain of a product’s beauty can elicit a positive affective response, resulting in pleasure, which can influence a user’s perception of usability, which lead to the claim “attractive things work better” (D. A. Norman, 1994). The affect-driven evaluative response to the beauty of a product and its possible influence on usability became an important area of UX research.

Positive affect has been found to have various positive influences on cognition, such as enabling effective decision-making (Isen, 1995), and have been explored in a number of UX studies. A series of studies by Sheldon, Elliot, Kim, & Kasser, (2001) showed a clear link between need fulfilment and positive affect in life events. Similar studies by (Hassenzahl, Diefenbach, & Göritz, 2010; Hassenzahl, 2008b) showed these results were generalizable to positive experience with technology, where need fulfilment (relatedness) clearly linked to hedonic quality perceptions and positive affective experiences with technology. These studies were extended by Partala & Kallinen, (2012) by assessing negative affective experience to provide insights into what needs and affective experiences are most important to both negative and positive experiences. All these studies used the PANAS (Positive and Negative Affect Scale), by Watson, Clark, & Tellegen, (1988), that measures ten positive, and ten negative affect items. However, Partala & Kallinen's, (2012) study also collected qualitative data and noted “current qualitative results highlight the need for further developing systematic methods for qualitative reflections on personal experiences” (P. 32), thus suggesting we can learn more from the content of narratives (Tuch, Trusell, & Hornbæk, 2013).

### 2.4. Aesthetics and Usability

“Aesthetics is clearly of vital importance to the human sense of well-being” (Porteous, 1996, P.5)

The importance of aesthetic is not new, although it was initially given low priority in HCI research due to the difficulty in accessing the subjective perception of aesthetic value. However, the importance of everyday aesthetics was promoted in the influential paper from Dion & Bersheid, (1972), who examined the psychological stereotype ‘what is beautiful is
good’, demonstrating that a persons’ physical appearance can influence social interaction. For example, more attractive people are perceived as being more intelligent, have successful jobs, and happier social lives, which is often referred to as the ‘halo effect’ (Meiners & Sheposh, 1977; Thorndike, 1920). This notion was also applied to product design, where the products visual attractiveness is determined by its aesthetic characteristics; in other words, the more beautiful the product is, the more it may appeal to the consumer. Aesthetic design became important within new product development, forming a large part of retail marketing strategies, used to gain strategic competitive advantage (Kotler & Rath, 1984; Whitney, 1988), where the “design of a product is an unquestioned determinant of its market success” (Bloch, 1995, p16). Although HCI was slow to take up the importance of aesthetics, it soon became evident that the design of computing products, (e.g., Apple Computers), and enhanced website design provide a competitive edge, form an important part in shaping users attitudes towards purchasing and using these products (Lavie & Tractinsky, 2004). One such study by Kurosu & Kashimura, (1995) considered that products should be first appealing, (apparent usability) as well as inherently usable, and found that aesthetic judgement (of several ATM cash machines) strongly correlated to inherent usability. This study was replicated by (Tractinsky, 1997) and started the debate around the relative influences of usability and aesthetics.

Aesthetics has become one of the most frequently researched dimensions in the field (Bargas-avila & Hornbæk, 2011), with much debate and conflicting viewpoints about its relative influence to usability. In accordance with the halo effect, several studies suggested a correlation between the aesthetic quality of an interface and its perceived usability. In a before and after experiment where users rated the changing layout of 9 different ATM interfaces Tractinsky et al., (2000) found that perceived aesthetics were strongly related to perceived usability, resulting in the claim that ‘what is beautiful is usable’. Similar correlations were found by Lavie & Tractinsky, (2004) who outlined a multi-dimensional model of usability and aesthetics, which was further expanded by Tractinsky & Zmiri, (2006) to include symbolism as an additional aspect of quality. Further experimental studies supported Tractinsky’s (1997, 2000) claim, showing that aesthetic design influenced usability ratings (Ben-Bassat, Meyer, & Tractinsky, 2006; Mahlke & Thüring, 2007). While other studies showed the opposite, that ‘what is beautiful is usable’ can be reversed, meaning that a more usable interface can lead to higher aesthetic ratings (S. Lee & Koubek, 2010a; Tuch, Roth, Hornbæk, Opwis, & Bargas-Avila, 2012). Therefore, it seems that the notion ‘what is beautiful is usable’ was only partially supported, as in some cases it can be described as ‘what is usable is beautiful’.

39
One confounding issue with these studies is the manipulation of either the aesthetics (through changing the interface design), or altering the interfaces’ usability factor, which can cause problems when making causal type interpretations, so no clear conclusions can be drawn. Some studies found no correlations between perceived usability and aesthetics, and the two concepts were independant (Hassenzahl, 2004; Lindgaard & Dudek, 2003; van Schaik & Ling, 2009). Further suggestions were that the original studies (Kurosu & Kashimura, 1995; Tractinsky, 1997) suffered from methodological issues due to the ‘improvised’ design which measured only the spatial layout (without changes to form, colour or style), making generalisation to real products difficult (Hassenzahl, 2004, p325).

2.5. Hedonics and Usability

Adopting a social psychology perspective Hassenzahl, (2001) observed that usability and aesthetics alone could not explain users' preferences and experiences with interactive products, as they neglect the subjective nature of ‘appealingness’ and pleasure. Hedonics was introduced as a non-task quality dimension to determine the subjective judgements of product appealingness (Hassenzahl, Platz, Burmester, & Lehner, 2000). In a later study using the aesthetic quality of MP3 player skins (Hassenzahl, 2004) considered the relationship between perceived Pragmatics (usability), Hedonics (identification and stimulation), along with two high-level evaluational constructs Goodness (value) and Beauty (appeal). Pragmatic quality refers to the products ability to enable “effective and efficient goal-achievement” (ease-of-use) while hedonic quality relates to the “users’ self”, and refers to the products capacity to provide, “stimulation by its challenging and novel character, or identification by communicating important personal value to relevant others” (Hassenzahl, 2004, P 7). Goodness refers to the overall evaluation of a product (value), while beauty is defined as “predominately affect-driven evaluative response to the visual Gestalt of an object” (Hassenzahl, 2008a, p. 291), where visual judgements can be made instantly, within milliseconds (Lindgaard, Dudek, Sen, Sumegi, & Noonan, 2011).

The interplay of these aspects are complex. Hedonics was found to be strongly related to beauty, while goodness relates to “objective” pragmatics (actual usability, rather than perceived usability), induced by the mental effort of experiencing usability problems and task performance (see Figure 6). While percieved usability (perception of pragmatic quality), are important contributors of goodness, no relation between actual and perceived usability and beauty were found (Hassenzahl, 2004).
Figure 6. Hassenzahl’s, (2004) summary of the relationships between attribute groups (pragmatics, hedonics) with evaluational constructs (goodness, beauty) and experience (mental effort) (P. 342).

The subjective perceptions and experiences are separate from evaluations, which allows for users to find a product novel (attribute), but not necessary like it (an evaluation), so perceptions of hedonic or pragmatic qualities could lead to positive evaluations, but is not gaurenteed. So according to Hassenzahl (2004), perceived usability and goodness are affected by experience, so subject to change, whereas hedonic attributes and beauty ratings remain stable over time (Hassenzahl & Monk, 2010; Hassenzahl, 2004). In addition, the task type and context of use can influence perceived product quality, for example, pragmatic quality will be more important when the user is in goal-mode (focusing on achieving a particular goal), whereas action-mode (user is focused on actions), then hedonic quality is more important. In other words, if a user is accessing a e-commerce website to make a specific purchase, then pragmatic quality of the website is more important, whereas if the same user is accessing the website to play a game, then hedonic quality will take priority (van Schaik & Ling, 2008).

2.6. Defining User Experience

"User experience focuses on interaction between a person and something that has a user interface" (Law et al., 2009).

User experience evolved from the concerns that traditional usability focused on the objective measures of user performance and did not account for the emotional and affective responses to the aesthetics of an interactive product. Similar to usability, the ability to evaluate UX has become the focus of both the HCI research community and industry practitioners alike. Despite the term ‘user experience’ being widely used and readily accepted, there is still uncertainty about what it actually means. Due to the multidisciplinary nature of UX research, which draws from computing, psychology sociology, and design and beyond, it’s not
surprising that no clear unified theory of user experience exists. Only a few attempts at providing a definitive definition have been made. For example, a survey conducted by Law et al., (2009) asked 275 researchers and practitioners within the field of UX to come up with a common agreement about what UX is. The result outcome was that UX is an individual experience that is “dynamic, context-dependent, and subjective” (p.719). In a more recent white paper on UX, the term is defined as encompassing three different perspectives; ‘experiencing’ which refers to “the individual and dynamic nature of experiencing the encounter with a system”; ‘a user experience’, which is a retrospective perspective, based on the “outcomes and memories of an experience”, and ‘co-experience’, which refers to “experiencing a situation together” (Roto et al, 2011, p.7). Although this goes some way towards identifying variances within UX, there is still a lack of clarity in understanding what UX is. Despite this, there is a vast array of UX research with many suggested theories and frameworks, which have made some progress in teasing out some of the complex intertwined aspects that embody UX.

2.7. UX Theories, Frameworks and Models

Since the 1980s there have been numerous theories and models produced in order to understand, evaluate and design better interactive technologies. The shift from usability focused methods to include a more holistic understanding of user experience to include concepts of fun (Monk et al, 2002), pleasure (Jordan, 2000), aesthetics (Tractinsky et al, 2000) and hedonic qualities (Hassenzahl, 2004) have added to the complexity. However no one consensual theory has emerged (E. L.-C. Law, van Schaik, & Roto, 2014; E. L.-C. Law & van Schaik, 2010). This is partly due to the variety of diverse perspectives, which can be roughly divided into two viewpoints, the objective or ‘reductionist’ approach that have their roots in cognitive psychology, or the subjective ‘experimental’ view that is rooted in pragmatist philosophy and views UX as a ‘holistic’ experience (Blythe, Hassenzahl, Law, & Vermeeren, 2007).

2.7.1. Reductionist verses Holistic

The holistic approach to user experience draws from the pragmatist philosophy of John Dewy, who focuses on the experience as a whole (person and environment), which is integrated and therefore inseparable. “It is not possible to divide in a vital experience the practical, emotional, and intellectual from one another and to set the properties of one over against the characteristics of the others. The emotional phase binds parts together in a single whole”, (Dewey, 1934, P. 55). Holistic approach to UX focus on the qualitative and complete aspects of experience as Buchenau & Suri, (2000) stated that “experience is a dynamic
relationship with other people, places and object". Wright, Blythe, & McCarthy, (2006) adopted this holistic approach to experience calling it the ‘felt life’ of individuals to which it has “a unity or wholeness that is fulfilling” (P.5)

In contrast, the reductionist approach contends that “complexity and richness of user experience [...] can be reduced to a set of manipulable and measurable variables” to create “abstract models and classifications” (Blythe et al., 2007, P.68). The reductionists embrace the measurement modelling approaches to understanding UX, which has predominantly focused on quantitative methods (Law & van Schaik, 2010). The holistic group argue that the reduction of experience “maybe useful for experimental analysis but they can miss some of the insights available in accounts that resist reduction” (Swallow, Blythe, & Wright, 2005, P.91); while conversely the reductionists highlight the benefits of using models, as “meaningful and valid measures of these constructs are deemed necessary and useful for comparing as well as predicting the quality of the artefact contributing to the experience being measured” (Law & van Schaik, 2010, P.315). There is a need for a mutual recognition of the strengths and weakness of the related approaches and values in order to benefit the advancement of emerging UX research. One main focus of this thesis is to bridge the gap between the objective and subjective approaches by combining both qualitative and quantitative methods in order to enhance our understanding of the nuances of UX.

2.7.2. Early UX Frameworks

Some of the first theoretical UX frameworks came from the holistic approach which focused on how experience is formed, aiming at interaction design. For example, Forlizzi & Ford, (2000) presented an initial framework of UX that combines the user, product and context, where “the type of experience has a beginning and an end, and changes the user, and sometimes, the context of the experience as a result” (Forlizzi & Ford, 2000, P420). They describe an experience as a “constant stream of “self-talk” that happens when we interact with products”, and these naratives represent the subjective side of experience, best captured through qualitative methods. (Forlizzi & Battarbee, 2004). Their framework was modified later to include ‘co-experience’, the social context of an experience (Battarbee, 2003; Forlizzi & Battarbee, 2004). The importance of context, culture and social factors was also shown in Arhippainen & Tähti’s (2003) framework generated from interviews and observations of users interacting with mobile PDA prototype devices. This contextual approach was also conceptualised by (Wright et al., 2006) in four intertwining threads, sensual, emotional, compositional and spatio-temporal, where the whole experience is situated in a place and time, involving both visceral and emotional ‘felt’ sensations (Wright, Wallace, & McCarthy, 2008).
2.7.3. Modeling UX

It is recognised by both approaches (holistic and reductionists) that there is a need for theories, models and frameworks to capture and communicate ideas about how best to design for UX. Theoretical research frameworks are used to conceptualise UX, and can comprise of either instrumental (objective) or non-instrumental (subjective) elements, or both. Where models can either be structural, used to establish causal relations between constructs (similar to frameworks), or measurement models that are used as a means to measure various constructs to aid evaluation and improve understand (Law et al., 2014).

UX models tend to take a reductionist view, with an emphasis on measuring UX constructs and evaluating their relationships. Numerous UX models have been proposed, many with differing approaches; however, these can be distinguished into three main clusters.

2.7.3.1. Construct Models

The multiple construct approach used to evaluate UX by Lavie & Tractinsky, (2004) developed a measurement model of users’ perceptions of aesthetics within websites. These came from initial exploratory studies (Lavie & Tractinsky, 2004; Tractinsky, 1997, see section 2.5), that used factor analysis to identify ‘a set of latent constructs’, that consisted of five factors; classical aesthetics, expressive aesthetics, perceived usability, pleasure and service quality (utility). Classical aesthetics related to the traditional notions of aesthetics (clear, symmetrical, well organised), while expressive aesthetics goes beyond classical principles and represents creativity (originality, fascinating design) (Lavie & Tractinsky, 2004). This was further developed by (Porat, Liss, & Tractinsky, 2007) who adopted the environmental psychological M-R model (Mehrabian & Russell, 1974) used extensively in the areas of marketing, which considers that perceptions of the environment, which affect the emotional state of a user. The emotional states are captured using bi-polar dimensions, pleasure (satisfaction), arousal (affect), dominance (user-control) (Porat & Tractinsky, 2012).

Findings showed that both classical aesthetics and expressive aesthetics influenced arousal and pleasure, which in turn influence attitudes to the product (e-store), and usability influence pleasure and dominance, (see Figure 7). However, the high-touch experiential products (e.g., luxury goods) produced better ratings of expressive aesthetics, so more positive emotions, while low-touch search based product (e.g., ebook stores) were evaluated somewhat more usable, indicating the products content more likely biased their judgement towards the aesthetic design (Porat et al., 2007; Porat & Tractinsky, 2012).
2.7.3.2. Inference Models

The ‘inference perspective’ is grounded in psychological theory of judgement, based on ‘probabilistic consistency’ (Ford & Smith, 1987) when product information is incomplete, consumers’ judgement are based on inferences about missing information. Similar to the construct model Hassenzahl, (2004) suggest that perceived hedonic and pragmatic quality constructs influence global evaluations of the products goodness and beauty (see Section 2.5). The link between pragmatic quality and beauty is indirect, and mediated by goodness (product value), whereas hedonic quality is directly inferred from beauty (see Figure 8).

Hassenzahl’s (2004) inference model posits users infer unavailable attributes (actual usability), directly from specific available attributes (beauty) and infer an overall quality (goodness), as per the claim “beautiful is good”, (which corresponds to Tractinsky, (1997), who suggests beauty influences perceived usability). This theory has been extensively tested using the AttrakDiff\(^1\) questionnaire that measures perceived hedonic and pragmatic quality, along with the global constructs beauty and goodness, across a range of products and user groups (Hassenzahl & Monk, 2010; van Schaik et al., 2012). This model could be considered similar to the multiple construct theory, by which people judge the product on their initial

\(^1\) http://attrakdiff.de/index-en.html
perceptions, however the inference perspective acknowledges that some information maybe absent (usability requires interaction with the product). This model is also subject to contextual effects, for instance Diefenbach & Hassenzahl, (2011) explored the notion of Hedonic Dilemma showing that the beauty of a product (mobile phone) was valued more, when in a forced choice situation people choose the more usable product, above the more hedonic. Reasons suggested were that people choose what is easy to justify (pragmatic), and not what they enjoy most (hedonic) even if they would feel better with the hedonic choice. Similar trade-off choices were found by Hartmann, Sutcliffe, & De Angeli, (2008), between more aesthetic and customised designs in mobile phone applications. In a similar study user judgement of aesthetics and overall preferences varied depending on the prior information given (Hartmann, Sutcliffe, & De Angeli, 2007), and even the order in which the positive and negative information is presented, indicating a framing or priming effect (Bloch, 1995).

Similarly, experiences of goal-orientated tasks (influenced by the users goal-mode or action-mode), and users background can influence ratings of pragmatic qualities (van Schaik et al., 2012), and their framing effect through different motivating scenarios (good/bad experiences) (Hassenzahl, 2008b). User perceptions of need fulfilment was also explored by Hassenzahl, Diefenbach, & Göritz, (2010) in a study that collected over 500 positive experiences with interactive products, with results showing clear relationships between need fulfilment (e.g., competence, stimulating, relatedness) and positive affect. Hedonic quality perceptions were found to be strong motivators due to their ability to create positive experience, while pragmatic quality was seen as a hygiene factor, as it enabled fulfilment of needs, but was not a source of positive experience itself. As UX becomes more diversified the examination of various variables to include in a model becomes a greater challenge.

2.7.3.3. Process Models
The third group are process models, of which some describe a dual-process that combines motivation and cognitive ability (Metzger, 2007), while others a three-phase process (Ou & Sia, 2010). The process model views UX as a cognitive process that can be modelled and used to measure or evaluate changes in perception and judgement over time. The integrated CUE (Components of User Experience) model proposed by Mahlke & Thüring, (2007) distinguished three UX components, instrumental, non-instrumental and the emotional reactions of the user. The perception of instrumental qualities (usability and usefulness) and the non-instrumental (visual aesthetics) qualities are influenced by emotional reactions, which all lead to an overall appraisal and judgement of a system (preference, behaviour, choice). The ‘interaction characteristics’ is the users’ perception of the instrumental and non-instrumental qualities, are influenced by the systems properties, task and context and users skill and knowledge, as shown in Figure 9.
Some researchers have explored the very early stages of the process, when users first form aesthetic judgements. Lindgaard, Fernandes, Dudek, & Brown, (2006) recognised that first impressions matter, demonstrating that perceptions after a brief exposure (50 msec), create an aesthetic impression, which correlate with those without a time limit, which positively influence the user’s evaluation and attitude towards the product (website). This was based on the ‘mere exposure effect’ experiments conducted by Zajonc, (1980) who claimed that initial visual perceptions are an affective response that occurs prior to the cognitive process.

In a further two experiments Lindgaard, Dudek, Sen, Sumegi, & Noonan, (2011) observed the response time for aesthetic judgements was much faster than for usability or trustworthiness, where a certain amount of ‘cognition reflection’ is needed, and that visual appeal influenced first impression judgements of other characteristics, such as perceived usability and trustworthiness, and the more aesthetically design product was favoured.

Lindgaard et al.’s, (2011) model shows that user judgement is driven by the initial visual appeal (aesthetics) of a product, followed by the cognitive reflection of usability, utility and trustworthiness, which occurs later due to the extra time needed to process (see Figure 10). Therefore aesthetics can create a halo effect, influencing the overall judgement of the product. This model has similarities to Porat & Tractinsky, (2012) who suggest that perceived aesthetics (classical and expressive) induces certain emotional states (arousal and pleasure), which in turn affects the users attitudes towards the product.
Hartmann et al., (2008) proposed a process model which focuses on the sequence of interactions when people make quality judgements within one session. Based on the adaptive decision making theory (ADM) by Payne, Bettman, & Johnson, (1993) which posits that user judgements depend on the context and task at hand, and individuals adapt their responses to different decision situations. Hartmann et al., (2008) proposed that people make decisions by adapting their strategies depending on the task in hand, their background (knowledge/skill) and the criticality of the decision, (see Figure 11). User judgement will also depend on the interactions between the decision-making criteria (design qualities such as usability, content, aesthetics, brand and customisation) that are modified by the context (serious vs not serious). Hence, user experience is determined by the design qualities biased by context, task and user background. In addition the relative importance of the judgement criteria changes during session, as demonstrated by the role of aesthetics in initial and later exposure to websites (Lindgaard et al., 2006).
2.7.3.4. **Technology Acceptance Models (TAM)**

Process models share the same approach as the widely used TAM (Technical Acceptance Models) originated by Davis, (1989), whose purpose was to model users’ intention to accept and use a new technology, taking a longitudinal perspective. TAM is based on the principles of reasoned action (Ajzen & Fishbein, 1980), and suggests that an individual’s intention to use a technology is jointly determined by their perception of the technology’s perceived usefulness (PU) and its perceived ease of use (PEoU). Although there have been numerous studies that have validated and extended the original TAM framework, it has suffered much criticism for what is considered its strength, its simplicity (parsimony), as explained by Bagozzi, (2007), “It is unreasonable to expect that one model, and one so simple, would explain decisions and behaviour fully across a wide range of technologies” (P.244). The problem could be due to preceding researchers reliance on TAM, as Benbasat & Barki, (2013) outlines “study after study have reiterated the importance of PU, with little research effort going into investigating what actually makes a system useful” (P.212).

The TAM model was expanded by Venkatesh, Morris, Davis, & Davis, (2003) in their Unified Theory of Acceptance and Use of Technology (UTAUT) model, by adding eight independent variables for predicting behaviour, which include user characteristics (age, gender, experience), and the social influence. Additional attributes were also explored over time by Magni, Susan Taylor, & Venkatesh, (2010), whose model included two hedonic constructs (personal innovativeness, and cognitive absorption) along with the instrumental factors (performance, enhancement). Other models proposed that design attributes such as perceived interactivity influenced judgements of efficiency, effectiveness, which in turn determined product preferences and enjoyment (Cyr, Head, & Ivanov, 2009; S. Lee & Koubek, 2010b; Teo, Oh, Liu, & Wei, 2003). Although these models share the same inference perspective as Hassenzahl, (2004) framework, they are difficult to compare, due to their different interpretation if constructs and variables. For example hedonics is treated as one unit by Magni et al., (2010), as a value by Teo et al., (2003) that contributes to overall satisfaction. In addition, the element of time (long-term or short-term) can add to the complexity, which can vary depending how the model is applied (Law & van Schaik, 2010).

2.8. **Clarifying User Engagement**

This thesis aims to expand existing models of UX by focusing on what makes users engage with technology by investigating the influence of interactive and design features on UX. User engagement (UE) can be defined as a more restricted view of UX, and refers to “the quality of the interactive experience rather than the whole life span experience of a product” (Sutcliffe,
2009, p1). UE refers to exploring why technologies attract people to use them and continue to use them within a session, while UX is reserved for the wider view encompassing what makes people adopt and continue to use technology over long-term multi-session use. To understand what attracts people to use, and continue to use interactive products encompasses a mixture of wider influences, beyond UX, such as aesthetics, interactivity, affect, presence and flow. Some of these constructs have been well used in UX research, such as aesthetics, affect and emotion, (see previous sections), while others have had little attention within HCI literature, such as interactivity, flow and presence.

### 2.8.1. Flow and Presence

The role of immersion (flow) and presence (involvement) have been well researched (Berlyne, 1960; Csikszentmihalyi, 1975, 1988), yet how these principles influence UX is poorly understood. Flow was defined by Csikszentmihalyi, (1990) as an ‘optimal experience’, described as "the state in which people are so involved in an activity nothing else seems to matter" (P. 4). To maintain flow requires optimal arousal, where curiosity is maintained through complexity and variety (Berlyne, 1960), by a mix of perceived challenge (difficulty) and user ability (skill). The focus is keeping the user in the flow zone, which has been a much used concept within game design (Sweetser & Wyeth, 2005). It was adapted in the Computer-Mediated Communication (CMC) context by Trevino & Webster, (1992) who defined four dimensions of flow; sense of control, attention focus, curiosity and intrinsic interest (with the last two being considered as a construct of cognitive enjoyment). Also in CMC, Hoffman & Novak, (1996) conceptualised flow in their process model, which combined interactivity, telepresence and flow (skill, control, challenge and attention). Agarwal & Venkatesh, (2002) provides a multi-dimensional framework of cognitive absorption, an extension of flow which includes enjoyment, curiosity, control and focused immersion.

Although flow show some parallels to immersion, it is considered more fleeting or transitory (Sanders & Cairns, 2010), while immersion is more long lasting, as described by Jennett et al., (2008), who view immersion as a graded three-step process that begins with engagement, then involvement to total immersion, with flow experienced at the extreme end of immersion.

The origins of presence come from virtual reality where the user is represented as an avatar (virtual character). Presence relates to the subjective experience of feeling situated in one place, defined succinctly as ‘the sense of being there’ (Lombard & Ditton, 1997). Sometimes presence is considered to be two related components, telepresence (the sense of feeling present in the mediated environment), and social presence (the sense of being there with another), whereas presence refers to the natural perception of the environment, (Biocca, Harms, & Gregg, 2001; Steuer, 1992). According to Witmer & Singer, Michael, (1998) presence is a multifaceted concept, where control and involvement are essential components,
although other factors (attention, naturalness, sensory experiences) are important, which like immersion are experienced in varying degrees. Due to the nature of presence, research studies are found mainly in virtual reality or gaming, with few studies exploring the effect of presence in other areas. In a study by Qiu & Benbasat, (2005), presence was explored using a with/without avatar as live-help aids to online shopping site, and found that 3D avatar improved users sense of telepresence, but not social presence. Whereas Yoon, Laffey, & Oh, (2008) compared 2D and 3D graphical representations of an online furniture store, and found users felt a higher sense of presence while interacting with the 3D store, which enhanced their sense of usefulness and usability.

2.8.2. Interactivity

The nature of interactivity has been widely discussed within the fields of advertising, marketing, communication, education and computer science, yet the concept still eludes a clear definition (Domagk, Schwartz, & Plass, 2010). An extensive review of the interactivity literature conducted by McMillan & Hwang, (2002) identified three different types of interactivity; process orientated (interchange or 2-way communication), user control (interactive features) and time perception (responsiveness), although all three dimensions often overlap and interrelate. However Rafaeli (1988) suggested a process-orientated definition, relating to the degree of responsiveness (reciprocal communication), which Hoffman & Novak’s (1996) based their feedback framework used within the computer-mediated environment (CME). A definition for perceived interactivity of websites proposed by Wu, (2006) combined three elements, perceived control, perceived responsiveness and perceived personalisation. Similarly, perceived interactivity was shown to influence trust within mobile e-commerce environment by Lee, (2005) whose framework of user control, user connectedness and responsiveness to the user was also used by (Cyr, Head, & Ivanov, 2009; Cyr, Head, Larios, & Pan, 2009). Kristof & Satran, (1995) suggest that interactivity motivated users by giving them self-directed control, where low and high levels of interactivity can influence engagement and cognitive processing. They developed a seven-level scale based on the varying levels of interactivity and user control, used by (Teo et al, 2003; Wang, Vaughn, & Liu, 2011). While Johnson, Bruner, & Kumar, (2006) examined perceived website interactivity in relation to nonverbal information, speed of response, responsiveness and reciprocity, arguing that interactivity involves a reciprocal communication process which depends on the system level of responsiveness to the user. Despite the past efforts of the last decade, where several explanatory frameworks have been proposed (Hoffman & Novak, 1996; Kristof & Satran, 1995; Lee, 2005; Wu, 2006), the concept of interactivity is somewhat elusive, with no general consensus about its definition (Johnson et al., 2006; Sohn & Lee, 2005).
2.8.2.1. Interactivity and HCI

Despite the importance of interactivity, the topic has been given little attention within the HCI literature, with most research originating from business and marketing, or computer science. There is also a growing body of research exploring the use of interactivity within museum and cultural spaces to aid learning and enhance engagement through multi-media exhibits and mobile technologies (Haywood & Cairns, 2006; Othman, Petrie, & Power, 2011). Also, the influence of interactive design features (i.e., virtual 3D environments and interactive avatars) used in games to enhance user engagement have been reported in several studies (Jennett et al., 2008; Schild, LaViola Jr, & Masuch, 2012). It is recognised that the interactivity of a website is an important factor in attracting and engaging users and can lead to increased satisfaction and sense of efficacy (Bargas-avila & Hornbæk, 2011; Venkatesh et al., 2003).

Previous research suggests that interactivity within websites can improve usability and user satisfaction (Lowry et al., 2006), encourage playfulness (Chen & Yen, 2004), enhance emotional pleasure (Fiore, Jin, & Kim, 2005) and engagement (Mollen & Wilson, 2010; Szuprowicz, 1995). However, interactivity within advertising was found to interrupt the process of persuasion, and therefore was not effective (Bezjian-avery, Calder, & Iacobucci, 1998). This disparity can be due to other confounding issues, such as the level of interactivity, type of interactive features, user characteristic and context of use (Teo et al., 2003). Lee & Koubek, (2010b) found the effect of perceived aesthetics and usability was stronger pre interaction than post, and that aesthetics had an influence on user preference. In a study that applied three levels of interactive communications, Teo et al., (2003) showed that high levels (online social feedback) had positive effect on users perceived satisfaction, effectiveness, efficiency and general attitude towards the websites. Although Porat & Tractinsky, (2012) found the aesthetics of online stores influenced consumers’ emotion (pleasure), affective (arousal) states, and attitudes towards the web stores, the influence of interactive features was not fully explored.

Perceived interactivity (user control, connectedness and responsiveness) was investigated within e-commerce by (Cyr, Head, & Ivanov, 2009) with various degrees of dynamic visualisation, where enhanced interactivity in visual information positively influenced users perception of effectiveness, efficiency, enjoyment and trust, leading to a greater website product loyalty. O’Brien (2010) conducted a survey of online shopping websites examining interactivity indirectly with various engagement constructs (focused attention, novelty, felt involvement etc.), which were positively influenced by social interaction facilities, while user attitudes (hedonic motivation) to the online shopping experience, (idea, adventure) may have been increased by the interactive features used to enhance product presentation and exploration. In contrast Yi, Jiang, & Benbasat, (2011) used two levels of triggered interaction
(where the user is prompted to interact) found that a more constrained level was more effective than unconstrained triggered interaction or a full video presentation, leading to stronger product seductiveness and more positive attitudes towards the website. Similarly, Jiang & Benbasat, (2007) compared static presentations of products with interactive versions and found that interactivity and vividness enhanced users intention to purchase, where vividness refers to the "richness of a mediated environment as defined by its formal features; that is, the way in which an environment presents information to the senses" (Steuer, 1992, P.81). While Chen & Yen, (2004) found that playfulness, connectedness and reciprocal communication are important predictors of website quality and preference, similarly De Angeli, Sutcliffe, & Hartmann, (2006) found that users experience (expressive aesthetics, pleasure) and overall preference was enhanced by the interaction style.

In summary, much of the research in flow and presence is found in virtual reality and games, while interactivity has been investigated (across many fields), with a particular focus on websites (Cyr, Head, & Ivanov, 2009; Cyr, Head, Larios, et al., 2009; O’Brien & Toms, 2013; O’Brien, 2010; Teo et al., 2003). The positive influence of interactivity on efficiency, effectiveness and satisfaction has been shown, as reflected in the TAM measures (Venkatesh et al., 2003), however the influence of subjective measures of UX (e.g., aesthetics, emotion and affect) on interaction has received less attention. In addition, the approach to interactivity has been based on Lee, (2005) and Hoffman & Novak, (1996), frameworks, which focus on user control, responsiveness and connectedness, rather then considering specific design features such as virtual environments, avatars, interactive media, that may influence UX measures. Therefore, the influence of design features (3D environments, interactive media and avatars) that may promote interactive experience and a sense of flow or presence have not been fully explored in relation to UX.

2.9. Expanding UX

Wider issues of context of use and task have been well recognised as contributing factors of UX (Bargas-avila & Hornbæk, 2011; Hartmann et al., 2008; Hassenzahl & Tractinsky, 2006; van Schaik & Ling, 2009), although the influence of individual differences has received less attention. Investigating UX over time is a growing area of research, although few studies have explored longer time frames involving multiple session use. This thesis further aims to explore the influence of individual differences upon quality judgements while using an interactive product, in both short time (in session) and longer time (6 months) experiences.
Chapter 2   Literature Review

2.9.1. Individual Difference

Understanding user difference has been a prominent area of psychology research over the past decade, but has remained marginal within the HCI field (Dillon & Watson, 1996). The ever widening diversity of user population makes it more difficult to design products based on one generic model. Although UX practitioners may have a vague sense of their intended users based on scenarios and personas (Carroll, 2000; Cooper, 1999; Pruitt & Grudin, 2003), which although build detailed pictures of their intended users, may not be identifying representative users, as they are based on subjective judgement of the designers or developers of the product (Sinha, 2003). There has been recent interest in understanding user diversity not only from a way of selecting appropriate users for research studies, but also from a cognitive-process perspective, where individuals differ (across and within) their subjective judgements of an interactive product (Karapanos & Martens, 2007; Kujala & Kauppinen, 2004). Although HCI literature advises the need for “a clear understanding of personality and cognitive styles” that can be “helpful in designing interfaces” (Shneiderman, 1992, P. 26), and suggests numerous lists of population profiles, such as age, gender, education, culture, skill, etc., the challenge lies in the ability to identify user types accurately (Karapanos & Martens, 2007; Kujala & Kauppinen, 2004). This has raised recent calls for the need for new methods that avoid the generic average of findigs at group level, that often mask individual difference diversity (Karapanos, Martens, & Hassenzahl, 2009). The use of the repertory grid technique has been used by Jordan & Persson, (2007), where different user groups responded differently to different product types, indicating that one-size fit design approach maybe inappropriate. Also Karapanos, Martens, et al., (2009a) showed diversity between different users and also within a single individual, who can often have conflicting views (beautiful and expensive).

Although the holistic approach recognised individual differences, where cultural, social factors, and contextual factors were seen as influences within the UX (Arhippainen & Tähti, 2003), there are few studies that have explored how user differences influence user judgement of an interactive product. Woszczynski, Roth, & Segars, (2002) developed a integrated theory of playfulness in computer interactions, using multiple constructs, of flow, personality and trait influences. They suggested that personality traits were antecedents to playful behaviour, which is closely related to flow, and playful behaviour can result in user satisfaction. Perceptions of design quality was shown to be influenced by user background, as people hold different rules in terms of valididy of beauty. In a consumer research study by Bloch, Brunel, & Arnold, (2003) individual difference was identified through using the ‘centrality of visual product aesthetics'(CVPA) scale, showing it was an important influence on the value of beauty. Individuals with higher CVPA (value, acumen, response), were more
likely to respond to beautiful things leading to a higher purchase intention than individuals with low CVPA. So some individuals may be more influenced in their judgements by beauty than others. Similar findings were gained from Hartmann, Sutcliffe, & De Angeli, (2008), who compared three different aesthetically designed websites that contained similar content and found that user judgement was influenced by their background, task and context, and that users ‘experiential predisposition’ towards aesthetics influenced their preference. Blom & Monk, (2003) suggest that personalisation of technology devices can affect users cognitively, socially and emotionally enabling them to reflect their personal, group and social identity. There is a growing demand for personalisation of user interfaces, represented by a growing interest in personalisation of UX and recommender systems that target group types (Karat, Blom, & Karat, 2004; Ricci, Rokach, & Shapira, 2011).

### 2.9.2. UX Over Time

Past HCI research has traditionally focused on ‘first-time’ experiences, with only 36% of research being conducted over long-term (Vermeeren et al., 2010). Despite recent calls for the need to consider the role of time in HCI research (Courage, Jain, & Rosenbaum, 2009; Hassenzahl & Tractinsky, 2006; Karapanos, Zimmerman, Forlizzi, & Martens, 2009a), few studies have investigated how time may shape individual motivations and behaviours, and how this may impact on UX (Zhang, Aikman, & Sun, 2008). This is due to the practicalities of conducting longitudinal research, which can be both time-consuming and costly due to the methods employed (Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, & Sinnelä, 2011; Law & van Schaik, 2010).

There are few theoretical frameworks that describe the temporality of UX. The first seeds evolved from a holistic approach, for example Forlizzi & Battarbee, (2004), describe the transformation of experience, forming first in the unconscious, then into a cognitive state, which manifests as an experience that can be shared by others through social interaction. While McCarthy & Wright, (2004) describe experience in terms of sense-making, developed through six activities from anticipation to recounting. In contrast, there is an established body of longitudinal research within MIS (Management Information Systems) focusing on product adoption (Davis, 1993; Karahanna, Straug, & Chervany, 2013; Venkatesh & Brown, 2001; Venkatesh et al., 2003).

Amongst the few published longitudinal research studies within UX, the focus has been mainly on short time durations, ranging from days to a few weeks, and often involving small participant numbers (Karapanos, Zimmerman, et al., 2009a; Kujala et al., 2011). There are however a few rare studies that have focused on the dynamic nature of UX over months. For example, Mendoza, Novick, & Paso, (2005) conducted an eight week longitudinal study to
understand how usability may change over time. They tracked middle school teachers’ experiences while they created websites, and found that frustration levels dropped as they moved from novice to experienced users. Karapanos, Hassenzahl, & Martens, (2008) conducted an in-depth five-week diary study following 10 participants use of an interactive TV pointing device, and found that initially ease of use and usefulness was important, but shifted to identification (social meaning) over time, mediated through goodness (satisfaction); while beauty initially related to stimulation (novelty), lost is appeal over time. Similar findings were found in a later study by Karapanos, Zimmerman, et al., (2009a) when following 6 participants for five-weeks after they purchased a mobile phone, where hedonic experiences (novelty) faded, being replaced with a different hedonic quality (meaning experience), and long-term usability and usefulness became more important. Von Wilamowitz-Moellendorff, Hassenzahl, & Platz, (2006) conducted a focused study (7 people) using a retrospective interview technique to elicit changes in perceptions towards their mobile phone usage, and found that pragmatic perceptions improved over time, while hedonics deteriorated, due to familiarity of use. Magni, Susan Taylor, & Venkatesh, (2010) examined PIIT (Personal Innovativeness in IT), users intentions to explore new technology, along with hedonic and instrumental factors over a one-year study that involved two time points (adoption and post adoption). They found that curiosity influences user intention to explore technology, and over time hedonic factors decrease, while utilitarian factors increases. In addition, users with high PIIT were more inclined to form an intention to explore novel technology, compared to low PIIT users. Similarly Karahanna et al., (2013) found that as users become more knowledgeable of technology over time, their attitudes are influenced more by instrumentality, enabling the discovery of new features and uses. In summary, UX research has focused on the initial experience with interactive products, with few longitudinal studies extending beyond a few weeks, often using small participant groups, where a lack of understanding of how various factors (e.g., interactivity, affect, user difference) may influence user satisfaction.

2.10. Methodology and Rationale
The dimensions and definition of UX still remains somewhat elusive, with varying opinions to the key characteristics that contribute to this multidimensional concept (Kujala et al., 2011). Similarly, the methodologies that are used to understand UX have also been debated with the long-standing rivalry between the quantitative and qualitative approaches still remaining (Bargas-avila & Hornbæk, 2011). The UX community is divided between those who model and measure UX (reductionist), and those who doubt the utility of dividing UX into separate
measurable attributes (holistic). However, there is a growing trend of those who affiliate to both perspective (Law & van Schaik, 2010) to which this research adheres, through adopting a dual-approach it employs a multi-methods measurement approach. Quantitative questionnaire measures will be used to evaluate UX at the population level to determine overall group trends and attitudes, which are combined with qualitative interviews, observations and diary studies, in order to understand the complexity of the single user’s judgement of a products quality at the individual level (Ricci et al., 2011).

Using a mixed methods research approach does not simple mean using two separate strands of quantitative and qualitative studies. There are different ways these two methods can be combined, and the stage of integration is important, which refers to when and how the mixing occurs. Creswell & Clark, (2011) identified six prototypical versions of mixed methods design (see Page 69-70 for overview), which depend on two aspects, the time ordering (concurrent or sequential) of the quantitative and qualitative stages, and degree of dominance of each of these methods (Wu, 2010). Overall this thesis adopts an exploratory sequential mixed methods design that will begin with the collection and analysis of the quantitative data, followed by the qualitative phase that will be used to explore the quantitative results in more detail, although flexibility will allow for emergent themes to be explored.

2.10.1. The Studies

This thesis reports on three studies that were conducted to investigate the effect of interactivity upon UX through using a mixed methods approach. The first two studies will use a traditional lab-based experimental design where a selection of quantitative questionnaires will be employed to capture users’ attitudes on different interactive websites, with the focus on evaluating group-level core constructs. This will be combined with post-test qualitative interviews that draw out the individual-level user judgement inferences. The results of the first study will be used to inform both the findings and method design (e.g., questionnaire design) of the subsequent second study, so it can be improved accordingly, as shown in Figure 12.

![Figure 12. The exploratory sequential mixed methods approach used for the first two studies](image-url)
The first ecological quasi-experiment (Chapter 3) selects three real-world websites chosen from the same domain to evaluate different levels of interactivity. Two bespoke websites will be evaluated in the second controlled study (Chapter 4), so any confounding variables (e.g., design, content) can be controlled, allowing to focus on the specific interactive features contained within the sites (e.g., avatar and videos), thus supporting and strengthen the findings from the first study.

The rationale for the third longitudinal study (Chapter 5 & 6) is to explore how interactivity and other variables (e.g., usability, aesthetics) may change over time. The study will adopt a multiphase design due to its longitudinal nature (Creswell & Clark, 2011), where a mix of concurrent research methods will be employed, (e.g., questionnaires, interviews and a diary study). Data collection and analysis will be broken into separate time frames and analysed using the sequential method similar to the first two studies, then each time frame is compared to generate the overall findings, as shown in Figure 13.

The longitudinal study addresses the lack of long-term UX research as identified by (Bargas-avila & Hornbæk, 2011; Kujala et al., 2011). There are limited opportunities to capture UX over long-term, and the iPad initiative (where third year Manchester Medical School students were given an iPad to support their learning during their first year out in hospital placements) provides a unique research opportunity to capture UX over time. This 6-month study aims to track users’ transitional experiences and activities of product use (iPad), from their initial perception (prior-use) through to product adoption or non-adoption. The iPad was considered a relatively new interactive device to focus upon, and provides a contrast to the first two studies, with a focus on ecological usage within a real-world setting and therefore should provide an interesting longitudinal UX study.

![Figure 13. The multiphase mixed methods design used for the longitudinal third study.](image-url)
2.10.2. Measures

Questionnaires are one of the most useful tools to assess attitudes towards a given product, as they are simple, quick and easy to apply (Devellis, 2003). Self-reporting Likert-type scales are the most commonly used measure, providing a range of quantitative responses to a given statement or question, typically on a 1-5, or 1-7 scale. However, their use has been criticised within the research community due to the confusion over the classification of the scale (Jamieson, 2004). Unlike nominal-level measures (unordered categories), Likert scales fall into the ordinal-level measures that contain categories that are “ordered along the same continuum”. However, Blaikie, (2003) contends, these categories are “not necessarily evenly spaced” (P.23), and Cohen, Manion, & Morrision, (2007) considers it an “illegitimate inference” (P.32) to assume that the intensity of feeling between ‘strongly agree’ and ‘agree’ is equivalent to all the other consecutive items within a typical Likert scale. The assumption that Likert scales are considered more as interval-level data is an important issue, as ordinal-data analysis (non-parametric) differs from interval-data analysis (parametric), where the use of parametric statistics to analyse Likert scales and been seen as poor practice. This view has been fiercely challenged as a misconception, as parametric methods can be utilised with a high degree of robustness (Carifio & Perla, 2007; Norman, 2010).

Similarly, there has been much debate regarding the best optimal number the scale should contain, with an overall consensus that 5 or 7-item are favoured. Furthermore, the interpretability of the scales middle position has been deliberated, with some suggesting it allows the respondent to give ‘no opinion’, while others content that removing this option (4 or 6-item scale) may force respondents to give an inaccurate result (Krosnick, Judd, & Wittenbrink, 2005; Krosnick, 1991). A mix of simple statement scales taken from existing sources and bipolar scales will be used within this research (see Appendix 2 to 7 for Affect, Design Quality, Usability, Content, Quality, Immersion & Presence, Usefulness, Overall Satisfaction, and Diary Study scales), administered using both pen and paper and an online survey tool (e.g., www.qualtrics.com). The semantic differential scales use single bipolar words that are quantified (numbered) to provide minimal semantic processing (cognitive effort) in order to illicit an immediate response (Devellis, 2003), while single statements will be kept short and concise. It is essential that Likert-type scales clearly capture the attitudes and opinions of the variable construct investigated, so where possible the use of well-established and reliable scales are adopted within this research. These include the AttrakDiff2 Scale (Hassenzahl & Monk, 2010), classical and expressive aesthetics scale (Lavie & Tractinsky, 2004), adapted immersion (Jennett et al., 2008) and presence scales (Witmer & Singer, Michael, 1998): and selected usefulness items from Technology Acceptance Model (TAM) scale (Davis, 1989).
2.11. Chapter Summary

User experience came from the recognition that usability did not account for the subjective nature of the experience of interacting with a technology. This brought a shift in HCI, which had traditional focused on the instrumental (task efficiency and learnability), to encompass the more non-instrumental aspects, such as the emotional, affective and hedonic qualities of use. The importance of aesthetics brought about much research into how various quality factors may influence user judgement, such as context and task.

There are a wide variety of theoretical UX frameworks and models that employ multiple constructs used to evaluate and measure UX. However, these models differ in their approach and manipulation of constructs, making comparisons difficult. The general consensus is that usability (pragmatics), utility (usefulness), aesthetics (beauty) and hedonics (satisfaction) are important, but context dependant, and their relative importance is not well understood. This research aims to investigate further these constructs in relation to user quality judgements of technology, along with in less well-researched factors of user engagement, interactivity, user difference and the temporal nature of UX.

It has been recognised that interactivity of a website is an important factor in engaging users, leading to increased satisfaction (Bargas-avila & Hornbæk, 2011). However, few studies have investigated the influence of interactivity in relation to usability, hedonics and UX. This research aims to understand what makes users engage with interactive technology, through the use of different interactive design features (3D environments, avatars), by exploring various constructs such as affect, flow and presence, more commonly researched in games (Jennett et al., 2008)). The relationship between interactivity and other quality ratings, such as usability, utility and aesthetics, will also be explored.

Few studies have explored how individual differences may influence UX, although it has been recognised it is an important factor in influencing user judgement (Arhippainen, 2003; Hartmann et al., 2008; Karapanos & Martens, 2007). This thesis aims to explore diversity within groups in order to avoid the generic averaging of findings, so greater clarity of user judgements can be gained at an individual level. There are few studies that have investigated the temporal nature of UX, with the majority of research focusing on short-time durations using few participants (Karapanos, Zimmerman, Forlizzi, & Martens, 2009b; Zhang et al., 2008). This research aims to conduct a mixed methods longitudinal research in order to explore the changing patterns of UX over time in relation to product satisfaction.
Chapter 3  Study One: Interactivity and Affect

This chapter outlines the first of three studies conducted as part of this thesis. It focuses upon the quality judgements of users interactive experiences at different stages (prior, within and post) during a session. The aim of the study is to explore affect, flow and presence through the responses to various interactive features and how this impacts upon other user quality judgement constructs (e.g., usability, aesthetics, hedonics).
3.1. Introduction

The basis for this study is to enhance the existing model of quality judgement by Hartmann et al., (2008) by adding interactivity. This multi-attribute model is based on the cognitive theories of judgement and decision making, adapted from Payne et al’s., (1993) Adaptive Decision Making Theory (ADM). The user selects different decision-making strategies depending on the task (goal) and context (seriousness), as user judgement depends on the prioritisation of various product qualities (usability, aesthetic, brand, etc.). Users also adopt trade-off strategies, for filtering or comparing different criteria, depending on their background experience and the criticality of the decision (Hartmann et al., 2008). Similar trade-off strategies are also found in (Diefenbach & Hassenzahl, 2009) between beauty and usability.

In this first study interactivity will be considered in relation to Hartmann’s et al., (2008) theory, as interactivity has been recognised as an important factor in attracting and engaging users in games and entertainment (Jennett et al., 2008), cultural spaces (Haywood & Cairns, 2006) and e-commerce websites (O’Brien, 2010). Although flow and presence are considered useful measures for immersion and engagement within the gaming domain (Jennett et al., 2008), they have not been used to evaluate the influence of interactivity on user experience in relation to specific interactive design features. This thesis aims to investigate how interaction may contribute to user experience, and the objective of this first study is to test the hypothesis that highly interactive websites provide positive user engagement. To maintain engaging experience during interaction users need to maintain high arousal and positive affect through interesting, stimulating and exciting interfaces. Positive affect has been shown to enhance problem solving and decision-making (Isen, 2001), and influence users’ intention to use an IT system (Zhang & Li, 2004). Therefore, this study posits that different interaction designs may promote positive affect and arousal through the serendipitous use of avatars and virtual environments, thus generating positive UX. It aims to test the following hypothesis:

H1: Websites with enhanced interactive features will provide a more positive user experience than standard websites containing standard interactive features.

It will be guided by the following research question:

RQ. How do different types of interactive features impact on user preference and design quality judgement?
Three websites that contain different interactive features are used to evaluate the hypothesis that interactive design features will prompt positive affect and generate positive user experience; and that the intensity of experience will increase after repeated exposure. The influence of other quality constructs such as aesthetics, usability, and brand, upon user judgement of quality will also be considered.

3.2. Materials

Three different real-world interactive art gallery websites were selected for their variation of interactive features: a traditional menu-link navigation interface verses two more audacious designs that use interactive guides, animations and 3D effects (see Figure 14). The application context of all three websites was in the same domain (art galleries), although they varied in interactive features employed, metaphor types, design and layout, as shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Interactive Features</th>
<th>Interactive Paradigm</th>
<th>Design Layout</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>National Gallery</strong></td>
<td>(Standard)</td>
<td>Menu-Link Navigation</td>
<td>Hyper-link Navigation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Traditional Grid Style</td>
</tr>
<tr>
<td><strong>Google Art Project</strong></td>
<td>(Enhanced)</td>
<td>3D Environment (Street View)</td>
<td>Egocentric Navigation</td>
</tr>
<tr>
<td><strong>Louvre</strong></td>
<td>(Enhanced)</td>
<td>Interactive Graphical Objects</td>
<td>Animated-Link Navigation</td>
</tr>
</tbody>
</table>

*Table 1. Summary of the three website features selected for the study*

The Louvre\(^1\) combined interactive graphical links with audio story telling that used 2D animation. These were provided through the metaphor of an avatar guide (a cartoon character) and interactive graphical objects (cartoon animated objects), that moved when users placed their mouse over them. Users could interact with the animated objects by clicking on them to navigate the gallery website. When any given object was activated the avatar guide gave short audio animated stories about the artefacts within the gallery space. The cartoon avatar had limited interactivity and only really acted as a guide as users could not ask questions. Users had the option to turn off the animated guide allowing them to freely navigate the website through a more traditional menu-based system.

\(^1\) [http://www.louvre.fr/en](http://www.louvre.fr/en) (Site design has since changed with avatar and animated links removed)
The Google Art Project\(^2\) differed from both the Louvre and National websites as it acted as a portal for a number of galleries. It was chosen for its use of Google Street Map technology that allows users to navigate in a 3D space using either the arrow keys on their computer or on the screen. This interactive fly-through metaphor created the feeling of a virtual 3D environment where users can walk around and interact with both 2D images within the gallery space.

The National Gallery\(^3\) website offered limited interactive features compared to both the Louvre and Google Art Project websites, so no expectations were sought and thus acted as a control. It provided a simple menu-link metaphor with a traditional grid style layout. All three websites differed in colour and layout. For example the National used a limited colour palette consisting of charcoal grey on a predominantly white background. The Louvre offered a muted colour palette for the main website, yet this was offset against the very colourful cartoon animations. The Google Art Project adopted an atypical layout that had a very large colourful image that filled the whole screen. The menu system was small in scale and streamlined, only occupying the top right hand corner of the home page. The internal pages followed the similar limited menu style, with options to hide or view the menu to allow for a full view of the gallery space or picture image. The colour scheme was predominately dark, as it used a dark grey background making it a stark contrast to both the Louvre and National (see Figure 14 for a visual comparison).

The Google Art Project website design style appeared more spacious, with minimal text compared to both the Louvre and National sites. This could be due to it being a web portal, where not all the information of a typical gallery website needed to be present. For example both the Louvre and National had pages relating to current exhibitions and visiting the gallery, this was not present within the Google Art Project. It is important to note that a variety of other factors (such as design, layout, familiarity etc.) may influence user judgement, although the focus of the study is to compare the individual interactive features with the sites in order to test the hypothesis’ that highly interactive websites promote positive UE.

\(^2\)https://www.google.com/culturalinstitute/project/art-project

\(^3\)http://www.nationalgallery.org.uk/
Chapter 3  Study One: Interactivity and Affect

The National Gallery Website: The left image is the home page, which shows the traditional grid style layout, and the right image is a gallery page that follows the same grid layout.

The Louvre Website: The home page (left) shows the animated avatar (cartoon character) within a room of various interactive objects. The second image shows an internal page (right) displaying the traditional menu layout style, with the animated avatar guide shown in the center of the page.

The Google Art Website: The home page (left) shows a large close-up image of a painting with a reduced dark grey menu box. The interior page shows the 3D virtual space with a limited top menu.

Figure 14. The three interactive websites used within the study. The first is the National Gallery website which acted as a control; the second is the Louvre, which provides an avatar (animated cartoon audio guide), and the third is the Google Art Project, a virtual 3D environment where users can navigate through the gallery space.
3.3. Method

The study was designed to test the hypothesis that websites with more interactive features will provide more positive user experience than the traditional menu-link design. To explore this a triangulation of mixed methods were employed, combining quantitative questionnaires with qualitative interviews and participant observations.

3.3.1. Experimental Design

The three websites were used in a within subjects experimental design, where the presentation order of the websites were counterbalanced across the participant group. A three-way repeated-measures design: 3 (websites) x 3 (tasks) for the affect scale only, and a 3x2 for website quality and criteria importance scale, and 3x1 for all other scales was used (see Table 2). The three activities were initial exposure (visual perception) Task One (initial interaction) and Task 2 (extended interaction) and remained constant for each site.

<table>
<thead>
<tr>
<th></th>
<th>Initial Exposure (Visual Perception)</th>
<th>Task One (Initial Interaction)</th>
<th>Task Two (Extended Interaction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Gallery</td>
<td>XX</td>
<td>X</td>
<td>XXX</td>
</tr>
<tr>
<td>Google Art Project</td>
<td>XX</td>
<td>X</td>
<td>XXX</td>
</tr>
<tr>
<td>Louvre</td>
<td>XX</td>
<td>X</td>
<td>XXX</td>
</tr>
</tbody>
</table>

*X = Affect Scale only (3x3), XX=Website Quality & Criteria Importance, XXX= All Other Scales (2x3)

Table 2. The 3 x 3 factorial experimental design

3.3.1.1. Procedure

The experimental procedure is illustrated in Figure 15, and described as follows:

1. **Briefing and Familiarisation:** after reading the information sheet, completing the consent form and demographic questionnaire, participants were given a short training session by showing them a similar art gallery website homepage (Metropolitan Museum Art Gallery), then asking them to complete the affect questionnaire. This prepared them for the initial exposure task (see no. 2) and allowed them to become familiar with the affect scale.
2. **Initial Exposure:** screen shots of each of the three websites homepage were captured from a Safari browser with a resolution of 150 dpi (1500x1030) as shown in Figure 14. These were presented to the users for 0.5 seconds using PowerPoint, which controlled the exposure time by replacing the image with a mask (grey screen). The purpose of masking was to cancel any image trace that enables further cognitive processing of the target image, as proposed by Lindgaard, Dudek, Sen, Sumegi, & Noonan, (2011). The sequence of exposure was counterbalanced across all the participants, and after each exposure participants completed both the affect and website quality scales.

3. **Task One:** users were asked to first visit the homepage and then explore one interactive feature in each website. The interactive feature for the Louvre was to follow the avatar guide to the ‘interactive room’ and click on an active object. For Google, users were asked to navigate the 3D fly-through space by clicking on the option to ‘explore the museum’, (which created the 3D view), and then navigate to the ‘Hall of Mirrors’ using the floor plan. The National Gallery task directed users to the interactive floor plan of the gallery, by selecting ‘Paintings Room by Room’ and then to select a specific art gallery room of their choice, (see Appendix 1 for more detail on all tasks).

4. **Task Two:** this exposed users to additional interactive features in each site and required a more detailed exploration of a specific art work. In the Louvre users navigated to a picture of Napoleon and then used the ‘view feature’, which gave users the option to zoom-in and listen to an audio voice-over explanation. In Google Art Project users was asked to navigate in fly-through mode to a painting (Brugel) and then select the 2D display of the painting, enabling users to zoom in and access it in more detail. In the National Gallery, users were asked to navigate via the floor plan map to ‘room 45’, then select the thumbnail image of a painting (Rousseau) and use the zoom to view it in more detail. All sites followed the same counterbalanced sequence of the initial exposure of the websites homepage. Tasks were given via written instruction sheets prior to each task, and were used to ensure participants experienced the same website interactive features. After completing each task (one & two), participants were given up to 3 minutes to explore anywhere within the site, without direction. Measures collected were: affect (3 times), website quality (2), immersion (1), usability (1), and overall experience (1). Criteria importance was collected twice, after each task.

5. **Interview:** a short interview using semi-structured questions was used to elicit participants’ preferences and experiences while interacting with each website (see Appendix 8). During the interview participants were asked to list the first two most remembered things relating to any of the three sites. At the end of the interview they were asked to rank the websites in preference order, which completed the study design.
The experiments took place within a university audio-visual lab so the task navigation time could be recorded. They were carried out over the summer of 2011 and all participants received a £10 Amazon gift voucher for their participation.

3.3.1.2. Participants

Forty participants (21 female) participated in the study, with age ranging from 18-25 (28%), 26-35 (53%), 36-45 (18%) and 46-55 (3%). The majority of participants were students (70%), while the remaining were university staff. All participants were educated to degree level and came from a variety of subject areas including business (37%), business computing (8%), computing (8%), humanities (25%), and science (23%).

3.3.2. Measurement Scales

Apart from the demographics, six questionnaire scales were administered and repeated after various tasks (see Figure 15), with all scales employing a 7-point Likert scale rating. Each of the questionnaire scales will now be discussed in further detail.

3.3.2.1. Demographics and Disposition

Demographic information such as age, gender, employment, education and degree subject were captured at the start of the study. Further questions checked for awareness of brand and brand quality rating for the three websites. Artistic preference checked for participants’ general interest in art, if they visited art galleries and how they rated themselves creativity.

To explore individual aesthetic disposition an adapted version of the Centrality of Visual Product Aesthetics (CVPA) scale was employed (Bloch, Brunel, & Arnold, 2003). This 11-item scale was used to determine if aesthetic disposition of a website had an impact on user experience.

3.3.2.2. Affect Scale

The affect scale aims to capture the first affective response to the website, which often occur without extensive perceptual or cognitive processing, and are made much quicker than cognitive encoding (Zajonc, 1980). So the initial affective responses can be made with greater confidence and so can potentially sway the cognitive user judgement (Derbaix & Pham, 1991).

An 9-item simple bi-polar affect scale was developed that focused on one emotive word in order to capture value-charged affect straight after each task (Karapanos, Martens, & Hassenzahl, 2010). The emotive words were drawn from a mix of sources on arousal, hedonics and emotion, as shown in Table 3. This 7-point semantic-differential scale was developed to be simple, relevant and quick to elicit affective responses straight after user exposure or interaction, and was preferred over the PANAS (Watson, Clark, & Tellegen, 1988).
scale, which has been considered to be somewhat ambiguous (Russell & Carroll, 1999). In order to capture change in affective responses over time this scale was completed three times after each task (initial exposure, task 1 & 2).

3.3.2.3. Website Quality

A 17-item semantic-differential 7-point scale was used to measure various website quality ratings, as shown in Table 4. This comprised of 11 items taken from the popular AttrakDiff2 scale (Hassenzahl, 2004). These include three subscales; Hedonic Quality-Stimulation (HQS), Hedonic Quality-Identification (HQI), and the Pragmatic Qualities (PQ) contained 3 items in each, and two global Evolutional Constructs (EC), of Goodness and Beauty.

Table 3. The Affect Scale individual items and origins (see Appendix 2 for example of Affect Scale)

<table>
<thead>
<tr>
<th>Affect</th>
<th>1-7 Scale</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mood</td>
<td>Very Calm</td>
<td>Very Excited</td>
</tr>
<tr>
<td>Fun</td>
<td>Not Fun</td>
<td>Highly Fun</td>
</tr>
<tr>
<td>Attractive</td>
<td>Not Attractive</td>
<td>Very Attractive</td>
</tr>
<tr>
<td>Curious</td>
<td>Not Curious</td>
<td>Very Curious</td>
</tr>
<tr>
<td>Interesting</td>
<td>Not Interesting</td>
<td>Very Interesting</td>
</tr>
<tr>
<td>Pleasurable</td>
<td>Not Pleasurable</td>
<td>Very Pleasurable</td>
</tr>
<tr>
<td>Absorbing</td>
<td>Not Absorbing</td>
<td>Very Absorbing</td>
</tr>
<tr>
<td>Engaging</td>
<td>Not Engaging</td>
<td>Very Engaging</td>
</tr>
</tbody>
</table>

Table 4. The Website Quality sub-scale items and origins (see Appendix 3 for Website Quality Scale)

<table>
<thead>
<tr>
<th>Website Quality</th>
<th>1-7 Scale</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>HQS Dull</td>
<td>Dull</td>
<td>Captivating</td>
</tr>
<tr>
<td>Conservative</td>
<td>Conservative</td>
<td>Innovative</td>
</tr>
<tr>
<td>Unimaginative</td>
<td>Unimaginative</td>
<td>Creative</td>
</tr>
<tr>
<td>HQI Cheap</td>
<td>Cheap</td>
<td>Premium</td>
</tr>
<tr>
<td>Tacky</td>
<td>Tacky</td>
<td>Stylish</td>
</tr>
<tr>
<td>Amateurish</td>
<td>Amateurish</td>
<td>Professional</td>
</tr>
<tr>
<td>PQ Confusing</td>
<td>Confusing</td>
<td>Clearly Structured</td>
</tr>
<tr>
<td>Complicated</td>
<td>Complicated</td>
<td>Simple</td>
</tr>
<tr>
<td>Unpredictable</td>
<td>Unpredictable</td>
<td>Predictable</td>
</tr>
<tr>
<td>EC Goodness</td>
<td>Bad</td>
<td>Good</td>
</tr>
<tr>
<td>EC Beauty</td>
<td>Ugly</td>
<td>Beautiful</td>
</tr>
<tr>
<td>PDQ Attention</td>
<td>Hardly At All</td>
<td>All The Time</td>
</tr>
<tr>
<td>Colour</td>
<td>Poor Colour</td>
<td>Good Colour</td>
</tr>
<tr>
<td>Images</td>
<td>Poor Images</td>
<td>Good Images</td>
</tr>
<tr>
<td>SDQ Harmonious</td>
<td>Design Discord</td>
<td>Design Harmony</td>
</tr>
<tr>
<td>Balanced</td>
<td>Balanced Design</td>
<td>Balanced Design</td>
</tr>
<tr>
<td>Spacious</td>
<td>Cluttered</td>
<td>Spacious</td>
</tr>
</tbody>
</table>

4 http://attrakdiff.de/index-en.html
Two further sub-scales were supplemented to these measures: Perceptual-Design Quality (PDQ), and Spatial-Design Quality (SDQ), each containing 3 items. SPD contained Attention, which was chosen to capture a user’s first impression (and was the only non bi-polar item). How well a website grabs a user's attention is a result of an individual’s sensory curiosity being aroused (Berlyne, 1960; Trevino & Webster, 1992), which can vary depending on the visual impact of the product, or the first impressions (Lindgaard, Fernandes, Dudek, & Brown, 2006). The remaining two PDQ items were general ratings of colour and image quality, inspired from Kim, Lee, & Choi, (2003). The SDQ sub-scale contained Harmonious, which were drawn from the gestalt tradition within cognitive psychology, which states that humans visually prefer things that are close together or that look as if they belong together, (Wertheimer, 1938), along with the visual qualities Balanced and Spacious, (Kumar & Garg, 2010; Lindgaard et al., 2011). To compare first impressions with post interaction this scale was applied twice.

3.3.2.4. Immersion and Presence

The 8-item scale was derived from a variety of measurements used to capture immersion, flow and presence, as shown in Table 5. Measures were derived from Csikszentmihalyi, (1975, 1988) who described flow as the state when people feel most enjoyment when they are fully focused on a stimulating and challenging task; and Witmer & Singer, Michael, (1998) who described presence, as the subjective experience of feeling present within one place, even when they may be physically situated somewhere different. Immersion is seen as more common, which can occur in varying degrees, in which both flow and presence contribute (Jennett et al., 2008). The combined immersion and presence scale were short questions (non bi-polar), and was applied once after interaction following the usability scale (see Figure 15)

<table>
<thead>
<tr>
<th>Immersion/Presence</th>
<th>1-7 Scale</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did Pace keep you interested</td>
<td>Not Interested</td>
<td>Interested all the time</td>
</tr>
<tr>
<td>Did you feel in Control</td>
<td>Not in Control</td>
<td>In Control all the time</td>
</tr>
<tr>
<td>How Challenged were you</td>
<td>Very Difficult</td>
<td>Very Easy</td>
</tr>
<tr>
<td>How fast did the Time go</td>
<td>Time was Slow</td>
<td>Time Flew</td>
</tr>
<tr>
<td>How Aware of other events</td>
<td>Very Distracted</td>
<td>Not Distracted</td>
</tr>
<tr>
<td>How Natural did you feel</td>
<td>Not Natural</td>
<td>Very Natural</td>
</tr>
<tr>
<td>How Involved did you feel</td>
<td>Not Involved</td>
<td>Very Involved</td>
</tr>
<tr>
<td>How Compelling was the site</td>
<td>Not Compelling</td>
<td>Very Compelling</td>
</tr>
</tbody>
</table>

**Table 5. The Immersion/Presence Scale questions and origins**

3.3.2.5. Usability Scale

An 4-item scale was used once to capture users perception of the functionality, navigation, learnability, and convenience while using the websites, in order to determine their overall
usability (Lavie & Tractinsky, 2004; Tractinsky & Zmiri, 2006). The 7-point (agree/disagree) scale was administered once after interaction (see Table 6).

<table>
<thead>
<tr>
<th>Usability</th>
<th>1-7 Scale</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convenient to Use</td>
<td>Strongly Disagree</td>
<td>Lavie &amp; Tractinsky, 2004</td>
</tr>
<tr>
<td>Clear Functionality</td>
<td>Strongly Disagree</td>
<td>Tractinsky &amp; Zmiri, 2006</td>
</tr>
<tr>
<td>Easy to Learn</td>
<td>Strongly Disagree</td>
<td>Tractinsky &amp; Zmiri, 2006</td>
</tr>
<tr>
<td>Easy to Navigate</td>
<td>Strongly Disagree</td>
<td>Lavie &amp; Tractinsky, 2004</td>
</tr>
</tbody>
</table>

*Table 6. The Usability Scale questions and origins (see Appendix 4 for example of Usability Scale)*

3.3.2.6. Overall Experience

An 3-item scale asked if participants would, *visit the site again; recommend the site to their friends* and how *stimulating their experience* was while using the site (O’Brien, 2010).

Participant free-exploration time during both tasks was recorded for further analysis, and at the end of the post-test interview participants were also asked to rate the sites in order of overall preference.

3.3.2.7. Criteria Importance

Participants were (verbally) asked to rate (on a 1-7 point scale) the importance of five criteria, *aesthetics, ease-of-use, content, engaging interaction* and *brand*, in influencing their overall judgement of the quality of the website to gauge if any of these criterion influenced their overall preference. This was administered twice, after initial exposure and then after task two.

3.4. Quantitative Analysis and Results

The analysis is split into two main areas; the statistical analysis (this section), which was conducted on the results from the questionnaires; and the qualitative analysis (Section 3.5), which used the data drawn from the participant interviews and observation while interacting with the websites.

The quantitative results were analysed using SPSS and report on the main questionnaire constructs used; affect, website quality, immersion (flow & presence), usability, overall experience, criteria quality judgement and exploration time. The general demographics were presented earlier (3.3.2.1.) while *Participant Disposition* (also included within the demographic questions) will be discussed in relation to the affect scale results.
3.4.1. Reliability

Questionnaire scores (affect, website quality, immersion and presence, usability and overall experience) were aggregated (averaging individual item scores), since the individual scales all produced high levels of internal reliability with Cronbach alpha (α) ranging from α=0.89 to 0.96, as shown in Table 7. Since Nunnally, (1978) suggests that correlations of 0.80 or higher are very good.

<table>
<thead>
<tr>
<th>Questionnaire Scales</th>
<th>National</th>
<th>Google</th>
<th>Louvre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>α=0.94</td>
<td>α=0.95</td>
<td>α=0.94</td>
</tr>
<tr>
<td>Website Quality</td>
<td>α=0.95</td>
<td>α=0.96</td>
<td>α=0.96</td>
</tr>
<tr>
<td>Immersion/Presence</td>
<td>α=0.93</td>
<td>α=0.89</td>
<td>α=0.94</td>
</tr>
<tr>
<td>Usability</td>
<td>α=0.94</td>
<td>α=0.97</td>
<td>α=0.96</td>
</tr>
<tr>
<td>Overall Experience</td>
<td>α=0.95</td>
<td>α=0.97</td>
<td>α=0.94</td>
</tr>
</tbody>
</table>

*Table 7. Internal reliability scores for questionnaire scales using Cronbach coefficient*

The affect scale was the only scale not taken from other validated sources, so its reliability was further investigated. It is recognised that a confirmatory factor analysis could be used to improve the validity of the affect scale, due to the small sample size (N=40) used. However, in certain conditions (low factor ratings and high loadings) using an exploratory analysis can yield reliable results (de Winter, Dodou, & Wieringa, 2009). A principal component analysis was run on the affect scale items using oblimin rotation to explore any cross-loadings. The analysis revealed that the affect scale was composed of one factor and had high loadings (0.6 to 0.9) for all three activities (initial exposure, task 1 & 2), as shown in Table 8.

<table>
<thead>
<tr>
<th>Affect Scale Item</th>
<th>Principle Components (X3 Websites)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor 1</td>
</tr>
<tr>
<td></td>
<td>Initial Exposure</td>
</tr>
<tr>
<td>Current Mood</td>
<td>.904</td>
</tr>
<tr>
<td>Fun</td>
<td>.900</td>
</tr>
<tr>
<td>Attractive</td>
<td>.900</td>
</tr>
<tr>
<td>Curious</td>
<td>.882</td>
</tr>
<tr>
<td>Interesting</td>
<td>.855</td>
</tr>
<tr>
<td>Pleasurable</td>
<td>.843</td>
</tr>
<tr>
<td>Absorbing</td>
<td>.829</td>
</tr>
<tr>
<td>Exciting</td>
<td>.764</td>
</tr>
<tr>
<td>Engaging</td>
<td>.669</td>
</tr>
<tr>
<td><strong>Engenvalue</strong></td>
<td><strong>6.15</strong></td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td><strong>68.3%</strong></td>
</tr>
</tbody>
</table>

*Table 8. Component matrix table for Affect Scale factor analysis across the three activities*

3.4.2. Affect Scale

The 9-item affect scale was analysed in two parts, first as an aggregated scale and then each item within the scale was explored individually.
3.4.2.1. Aggregated Affect Scale

A two-way repeated measure ANOVA was carried out on the aggregated Affect scales using website (3) and tasks (3) as within subject factors. Mauchly’s test indicated that the assumption of sphericity had been violated against task \(X^2(2)= 18.7, p < .001\) and the interaction effect between task * site \(X^2(9)= 45.8, p < .001\), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity \([\varepsilon = .72], [\varepsilon = .71]\) respectively. The results revealed there was a significant main effect of the task; \(F(1.4, 56.2) = 51.2, p < .001, \eta^2 = .57\), the website; \(F(2, 74.8) = 4.5, p < .05, \eta^2 = .1\), and the interaction between task * site \(F(3.9, 111.1) = 7.7, p< .001, \eta^2 = .16\), see in Figure 16.

Post hoc tests using Bonferroni correction indicated that participants’ affective responses increased significantly \((p < .001)\) for both the Google Art Project (GA) and the Louvre (L) site from Initial Exposure to Task 1, whereas no difference emerged for the National Gallery (NG) site, see Figure 16. Although the means shown in Table 9 indicate a slight decrease from Task 1 to Task 2 in both the Google and Louvre sites, there was no significant difference shown between Google/Louvre and National \((p=.74)\), despite both still showing higher mean scores.

3.4.2.2. Individual Affect Scale Items

Each individual item within the Affect scale was explored to find any interesting patterns. A comparison between Initial Exposure, Task 1 and Task 3 aggregated results (all three websites) is shown in Figure 17. Participants were more Curious and Interested in the
websites on initial exposure, which became less evident post interaction (Task 1 & 2), when all ratings clearly showing an increase as a result of interaction.

A repeated measure (3x3) ANOVA was carried out on each of the individual items within the Affect scale and all items showed a significant main effect for task \( (p<0.001) \) and interaction \( (p<0.01 \text{ to } p<0.001) \). Post hoc tests using Bonferroni adjustment showed all items proved significant from Initial Exposure to Task 1 \( (p<0.001) \), and from Initial Exposure to Task 2 \( (p<0.05 \text{ to } p<0.001) \). This suggests that when participants interacted with any of the websites their affective responses were enhanced through their interactive experience. No significant effect was found between Task 1 and Task 2, apart for Current Mood, which showed only a weak significant \( (p<0.05) \) difference. As participants’ affective responses did not change from task one to task two, apart from their overall mood, no effect of repeated exposure was apparent.

Four items showed significant main effect between websites, these were Current Mood: \( F(2, 78) = 7.54, \ p< .001, \) Fun: \( F(2, 78) = 11.27, \ p< .001, \) Curious: \( F(2, 78) = 4.7, \ p< .05, \) and Exciting: \( F(2, 78) = 4.43, \ p< .05. \) Post hoc showed significant differences \( (p<.05 \text{ to } p<.001) \) between National and Google/Louvre, but no significant differences were found between Google and Louvre. Comparing the mean graph patterns (see Figure 18), it clearly shows that on initial exposure the ratings for mood, fun and exciting were very similar across the three sites, but after interaction both Google and Louvre ratings increased far more than the National gallery. The Louvre website was initially favoured (after Task 1), but this dropped after Task 2 and became equal to Google.
The same pattern was seen in the item *absorbing* (although not significant for task), it clearly shows an increase post interaction, but little change after task two (see Figure 19). However interestingly, on initial exposure Google was rated higher for *curiosity* than the National and Louvre, which could be due to its atypical homepage layout. After interaction it followed the same pattern as the previous items (Task 1 Louvre was favoured, Task 2 no difference between Google or Louvre).

The remaining four items *attractive, interesting, pleasurable* and *engaging* show an interaction effect where the National was rated the highest on initial exposure but after interaction was favoured the least (see Figure 20). Both Louvre and Google were rated much higher than National post interaction (Task 1 & 2), with the Louvre site being slightly favoured over Google (although these showed no significant effect for site).
3.4.2.3. Summary of Affect Scale Results

Overall participants’ affective responses significantly increased for all three websites after interaction (from initial exposure to task one), and higher affect ratings were given for both Google and Louvre after interaction (task one), indicating that users’ affective responses increased as a result of their interactive experience, with the increase being shown far more for the interactive sites (Google and Louvre), due to the interactive features they offered.

No time effect was captured during interaction (from task 1 to task 2) for affect, apart from the individual item mood. Significant inter-site differences was found on the single items, mood, fun, curiosity and exciting, with both Google and Louvre being rated higher than National post-interaction, despite National receiving significantly higher ratings on initial exposure for attractive interesting, pleasurable and engaging, which could be attributed to its familiar style homepage; while Google’s atypical homepage resulted in higher curiosity ratings. Overall Louvre had slightly higher affect ratings than Google (although not significant), which was more apparent straight after interaction (task 1), but decreased over time (task 2).
3.4.3. Participant Disposition

The same 3x3 repeated measures ANOVA was conducted with gender as a between group factor, and no significant gender effect was found. Data drawn from the questions on prior knowledge, brand awareness, artistic appreciation and aesthetic disposition were analysed.

3.4.3.1. Prior Knowledge and Brand Awareness

To determine the participants’ prior knowledge and brand awareness three questions asked, (1) how well they knew the two art galleries, (2) what knowledge they have of the three chosen websites, and to (3) rate the brand quality of all three websites. The results are shown in Table 10, which displays the mean, standard deviation and number of respondents for each question across all three websites.

<table>
<thead>
<tr>
<th>Question</th>
<th>National</th>
<th>Louvre</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sig</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) Art Galleries</td>
<td>x</td>
<td>3.98</td>
<td>4.18</td>
</tr>
<tr>
<td>(2) Websites</td>
<td>x</td>
<td>1.68</td>
<td>1.58</td>
</tr>
<tr>
<td>(3) Brand Quality</td>
<td>√</td>
<td>3.88</td>
<td>4.65</td>
</tr>
</tbody>
</table>

Table 10. The sig. results, mean (M) and standard deviation (SD) for Brand Awareness

A one-way repeated measure ANOVA was carried out on each of the three scales and the only item that proved significant across site was, brand quality $F(2, 78) = 30.22$, $p < .001$, $\eta^2 = .44$. Participants showed little difference in their awareness between the two art galleries, with the Louvre art gallery (M =4.18) being slightly better known than the National Gallery (M =3.98).

Participants awareness of the three websites were much lower than the art galleries, with again little difference between sites, where National (M =1.68) and Louvre (M =1.58) awareness showing marginally higher than the Google website (M =1.38).

Although there was a significance difference across the three websites of participants’ rating of the main brand quality, with Louvre s rated the highest (M =4.65) compared to the National (M =3.88) and the Google (M=1.4); the ratings per site varied depending on how well the sites were known by the participants. For example, when participants did not know the site a rating of 0 was given, so Google was the least known site (13 ratings), followed by the National (28) and the Louvre (30). The low ratings score for the Google Art Project could be attributed to the site not representing an art gallery along with its lack of familiarity, as it was a relatively new site at that time.
3.4.3.2. Artistic and Aesthetic Appreciation

Three questions evaluated participants’ art appreciation; (1) *rate their interest in art*, (2) *if they visited art galleries*, and (3) *if they considered themselves creative*. Just over half the participants were interested in art ($M = 4.75, SD = 1.27$) and considered themselves to be creative ($M = 4.78, SD = 1.27$), although their frequency of regularly visiting art galleries ($M = 4.03, SD = 1.61$) was lower.

To explore if the *artistic appreciation* of participants influenced their affective responses, each of the three questions were added to the repeated 3x3 ANOVA on affect as separate covariates, but no significant effect was found, suggesting that artistic appreciation had no impact on user experience recorded by the affect scale.

The 11-item adapted *aesthetic disposition* scale derived from (Bloch et al., 2003) proved very good ($\alpha = .853, M = 4.52, SD = 1.02$) for internal reliability using Cronbach Alpha. The aggregated result from this scale was added as a covariant to the same repeated measures 3x3 ANOVA, which proved non-significant. This indicates that *aesthetic disposition* had no impact on the affect scales.

3.4.3.3. Participant Disposition Summary

There was little effect of *brand awareness*, although the Louvre and National art galleries were the better-known websites, although there was no inter-site difference. However the Louvre *brand quality* was rated the highest, above the National brand, and Google being rated the lowest, which could be due the lack of familiarity of the Google Art Project site as it was a relatively new website. No significance effect was found on artistic appreciation, aesthetic disposition and other general demographic scales (e.g. gender).

3.4.4. Website Quality Scale

The 17-item *Website Quality* scale was analysed in two parts, first by its aggregated sub-scales, and then each item within the scale was explored individually to see if there were any interesting patterns.

3.4.4.1. Subscale Website Quality Analysis

A repeated measures (3x3) ANOVA was carried out the *Website Quality* sub-scales, with *website* (3) and *task* (2) as factors (*Initial Exposure & Task 2*). Mauchly's test indicated that the assumption of sphericity had not been violated, so no correction was required. There were significant main effect for *task* on all measures ($p < .001$ to $p < .05$) and *interaction* between *task* and *site*, (apart from PQ, see Table 11). Three items showed significant differences between websites, for Pragmatic Quality (PQ), $F(1.8, 72) = 7.57, p < .001, \eta^2 = .16$, Hedonic
Stimulation (HQS), $F(1.8, 70) = 6.51, p<.001, \eta^2 = .14$ and Spatial-design Quality (SDQ) $F(1.9, 74) = 3.49, p<.05, \eta^2 = .82$.

<table>
<thead>
<tr>
<th>Website Quality Subscale</th>
<th>Task</th>
<th>Web-site</th>
<th>Task * Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pragmatic Quality (PQ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Confusing, Predictable, Complicated)</td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Hedonic Identification (HQI)</td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>(Cheap, Tacky, Amateurish)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hedonic Stimulation (HQS)</td>
<td></td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>(Creative, Captivating, Innovative)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial-design Qualities (SDQ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Spacious, Balanced, Harmonious)</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Perceptual-design Qualities (PDQ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Attention, Colour, Image)</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Goodness (EC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beauty (EC)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(* *** $p>.001$, ** $p>.01$, * $p>.05$)

Table 11. Website Quality Scales showing significant main effects and interactions

There was a strong (task*site) interaction for HQI and HQS ($F(2, 77) = 4.8, p<.05, \eta^2 = .11$), with a weaker interaction for SDQ, PDQ and the EC items Goodness and Beauty ($p<.05$). All the interactions showed the same pattern as HQS, with all ratings showing an increase post interaction, and Google and Louvre showing far higher ratings (for Task 2) than the less interactive National site (see Figure 21).

Figure 21. Mean patterns for Hedonic Quality Stimulation (HQS) left, and Pragmatic Quality (PQ) right, showing significant difference between website (3) and task (2)
Chapter 3  Study One: Interactivity and Affect

The order of means for Task 2 and significant post hoc test differences are given in Table 12. Both Google and National were considered more creative, captivating and innovate (HQS), than the National, and Google was rated more spacious, balanced and harmonious (SDQ), than the National and Louvre. However National was considered to be clearer, simpler and more predictable (PQ) than Google and Louvre, see Figure 21. The rank order of means showed Google was ranked marginally higher for 4 items (HQI, HQS, SDQ & Beauty), compared to Louvre’s 2 items (PDQ and Goodness), although these differences were minor and non-significant.

<table>
<thead>
<tr>
<th>Website Quality Subscales</th>
<th>Order of Mean Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PQ (Pragmatic Quality)</td>
<td>N (5.3)</td>
</tr>
<tr>
<td>HQI (Hedonic Identification)</td>
<td>G (5.9)</td>
</tr>
<tr>
<td>HQS (Hedonic Stimulation)</td>
<td>G (5.9)</td>
</tr>
<tr>
<td>SDQ (Spatial Design Quality)</td>
<td>G (5.6)</td>
</tr>
<tr>
<td>PDQ (Perceptual Design Quality)</td>
<td>L (5.8)</td>
</tr>
<tr>
<td>Goodness (Evaluation Construct)</td>
<td>L (5.8)</td>
</tr>
<tr>
<td>Beauty (Evaluation Construct)</td>
<td>G (5.6)</td>
</tr>
</tbody>
</table>

(Based on Mean Scores Post-Interaction)

Table 12. Website Quality rank order means & post hoc sig. results (PQ & HQS)

3.4.4.2. Analysis of Individual Items within Website Quality

A two-way ANOVA (2x3) was carried out for each of the 17 items within the Website Quality scale and all items proved significant (p<.001 to p<.01) on the main effect of task (apart for complicated), as reflected in the subscale results (see Section 3.4.4.1), indicating that after interaction, participants ratings’ for the website qualities increased.

Six items showed significant differences (p<.001 to p<.05) between site, and eleven items showed significant differences (p<.001 to p<.05) for interaction. Post hoc tests for these items revealed six items that showed significant differences, and their mean patterns were analysed to reveal four different patterns; (1), where all three sites increased to the same level, (2), both Louvre and Google sites increased more than National, (3) Google increased more than National, and (4) National increased more than Louvre and Google. These four patterns are discussed in relation to the post hoc results as shown in Table 13.
Table 13. Post Hoc results show sig. differences between sites for 6 items within the Website Quality Scale, where 4 mean patterns are revealed.

Pattern One (All sites improve)
This pattern shows all three sites increased to the same level after interaction (Task 2), as illustrated by the item Balanced (see Figure 22). The National Gallery site was rated much higher than the Google or Louvre at first glance, yet after interaction all three websites were viewed the same. This was the most common mean pattern and was reflected in five other non-significant items; Goodness, Harmony, Professional, Colour and Image.

Pattern Two (Favours Google & Louvre)
The second most common pattern shows the ratings for both Louvre and Google increased far more than National after interaction, as demonstrated by the two significant items Creative and Innovative, see Figure 23. Two additional non-significant items showed similar patterns, Beautiful and Captivating. This indicates that the sites with more interactive features were found to be more innovative and captivating post-interaction.
Pattern Three (Favours Google)

The third pattern shows the Google site receiving higher ratings than both the Louvre and National, for both initial exposure and Task 2, as illustrated by the item spaciousness, see Figure 24. This can be attributed to the site’s simple design layout and its 3D navigation metaphor adopted by the site that enhances the sense of space as users can navigate in a third dimension. Although not significant, a similar pattern was found in the item stylish, although the differences between the sites were not so apparent.

Pattern Four (Favours National Gallery)

The last remaining pattern shows the National being favoured the most, across both initial exposure and task 2, for two items: confusing and predictable, as shown in Figure 25. The familiar homepage and lack of interactive features within the National could attribute to it being considered the less confusing website and most predictable of all three. Although both Google and Louvre were considered less confusing after interaction, the Louvre was rated
more unpredictable post-interaction compared to Google, which could be attributed to the multiple interactive features employed (e.g., avatar, interactive links, animated movies, audio guide).

![Figure 25. Mean Pattern Four: Shows National being rated higher for Clear Structure and Predictable, and Google being rated more unpredictable post-interaction](image)

### 3.4.4.3. Summary of Website Quality Results

Both Google and Louvre received higher quality ratings post-interaction than National, due to the interactive features they provide, as reflected in the subscale results (HQS & SDQ) and individual item patterns 2/3, with Louvre and Google being rated more creative and innovative than National (pattern 2).

Little difference was found between the Google and Louvre sites, with Google scoring marginally higher (for 4 subscale items) than Louvre, although these scores were minor, with only one item showing a weak significance (SQD). Only one single item (within SDQ), spaciousness (pattern 3) proving significant, which can be attributed to Google's simple homepage layout and the 3D navigation metaphor enabling participants to navigate in a third dimension, thus enhancing their sense of space.

The National was ranked the highest for PQ (usability), and rated less confusing and more predictable (pattern 4), than Google and Louvre, which could be due to its simple homepage grid structure. Although both Google and Louvre were considered less confusing post interaction, Louvre was rated less predictable, which could be due to the multiple interactive features offered within the site (avatar, animated movies, interactive objects, audio guide). Whereas Google provided only one feature (3D street view navigation), so once users interacted with the site they soon became familiar with the interactive style, and so it became more predictable.
3.4.5. Immersion and Presence

The aggregated 8-item scale used to capture immersion, flow and presence was analysed using a one-way repeated measure ANOVA and no significant difference was found across the three sites. Only one of the 8 items within the scale proved significant when tested individually, that was involved, $F(1, 39) = 74.7, p < .001$, $\eta^2 = .13$, where both the Louvre (M=5.75) and the Google site (M=5.78) were rated higher than the National (M=4.97) for involvement.

3.4.6. Usability Scale

The aggregated 4-item usability scale was used to capture users' perception of functionality, navigation, ease of learning and convenience of use. A one-way repeated measure ANOVA yielded no significance difference between the three sites. Similarly, the individual items within the scale were also tested separately and all four items were non-significant.

3.4.7. Overall Preference

The experience scale results and forced ranking choice were analysed to determine participants overall website preference.

3.4.7.1. Experience Scale

A one-way repeated measure ANOVA was carried out on the aggregated 3-item Overall Experience scale, and each of the individual items within the scale (stimulation, visit again and recommend), and no significant effect was found across the three sites. The aggregated means show that the Google (M=5.56) was favoured slightly higher than Louvre (M=5.34) and National (M=5.18), which is reflected in the mean scores for each individual question within the scale, as shown in Table 14.

<table>
<thead>
<tr>
<th></th>
<th>Visit Again</th>
<th>Recommend a Friend</th>
<th>Stimulating Experience</th>
<th>Aggregated Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Google</td>
<td>M= 5.50</td>
<td>M= 5.60</td>
<td>M= 5.57</td>
<td>M= 5.56</td>
</tr>
<tr>
<td>(2) Louvre</td>
<td>M= 5.35</td>
<td>M= 5.15</td>
<td>M= 5.53</td>
<td>M= 5.34</td>
</tr>
<tr>
<td>(3) National</td>
<td>M= 5.28</td>
<td>M= 5.13</td>
<td>M= 5.12</td>
<td>M= 5.18</td>
</tr>
</tbody>
</table>

Table 14. Mean rank order scores for Overall Experience Scale for the 3 sites

3.4.7.2. Forced Choice Preference

Participants were asked to rank each website in order of preference and the results were weighted to give an overall percentage rating per site, where 100% resulted if all participants rating the site first preference. When participants were given a forced choice the Louvre site won the highest rating at 41 %, followed by Google at 32 %, and then the National at 26 % (as
shown in Figure 26). This post-test question was based on participants’ memory of the three sites, which may be influenced by brand preference, as the Louvre was rated the highest for brand quality and was rated the most familiar (see Section 3.4.3.1), closely followed by the National Gallery. This also corresponds to the affect results that favoured the Louvre site (Figure 16) out of the three, although the overall experience scales results (see Section 3.4.7.1) favoured Google over Louvre (although this did not prove significant).

![Figure 26. Overall Percentage score from Forced Rank choice](image)

### 3.4.8. Inter-variable Relationships (Regressions)

Affect, Hedonics (HSQ), and Pragmatics (PQ) showed inter-site and inter-task differences, suggesting these were indicators of UX. However, the influence of participants’ emotional and affective responses and their judgement of website quality on their overall experience remains unclear. To explore further, a Multiple Regression was conducted to test the hypothesis:

*Website quality (HQ & PQ), and Affect could influence users Evaluations of Goodness, Beauty and Overall Experience.*

The *Multiple Regression* analysis was used to assess the influence of three independent (predictor) variables on three dependent variables for all three sites, across pre and post interaction. Predictors used were: Affect, Hedonics, and Pragmatics; on the Dependent Variables (1) Beauty, and (2) Goodness and (3) Overall Experience (for post interaction only).

Preliminary analysis was conducted to ensure there was no violation of the assumptions of normality, linearity and multicollinearity, and all values were within recommended tolerance (> 10: 0.136 to 0.668), and VIF (< 10: 1.49 to 7.38), (Tabachnick & Fidell, 2007).
Pre-interaction results showed HQ predicted Beauty for National ($R^2=0.791$, $\beta=0.55$, $p<0.01$), and less strongly for Google ($R^2=0.726$, $\beta=0.51$, $p<0.05$), and Louvre, ($R^2=0.721$, $\beta=0.38$, $p<0.05$). PQ showed only a weak effect for Louvre ($\beta=0.25$, $p<0.05$), as did Affect for National ($\beta=0.28$, $p<0.05$). Both HQ ($R^2=0.894$, $\beta=0.43$, $p<0.001$), and PQ ($\beta=0.50$, $p<0.001$) predicted Goodness for National; and the same pattern occurred for Louvre, HQ ($R^2=0.781$, $\beta=0.53$, $p<0.01$), and PQ ($\beta=0.31$, $p<0.05$), predicting Goodness. In contrast, only HQ weakly predicted Goodness for Google ($R^2=0.758$, $\beta=0.58$, $p<0.05$).

Post interaction, HQ predicted Beauty consistently for National ($R^2=0.674$, $\beta=0.71$, $p<0.01$), with PQ showing a minor influence ($\beta=0.27$, $p<0.05$); and for Google ($R^2=0.760$, $\beta=0.58$, $p<0.001$). In contrast, Affect was the only predictor for Beauty in Louvre ($R^2=0.694$, $\beta=0.74$, $p<0.01$). Goodness was predicted by HQ and PQ for Google ($R^2=0.820$, $\beta=0.45$, $p<0.001$; $\beta=0.33$, $p<0.01$); with weak influences from HQ in National ($R^2=0.679$, $\beta=0.43$, $p<0.05$), and PQ for Louvre ($R^2=0.636$, $\beta=0.27$, $p<0.05$). All three variables predicted Overall Experience for Google ($R^2=0.753$, HQ: $\beta=0.37$, $p<0.01$, PQ: $\beta=0.31$, $p<0.05$, Affect: $\beta=0.29$, $p<0.05$). Affect also predicted Overall Experience for National ($R^2=0.764$, $\beta=0.79$, $p<0.001$), and a weak effect for Louvre ($R^2=0.705$, $\beta=0.50$, $p<0.05$), which was shared with PQ ($\beta=0.29$, $p<0.05$). The results are summarized in Table 15.

<table>
<thead>
<tr>
<th>SITES</th>
<th>Dependent (DV)</th>
<th>BEAUTY</th>
<th>GOODNESS</th>
<th>OVERALL EXPERIENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predictor (IV)</td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
</tr>
<tr>
<td>National Gallery</td>
<td>Affect</td>
<td>*</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Hedonics (HQ)</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Pragmatics (PQ)</td>
<td>ns</td>
<td>*</td>
<td>***</td>
</tr>
<tr>
<td>Google Art Project</td>
<td>Affect</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Hedonics (HQ)</td>
<td>*</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Pragmatics (PQ)</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Louvre</td>
<td>Affect</td>
<td>*</td>
<td>ns</td>
<td>***</td>
</tr>
<tr>
<td></td>
<td>Hedonics (HQ)</td>
<td>*</td>
<td>ns</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Pragmatics (PQ)</td>
<td>*</td>
<td>ns</td>
<td>*</td>
</tr>
</tbody>
</table>

***=p <0.001, **=p <0.01, *=p <0.05

Table 15. Summary of the regression analysis results for the three sites (pre & post interaction).

The results suggest that HQ is a significant predictor for both Beauty and Goodness for all three sites, however PQ influenced Goodness post interaction for the enhanced interactive sites, and for National before interaction, even though the means and ANOVA showed that PQ for National was superior to Louvre and Google post interaction. Affect showed to be the main predictor for Overall Experience, and a weak predictor for Beauty for Google site only (see model summary in Figure 27).
The hypothesis model posits that users’ judgement of product quality, measured by HQ and PQ influences their overall judgement of global qualities, such as Goodness and Beauty. In contrast, the inference model (Hassenzahl & Monk, 2010; Hassenzahl, 2004) posit the opposite causal influence, i.e., that judgement of product qualities are inferred from general perceptions of Goodness and Beauty. Therefore, further regressions were run to investigate the inference model hypothesis, using Goodness and Beauty as predictors for HQ and PQ.

<table>
<thead>
<tr>
<th>Predictor (IV)</th>
<th>HEDONICS (HQ)</th>
<th>PRAGMATICS (PQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Gallery</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Goodness</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Beauty</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Google Art Project</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Goodness</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Beauty</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Louvre</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>Goodness</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>Beauty</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

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

Table 16. The general to specific (inference) model regression analysis results for 3 sites

The results summarized in Table 16, partially agree with the inference model pattern that Goodness and Beauty predict HQ, although Goodness predicting PQ was inconsistent between the sites. Beauty was the main predictor for National, and Goodness was the main predictor of PQ, although this relationship was less consistent across sites.

The models generated from the two regressions are summarized in Figure 27. Relationships between HQ, Beauty and Goodness shows both direction casual influences, for either the general to specific hypothesis or the general to specific inference model approach, illustrating the problem of assigning casual influences from correlation based analysis. The relationship between PQ and Goodness was less consistent in both models, reflecting pre/post interaction effects.
3.4.9. Criteria Importance

Participants were asked to rate the importance of how five different criteria (overall aesthetics, ease-of-use, content, engaging interaction and brand) may have influenced their judgement of quality for each website. Participants rated each criterion on a 1-7 scale for importance of judgement after initial exposure and task two.

Two-way repeated measures ANOVA was run on each criterion for task (2) and site (3). Four items showed a significant main effect for task; ease of use; \( F(1, 39) = 18.72, p<.001, \eta^2 = .32 \), content; \( F(1, 39) = 26.7, p<.001, \eta^2 = .41 \), engaging experience; \( F(1, 39) = 13.66, p<.001, \eta^2 = .26 \), and brand; \( F(1, 39) = 14.13, p<.001, \eta^2 = .27 \), indicating participant ratings for these four items increased after interaction. No significant differences were found between sites, and two items show weak significant interaction (site*task) effect; engaging interaction; \( F(2, 78) = 4.15, p<.05, \eta^2 = .056 \), and aesthetics importance; \( F(2, 78) = 3.49, p<.05, \eta^2 = .027 \). The means (see Figure 28) for engaging interaction, show the National being ranked first on initial exposure, but after interaction was ranked last, where Google was favoured over Louvre. Google clearly was rated the highest for aesthetic importance for both initial exposure and after interaction, whereas National ratings dropped post interaction. However Louvre showed the greatest increase as a result of interaction, where ratings post interaction were in line with Google. This clearly indicates that participants’ judgement of aesthetic importance increased after interaction for the sites that offered more interactive features.

![Figure 28. Mean Patterns for Engaging Interaction (left) and Aesthetic Importance (right).](image)

The mean criterion importance ratings and rank order for each website are shown in Table 17, which shows the National being ranked first for all items on initial exposure, apart from aesthetic importance, however this changed after interaction, where the National only remained first for ease of use, and was ranked last on all other items. This indicates that ease of use was an important criteria for participants when they judged the website quality,
although it showed no significant inter-site differences, which supports the non-significant finding for the usability scale (see Section 3.4.6). Little difference was found between Louvre and Google post interaction, with rank order being equally shared.

<table>
<thead>
<tr>
<th></th>
<th>Initial Exposure</th>
<th>Task Two</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>National</td>
<td>Google</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>M= 5.72</td>
<td>5.85</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>M= 5.80</td>
<td>5.15</td>
</tr>
<tr>
<td>Content</td>
<td>M= 5.17</td>
<td>4.83</td>
</tr>
<tr>
<td>Engaging</td>
<td>M= 5.15</td>
<td>5.05</td>
</tr>
<tr>
<td>Brand</td>
<td>M= 4.13</td>
<td>4.13</td>
</tr>
</tbody>
</table>

Table 17. Mean scores and ranking order of websites for Criteria Importance

### 3.4.10. Exploration Time

For each task participants had up to 3 minutes free time to interact with each of the websites. Time logs were recorded and a two-way (2 task * 3 site) repeated measures ANOVA was performed to see if there was an effect of time for each website. A significant main effect was found on task $F(1, 39) = 7.72, p < .001$, $\eta^2 = .17$, (Greenhouse-Geisser correction), but not for site or the interaction (site * task). Time dropped for both the National (T1: M= 2.43 to T2: M=2.16) and the Louvre site (T1: 2.54 to T2: M=2.35) from task one to task two, but remained the same for the Google (T1: M=2.30 to T2: M=2.31), as shown in Figure 29. This indicates that Google maintained the participants’ interest longer than the other two sites, although it is unclear if this is an interaction effect, or a result of the Google site being a portal where users had the option to explore other galleries. However, overall participants spent more time exploring Louvre (M=2.5), compared to Google and National (M=2.3).

![Figure 29. Mean Patterns of Time Exploration](image-url)
3.4.11. Summary of Quantitative Results

To summarize the quantitative results, repeated exposure to all three sites produced a clear increase in affect and website quality ratings after the initial exposure (homepage), with the increase being more apparent for the interactive sites (Google and Louvre). This confirms the original hypotheses that websites with enhanced interactive design features improve user engagement.

<table>
<thead>
<tr>
<th>Quantitative Measures</th>
<th>Order of Mean Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>L G N</td>
</tr>
<tr>
<td>HQS (Creative/Innovative)</td>
<td>G L N</td>
</tr>
<tr>
<td>PQ (Clear/Predictable)</td>
<td>N G L</td>
</tr>
<tr>
<td>SDQ (Spacious)</td>
<td>G N L</td>
</tr>
<tr>
<td>Goodness</td>
<td>L G N</td>
</tr>
<tr>
<td>Beauty</td>
<td>G L N</td>
</tr>
<tr>
<td>Overall Experience</td>
<td>G L N</td>
</tr>
<tr>
<td>Forced Ranking</td>
<td>L G N</td>
</tr>
<tr>
<td>Exploration Time</td>
<td>L G N</td>
</tr>
</tbody>
</table>

*Table 18. Summary of rank order across quantitative measures*

Little difference was found between Google and Louvre, with first rank orders being split between the two (see Table 18). Although Google gaining a slight advantage in overall experience and the website quality (HQS, SDQ) ratings, which can be attributed to its simple homepage layout and the 3D navigation metaphor enabling participants to navigate in a third dimension; Louvre scored higher affect ratings, and participants spent more time exploring the site, and it was their overall preferred (forced) choice. Mean ratings showed no difference between Google and Louvre for Beauty or Goodness ratings. The National trailed behind in 3rd position on all ratings, although it scored the highest for PQ, due to the familiar grid style layout, making it appear clearer and more predictable to use.

Participants’ judgement of quality items was enhanced through interaction, where ratings increased for engaging interaction and aesthetic importance for those sites that contained interactive features. Participant background (artistic disposition, awareness, gender) had no impact on the results. No difference was found between sites for brand awareness, however Louvre scored slightly higher for brand quality, possibly reflected in participants overall forced choice preference.

Although both Google and Louvre was rated higher for *involvement*, no other immersion or presence results were found. No time effect was captured during interaction (from task 1 to task 2) on all measures, however participants spent more time exploring the more interactive
sites (Google and Louvre), indicating that overall participants preferred the sites that provided more enhanced interactive features.

The reasons for these judgements are explored in the following section that reports the qualitative data analysis.

### 3.5. Qualitative Analysis and Results

The qualitative data consists of the participant interviews and observations during the experiment and will be used to further understand reasons behind participants’ judgement of the three websites. A combination of thematic identification and enumeration was used to analyse the qualitative data.

#### 3.5.1. Qualitative Analysis

The interview data was analysed using a three-step process, as shown in Figure 30. First was a directed content analysis approach (top-down), where predetermined codes were generated from the questionnaire constructs (website quality, usability, interactive features etc.), which provides an initial guide to the analysis (Hsieh & Shannon, 2005). Then a summative content method (bottom-up) was used where occurrence of key words and phrases were identified (affect, presence, flow etc.) using NVivo, and open coding was applied. This process resulted in multi-allocation of comments across themes, for example a usability comment often elicited frustration, so was coded twice (usability/frustration). Total frequency counts for each theme was used to direct the final analysis.

![Figure 30. The qualitative data analysis method](image)

All comments were assigned a valency, either positive when praise or good qualities were described, or negative when comments involved criticism, or neutral as a default. This
deductive enumeration of comments was used as a basis for theme comparison across the three websites themes where the summative findings can inform the quantitative results (Goetz & LeCompte, 1981).

### 3.5.2. Memory Recall

Based on an adapted critical incident technique (Cassell & Symon, 2004), at the start of the interview participants were asked to list the most important aspects of the websites they had visited. The first two recalled ‘aspects’ were recorded as either a positive or negative comment. From the 80 aspects (N=40x2) a total of 11 different categories emerged.

Google accounted for 42% of the total recalled aspects, closely followed by Louvre at 40%, with both being cited more frequency than National, at 18%. Google was remembered mainly for its 3D Virtual Environment (VE) (accounting for 36% total aspects), while the Louvre features were spread between the cartoon guide (10%), animation (14%) and interactive room (10%). The National was primarily remembered for its ease-of-use (10%) and content (8%). It is worth noting the balance of positive and negative results for each aspect (prior to calculating the NV value), shown in Figure 31. Interestingly the same features that elicited high saliency positive scores (3D VE, Cartoon, Animation and Interactive Room) also received the highest negative scores, which were associated with Louvre and Google sites only. The Louvre received the highest number of negative comments collectively for the cartoon, animation and interactive room, compared to the single item 3D view attributed to the Google, while the National received no negative comments.

![Figure 31. Frequency of the first two most recalled items (showing positive & negative items)](image-url)
Net valencies (NV) were calculated (by subtracting the negative from the positive frequencies, and discarding neutral items), to assess the participants’ attitude towards the recalled aspects. Overall, the interactive features were more frequently recalled with positive memories for both Google (NV+18) and Louvre (+19), while National had positive recall for content and ease of use (both +4) see Figure 32. All other categories had low volume and near neutral (+1) recall. Memory of Google and Louvre was equally positive (+21/+20), with NG trailing in third place (+9).

![Figure 32. The NV scores for the two most remembered aspects across the three sites](image)

### 3.5.3. Interview Comments

Three coding frameworks were identified, Website Interactive Features, Engagement (Affect and Immersion), and General Impressions.

#### 3.5.3.1. Website Features

A total of 366 interview comments were coded first into 17 relevant website feature sub-themes, which were then grouped into eight key main themes, as shown in Table 19. The total frequency of comments was recorded per site according to the website feature themes and the overall percentages and NV scores are now reported.

The Louvre site received the highest frequency (44% of total) of feature comments and the highest NV score (+46), while Google had almost as many comments (36%), which tended to be more negative (NV +9). The National received fewer comments (20%) although most were positive (NV +37).
Before discussing the main results it is worth reviewing the positive and negative comments recorded for the sub-themes (website feature themes) per site, as shown in Figure 33. The graph clearly illustrates that the highest positive comments were attributed to the interactive features, with Louvre receiving the highest frequency (80% of both positive and negative comments) compared to Google (20%). However, Google received higher negative comments for navigation, usability and content, which will now be discussed in relation to the NV scores.

The NV scores are shown in Table 20, which are split across the eight main website themes. The Louvre received the highest frequency (56% of Louvre comments) for the interactive features (avatar, interactive objects, animation and audio guide) yielding the best overall positive NV (+32) as illustrated by the following excerpts:

---

Table 19. The 8 key main themes and 17 sub-themes that formed the coding framework

<table>
<thead>
<tr>
<th>8 KEY MAIN THEMES</th>
<th>17 WEBSITE FEATURE SUB-THEMES</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Interactive Features</td>
<td>3D Virtual Environment</td>
</tr>
<tr>
<td></td>
<td>Animation</td>
</tr>
<tr>
<td>• Usability</td>
<td>Cartoon Character</td>
</tr>
<tr>
<td></td>
<td>Utility</td>
</tr>
<tr>
<td>• Design Qualities</td>
<td>Interactive Objects</td>
</tr>
<tr>
<td></td>
<td>Zoom</td>
</tr>
<tr>
<td>• Navigation</td>
<td>Design</td>
</tr>
<tr>
<td></td>
<td>Map</td>
</tr>
<tr>
<td>• Content</td>
<td>Colour</td>
</tr>
<tr>
<td>• Ease of Use</td>
<td>Information</td>
</tr>
<tr>
<td>• Presentation Structure</td>
<td>Portal</td>
</tr>
<tr>
<td></td>
<td>User Friendly</td>
</tr>
</tbody>
</table>

---

Figure 33. The Website Feature frequency of Positive & Negative comments across the 3 sites

The NV scores are shown in Table 20, which are split across the eight main website themes. The Louvre received the highest frequency (56% of Louvre comments) for the interactive features (avatar, interactive objects, animation and audio guide) yielding the best overall positive NV (+32) as illustrated by the following excerpts:
[P25, Louvre Avatar]: “I liked the little character... he made it a bit more interesting... made the story alive behind the works of art... it made me smile”, [P23, Active Objects]: “you could click on things that interested you... that really appealed to me... that is why it engaged me more”, [P22, Louvre Animation]: “I remember more about the animations, which were colourful and bright”, and [P21, Audio Guide]: “I really enjoyed the part where there was an audio description... its quite interesting”.

<table>
<thead>
<tr>
<th>Main Website Feature Themes</th>
<th>National NV</th>
<th>Google NV</th>
<th>Louvre NV</th>
<th>Total NV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive Features</td>
<td>-</td>
<td>+17</td>
<td>+32</td>
<td>+49</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>+10</td>
<td>+10</td>
<td>+9</td>
<td>+29</td>
</tr>
<tr>
<td>Content</td>
<td>+10</td>
<td>-5</td>
<td>+12</td>
<td>+17</td>
</tr>
<tr>
<td>Usability</td>
<td>-2</td>
<td>-17</td>
<td>-4</td>
<td>-23</td>
</tr>
<tr>
<td>Presentation Structure</td>
<td>+10</td>
<td>+4</td>
<td>-4</td>
<td>+10</td>
</tr>
<tr>
<td>Design Qualities</td>
<td>+3</td>
<td>+7</td>
<td>+7</td>
<td>+17</td>
</tr>
<tr>
<td>Navigation</td>
<td>+6</td>
<td>-7</td>
<td>-4</td>
<td>-5</td>
</tr>
<tr>
<td>Total NV</td>
<td>+37</td>
<td>+9</td>
<td>+46</td>
<td>(366)</td>
</tr>
</tbody>
</table>

*Table 20. NV’s for the main Website Feature themes by site*

Google comments were more evenly spread among navigation, design quality, content and interactive features (ranging from 17 to 21% of the Google total), of which those related to its main interactive feature (3D VE), which were mainly positive (+17), as reported:

[P37, 3D Virtual Environment]: “Being able to walk around the gallery was interesting... it was visually stimulating”, [P30]: “the way you could move around so easily and everything felt very instinctive”, [P20]: “it was kind of a virtual tour... the way you navigate to different rooms”.

Although Google received positive ease-of-use comments (+10) it had the worst usability NV (-17 for 17 comments) and navigation (-7) caused by poor image quality and difficulty moving from the 3D to 2D space, as explained by these extracts:

[P21, Google Usability]: "The poor quality of images, especially zooming... after two zooms you couldn’t see anything”, [P15]: “it wasn’t so intuitive that you could click on those little white plusses to view a picture”, [P25, Navigation]: “I didn’t like it... it was really hard to navigate and I just kept getting really confused as to where I was, what I was doing”.

Google also received a high number of negative comments for content, giving an overall negative NV score (-5) as a few participants explained:

[P5, Google Content]: “the Google one didn’t necessarily have a lot of information, like when I clicked on artist information, it didn’t really have that much there” and [P12]: “you could only
get information on some of the paintings and not all of them”, and there was “a lack of information and description of the paintings were very short”.

However the Louvre and National received a positive NV (+12/+10) for content, with National gaining the advantage for its presentation structure (+10), as reported:

[P30, Louvre Content]: “I felt there was a lot of rich content on the Louvre site”, and [P30, National Content]: “there’s a greater depth of information on the National”. [P31, National Structure]: “its very well structured and laid out”, and [P15], “its quite well balanced”.

Little difference was found between the sites for design quality, with Louvre and Google gaining a slightly higher NV (+7), than National (+3), with comments often relating to their novelty of the sites:

[P30, Google Design]: “it just felt really modern and sort of cutting edge”, and “[P15] “its quite a novel look and didn’t have the stereotypical outline”. [P20, Louvre Design]: “It was engaging, the colour, design the shape and the layout was pretty much kind of fascinating”, and [P37]: “it was quite creative and stimulating”

3.5.3.2. Engagement
A total of 181 engagement related phrases were collated into 13 themes through conducting a word search using NVivo. Themes were split into two groups, (1) immersion/presence, and (2) positive and negative affect.

Immersion/Presence
A total of 80 immersive or presence related phrases or words captured during the interviews were coded into 6 themes, (see Table 21).

<table>
<thead>
<tr>
<th>Immersion/Presence Themes</th>
<th>National</th>
<th>Google</th>
<th>Louvre</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Presence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being There</td>
<td></td>
<td>-</td>
<td>+10</td>
<td>12</td>
</tr>
<tr>
<td>Interactive Exp.</td>
<td>Dynamic</td>
<td>-7</td>
<td>-4</td>
<td>-3</td>
</tr>
<tr>
<td>Pace</td>
<td>Time, Slow, Quick</td>
<td>+1</td>
<td>+2</td>
<td>-6</td>
</tr>
<tr>
<td>Control</td>
<td>Manage, Master, Grasp</td>
<td>-</td>
<td>+6</td>
<td>- 2</td>
</tr>
<tr>
<td>Natural</td>
<td>Instinctive</td>
<td>-</td>
<td>+3</td>
<td>-1</td>
</tr>
<tr>
<td><strong>TOTAL NV</strong></td>
<td>-6</td>
<td>+17</td>
<td>+1</td>
<td>(80)</td>
</tr>
</tbody>
</table>

Table 21. NV's for the Immersion/Presence themes by site

Google received the highest frequency (56%) of immersive/presence comments; followed by Louvre (31%), with National coming in third (13%). Presence and Interactive Experience themes were the most frequent (35%, 34% total), with Google accounted for most presence
comments (93%), which generated an overall positive NV (+10), with comments relating to participants’ sense of being in the 3D space:

[P10, Google Presence]: “It gives you a sense of being there, as if you are walking around”,
[P18]: “it gives you a feeling as if you are in the museum itself, and you are involved and engaged in it”, and [P27] “it’s quite a real experience”.

Participants also referred to feeling natural and in control while navigating the 3D space:

[P35, Google Natural]: “It felt very natural as if I was really there”, [P30] “everything felt very instinctive”. [P2, Google Control]: “you feel really in control of the situation, it makes you feel you can go in every room and see every corner”.

Figure 34. Immersion/Presence frequency of Positive & Negative comments across the 3 sites

However Google also received negative Presence comments (see Figure 34), as “it gets a bit boring after a while, I have much more enjoyment actually being there” [P34]; as well as receiving a negative NV (-4) for interactive experience, as it was considered “not as interactive compared to Louvre” [P18], since it only offered one feature (3D VE).

As expected the National Gallery received a negative NV (-7) for the lack of interactive features, [P40]: “there is nothing, no multimedia interaction... after a while you can get a bit bored”, and [P25]: “its pretty average, nothing spectacular”. In contrast, Louvre received a high positive NV (+8) for interactive experience, despite scoring a negative NV for pace (-6), as reported here:

[P39, Louvre Interaction]: “there were different methods of interacting with it, there was the films, the story... the character”. [P24, Louvre Pace]: “I found the pace of speaking a little bit slow”, and [P.26]: “I think the slowness of it... at having to wait... felt like I wasn’t in control”.

97
Affect

The affect group consists of 101 phrases generated from key word searches, which were divided into 8 themes (4 positive and 4 negative) as shown in Table 22.

<table>
<thead>
<tr>
<th>Positive &amp; Negative Affect</th>
<th>Key Words</th>
<th>National</th>
<th>Google</th>
<th>Louvre</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engaging +</td>
<td>Captivate, Absorb, Involv, Grab</td>
<td>+2</td>
<td>+4</td>
<td>+17</td>
<td>23</td>
</tr>
<tr>
<td>Fun +</td>
<td>Funny, Enjoyable, Comical</td>
<td>-</td>
<td>+1</td>
<td>+16</td>
<td>17</td>
</tr>
<tr>
<td>Attractive +</td>
<td>Appeal, Beautiful, Draw, Pulls</td>
<td>-</td>
<td>+3</td>
<td>+2</td>
<td>5</td>
</tr>
<tr>
<td>Exciting +</td>
<td>Stimulating, Intriguing</td>
<td>-</td>
<td>+3</td>
<td>+3</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td><strong>Total Positive NV</strong></td>
<td><strong>+2</strong></td>
<td><strong>+11</strong></td>
<td><strong>+38</strong></td>
<td><strong>(51)</strong></td>
</tr>
<tr>
<td>Confusing -</td>
<td>Distract, Lost, Complicate, Disjoint</td>
<td>-2</td>
<td>-3</td>
<td>-8</td>
<td>15</td>
</tr>
<tr>
<td>Annoying -</td>
<td>Irritating</td>
<td>-1</td>
<td>-11</td>
<td>-1</td>
<td>12</td>
</tr>
<tr>
<td>Boring -</td>
<td>Dull, Tedious, mundane</td>
<td>-8</td>
<td>-7</td>
<td>-1</td>
<td>16</td>
</tr>
<tr>
<td>Frustrating -</td>
<td>Disappointing</td>
<td>-4</td>
<td>-4</td>
<td>-4</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td><strong>Total Negative NV</strong></td>
<td><strong>-10</strong></td>
<td><strong>-16</strong></td>
<td><strong>-24</strong></td>
<td><strong>(50)</strong></td>
</tr>
<tr>
<td>NET TOTAL NV AFFECT</td>
<td>-8</td>
<td>-5</td>
<td>+14</td>
<td>(101)</td>
<td></td>
</tr>
</tbody>
</table>

Table 22. NV’s for the Positive and Negative Affect comments

The frequency of positive (51%) and negative affect (49%) comments was almost the same. Louvre received more overall affect comments (61%) than Google (27%) and National (12%). The most frequent positive affect categories were engaging (45 % of total positive affect +ve) and fun (33 %), and these were concentrated on the Louvre site (33%, 31% of total +ve affect), (see Figure 35), with comments relating to its interactive features:

[P33, Louvre Engaging]: “It draws you more into art, to make it more fun and engaging”, and [P15]: “the comic guide was quite funny and engaging”. [P14, Louvre Fun]: “the website was very clever I thought, and quite funny”, and [P29]: “where you could click on the different artifacts and look at them, that was a nice, fun, engaging part of the website”.

Louvre also received the highest negative affect score (-24, 48% total negative affect -ve) for being the most annoying (-11) and confusing (-8), which was especially due to the avatar guide and site layout, as these excerpts explain:

[P11, Louvre Annoying]: “it was interesting at first, but it quickly got annoying because it felt a bit childish”, [P17]: “I think that got a bit irritating after a bit... it’s a bit gimmicky”, and [P18]: “its quite distracting for me because I’m trying to read information about the picture and... all the time this cartoon is moving about... it’s a bit annoying”. [P5, Louvre Confusing]: “It was kind of complicated...the Louvre one... it seemed a little bit all over the place and confusing”.

98
Similarly, Google received positive NV (+4/3) for engaging, exciting and attractive, but also a high frequency of negative comments (see Figure 35), for boring and frustrating (-7/6), with comments relating to users’ interactive experiences:

[P 33, Google Boring]: “It didn’t engage me, it just wasn’t engaging enough, it wasn’t compelling”, and [P34]: “once you have seen a couple of things ... it gets a bit boring after a while, pretty quickly”.

Although the National received the fewest comments, it was referred to the most for being boring, due to its lack of interactive features, as reported:

[P9, National Boring]: “it kind of gave me the perception of what a gallery would be like, a bit boring... it was soul-less”, and [P33]: “it wasn’t dynamic enough, it wasn’t compelling, there wasn’t a personality to it... its quite drab I thought”.

### 3.5.3.3. General Impressions

A total of 56 phrases and key words were coded into 5 themes that related to general impressions of which creative (41%) and narrative (34%) were the most frequent.

<table>
<thead>
<tr>
<th>General Impression</th>
<th>Key Words</th>
<th>National</th>
<th>Google</th>
<th>Louvre</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NV</td>
<td>NV</td>
<td>NV</td>
<td>NV</td>
</tr>
<tr>
<td>Creative</td>
<td>Innovative, Initiative, Clever</td>
<td>-2</td>
<td>+7</td>
<td>+8</td>
<td>13</td>
</tr>
<tr>
<td>Narrative</td>
<td>Story, History, Picture</td>
<td>-</td>
<td>+1</td>
<td>+10</td>
<td>11</td>
</tr>
<tr>
<td>Childish</td>
<td>Silly, Foolish, Frivolous</td>
<td>-</td>
<td>-</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>Professional</td>
<td>Amateurish</td>
<td>-1</td>
<td>-</td>
<td>-3</td>
<td>+4</td>
</tr>
<tr>
<td>Traditional</td>
<td>Conservative, Conventional</td>
<td>+2</td>
<td>-</td>
<td>-1</td>
<td>+3</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>TOTAL NV</strong></td>
<td><strong>-10</strong></td>
<td><strong>+18</strong></td>
<td><strong>+14</strong></td>
</tr>
</tbody>
</table>

*Table 23. NV’s for the General Impression themes by site*
Louvre received the highest number of general impression comments (45%), followed by Google (33%) and then National (22%). Louvre accounted for most of the narrative and creativity comments (90%, 61% of category totals), with positive NVs (+8, +10, see Table 23) where comments related to the audio guide feature:

[P33, Louvre Narrative]: “Its really effective, because it told you the story and it allowed you to picture more what was happening at the time” [P39], and “I think it’s the most innovative, fun to use”. [P12, Louvre Creative]: “I thought the Louvre looked the most premium and stylish with their design”, and [P37]: “it was well designed and quite creative”.

Google comments were slightly less positive (+7, +1), and considered “quite clever” and “innovative” due to “being able to walk around” the 3D VE, however several participants were familiar with the technology, [P29]: “I am used to that functionality… I’m really used to that, so you kind of felt at home with Google”.

The National received the lowest frequency of comments (22%), and was considered not “very creative or interesting” and “pretty conventional and traditional”, also participants experienced difficulty in recalling the website, [P30]: “I am struggling to remember the middle one” [NG].

3.5.4. Qualitative Results Summary

In the quantitative analysis, the websites that contained the interactive features (Google and Louvre) generated more positive affect and website quality ratings than the National site that offered limited interactivity. No significant difference was found between the two interactive sites, although Google was marginally favored for hedonic experience, Louvre generated more affective responses and participants explored this site longer, while the National was preferred for its pragmatic qualities.

However, the picture changes in the qualitative analysis (see summary Table 24). Although Google and Louvre were close in immediate memory recall both for volume and valency, and were remembered more for their interactive features, with National trailing in 3rd place. During the interviews participants had more opportunity to reflect upon their judgement through their verbalization, and the Louvre site was more favored in comment volume and valency for three of the themes, website features, positive affect and general impressions, with Google only remained ahead for immersion and presence.
The reasons for these judgements appear to be at the design feature level, where the 3D interactive metaphor for Google and Louvre’s combination of interactive media (avatar, interactive media, animation etc.) generated most of the comments. Both Google and Louvre received the most negative comments. Google scored poorly on usability and navigation (consistent with the low PQ score), along with negative affect comments relating to some participants’ frustration/boredom navigating the 3D environment. Louvre also received high negative affect comments and pace, which can also be attributed to poor usability (also inline with the low PQ score), arising from the audio guide, the avatar (for some users), and pace of interaction. Although National received positive comments for ease-of-use, content, structure and navigation (as reflected by its high PQ rating), it received negative affect comments for being boring and lacking an interactive experience.

Interestingly when users’ asked to state a preference rank order (forced choice) at the end of the interview, Louvre came out ahead with 41% first choices, followed by Google (32%) and NG languishing in third position (26%). A conjecture is that immediate memory recall was dominated by affective impressions of the users’ interactive experience that typically come quickly and automatically (Russell & Carroll, 1999). But when participants are given time to reflect on their experience (Norman, 2002), as during the verbal interviews, they became more critical of Google and better disposed towards Louvre.

The qualitative analysis not only revealed some of the reasons for users’ judgement at the feature level but also demonstrated that judgement is inconsistent across the population, as reflected in the mix of positive and negative NV scores. For example, some participants
thought the Louvre’s avatar guide was fun and engaging, while others found it childish and annoying. Similarly, for some the audio guide and animations were informative and enjoyed for its narrative; however, others found them distracting and preferred just to read the content. Although NG was third in most measures some participants preferred its ease-of-use and simple structure. The following section will investigate individual differences to assess possible variations in the way people make quality judgments.

### 3.6. Individual User Experience

As quite a few of the interactive feature-related comments were polarized between the participants, even on the same site, (e.g., participants were split between the Louvre avatar and the Google 3D environment), further exploration was conducted to identify any subgroup preferences for either of the sites.

Adopting a user-type profiling approach (Brandtzæg, 2010) similar to Bartle, (1996) who identified player types within game design, the data was explored using *Cluster Analysis (CA)* to identify possible individual differences. *Cluster Analysis* is an exploration data analysis tool used to categorize data into meaningful taxonomies, groups or clusters. It is used in market segmentation where it can be useful to identify customer ‘types’ in order to market products aimed specifically at customer needs (Burns & Burns, 2008). *Cluster Analysis* is used within this research to identify homogenous groups of different types of users, but differs from *Factor Analysis (FA)*, which reduces the number of variables into factors; *Cluster Analysis* reduces the number of cases by grouping them into smaller cluster sets using a measure of similarity or distance.

To investigate whether there were any user groups manifest in the user judgement and attitudes, a *Hierarchical Cluster* analysis was conducted using SPSS on the quantitative data sets that had discriminated between the sites (*Hedonics* and *Usability*), along with the *Overall Experience* data and the aggregated NV score taken from the qualitative data. A two-step CA was conducted; first a *Hierarchical Cluster* analysis using Ward’s method (based on squared Euclidean distances and similarity measure) was used to determine the optimal number of clusters, which was then subsequently used as an input for conducting a *K-means Clustering*, which uses a partitioning (within cluster measure) method of analysis. Using two methods provides a validity check that increases the reliability of the final cluster groups. Although there is no adequate method for choosing the number of clusters (Hair, Tatham, Anderson, & Black, 2006), inspection of the dendrograms from the *Hierarchical Cluster* analysis showed three upper-level clusters for all three websites, (K-means analysis) suggesting a three-cluster solution, and the results for the three site cluster groups shown in Table 25.
For the National site a two cluster-groups were apparent, after one individual outlier was eliminated. Only four participants ranked National first on the aggregated preference measures, so both of these clusters were composed of users’ ratings for their second or third choice site. The most positive groups (1) for each site are likely to be enthusiasts for that site, with the second group (2) being early adopters (Norman, 1998), which reflects the majority, while the last group with the lowest ratings (3) appear to be non-adopters for that site. The larger National group of adopters (N=29) probably reflects second choices by the Louvre/Google enthusiasts.

<table>
<thead>
<tr>
<th>Cluster Groups</th>
<th>National</th>
<th>Google</th>
<th>Louvre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hedonics (HQ)</td>
<td>5.64</td>
<td>6.34</td>
<td>6.26</td>
</tr>
<tr>
<td></td>
<td>4.21</td>
<td>5.42</td>
<td>5.52</td>
</tr>
<tr>
<td></td>
<td>3.34</td>
<td>3.34</td>
<td>4.09</td>
</tr>
<tr>
<td>Usability (PQ)</td>
<td>6.13</td>
<td>6.06</td>
<td>6.41</td>
</tr>
<tr>
<td></td>
<td>3.88</td>
<td>5.36</td>
<td>5.54</td>
</tr>
<tr>
<td></td>
<td>1.94</td>
<td>1.94</td>
<td>3.91</td>
</tr>
<tr>
<td>Overall Experience</td>
<td>5.86</td>
<td>6.19</td>
<td>6.33</td>
</tr>
<tr>
<td></td>
<td>3.03</td>
<td>5.68</td>
<td>5.61</td>
</tr>
<tr>
<td></td>
<td>2.25</td>
<td>3.37</td>
<td>3.37</td>
</tr>
<tr>
<td>NV</td>
<td>11</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>N members</td>
<td>29</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>9</td>
</tr>
</tbody>
</table>

(1)=Enthusiasts, (2)=Adopters, (3)=Non-Adopters

**Table 25. Cluster group means for three sites: National, Google and Louvre**

The Google enthusiasts (cluster group 1) were larger (N=17) than Louvre (N=13), although the mean scores were slightly lower, while the non-adopter (cluster group 3) shared similar means between Google and Louvre, although Louvre had a higher group membership (N=9/4). From the 17 members in Google’s’ enthusiasts cluster group, 10 individuals rated Google first on preference, with the remaining 7 shared equal preference with Louvre (4) or National (3).

From the cluster means the Louvre enthusiasts (cluster group 1), 14 individuals rated Louvre first on preference, however a few of the adopters (cluster group 2) also favoured Louvre equally with Google (4 ties) and National (2 ties). Although the enthusiasts rated their preferred site better than the adopters, the inter-site differences were not significant for any of the groups. It appears that participants split into enthusiasts for a particular site, but their second choice also received reasonable ratings (adopters), so most would probably tolerate their second preference. The third choice for all sites was rated much worse, which may possibly result in non-adoptions of the site. Louvre attracted more non-adopters (N=9) compared to Google (N=4), even though it was favoured the most overall, which possible reflects the polarised preferences towards the avatar as highlighted in the qualitative findings.
3.7. Chapter Summary

This study demonstrates the importance of interactivity in positively influencing user experience, and shows that user preference between interactive websites can be explained through user types. However, a caveat is that the product domain (leisure/entertainment), and task-type (action-mode) may have a contributory factor affecting the generalizability of this finding (Hartmann et al., 2008; Hassenzahl & Ullrich, 2007).

The original hypothesis was proven, where interaction with all three sites produced an increase in affect and hedonics after initial exposure, where the increase was more evident for the interactive sites (Google and Louvre) due to their interactive features employed. Affect and hedonics increased for all three sites demonstrating a powerful effect of interaction on UX. Initial interaction showed the most dramatic effect, which dropped slightly after the second interactive task, showing some signs of waning. However, little difference was found between the interactive sites (Google and Louvre), although Google gained a slight advantage over Louvre, while the National site scored well for ease of use (PQ), it remained in third position due to its limited interactive features.

The quantitative results were supported by the qualitative findings, where Google and Louvre received higher positive comments on interactive website features, affect and general impressions compared to the National site. These findings are consistent with previous studies that have suggested the interactivity is an important factor in attracting and engaging users (Bargas-avila & Hornbæk, 2011; Venkatesh, Morris, Davis, & Davis, 2003), and it can be linked to specific interactive design features (Hartmann et al., 2008).

<table>
<thead>
<tr>
<th>Measures</th>
<th>Google</th>
<th>Louvre</th>
<th>National</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative Measures</td>
<td>Rank Order of Means</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affect (GA-L not sig)</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Hedonic Quality (HQS)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Pragmatic Quality (PQ)</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Overall Experience (GA-L not sig)</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Qualitative Results</th>
<th>Rank by (Volume) and NV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive Features</td>
<td>(2) 3</td>
<td>(1) 1</td>
</tr>
<tr>
<td>+Affect</td>
<td>(2) 2</td>
<td>(1) 1</td>
</tr>
<tr>
<td>-Affect</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Immersion/Presence</td>
<td>(1) 1</td>
<td>(2) 2</td>
</tr>
<tr>
<td>Impressions</td>
<td>(2) 1</td>
<td>(1) 2</td>
</tr>
<tr>
<td>Exploration Time</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Memory Recall</td>
<td>(1) 1</td>
<td>(2) 2</td>
</tr>
<tr>
<td>Rank Preferences</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

*Table 26. Summary of mean ranking & frequency of quantitative and qualitative results*
Quantitative ratings were captured immediately after interaction via the questionnaires, with Google scoring the highest for hedonics and (although not significantly), showed a slight advantage in overall experience ratings (see Table 26). A shift in users’ preferences was found, when users were encouraged to articulate their opinions in the post-test interviews, where a preference for Louvre emerged supported by more positive comments on interactive features and general impressions. One explanation for this shift is that users make different types of judgement depending on the measures employed. For example quantitative questionnaire responses maybe more of a affect-based style of judgement, generated by a holistic “gestalt of an objects” initial impression (Hassenzahl, 2008a, P6; Lindgaard et al., 2011); whereas interviews may stimulate more cognitive reflection, and hence produce more detailed, refined judgement. Clearly this view has important implications for user research and requires further, more comprehensive investigation.

This study showed clear differences in UX ratings and user preference between websites with differing degrees of interactivity; however although these three sites were from the same domain with similar content, their content was not controlled. The motivation for the second study (reported in Chapter 4) is to conduct a controlled experiment to test specific interactive features of two websites with identical content. The effect of longer-term interactivity will be investigated in Chapter 5 & 6.
Chapter 4  Study Two: Interactivity - Avatars & Media

This chapter presents the second study within this thesis. It aims to verify the same hypothesis that directed Study One; that more interactive websites will generate higher positive user experiences. However, the study design differs from the previous study as it uses a controlled experiment that manipulates different interactive features in order to show clear differences in user preference and user experience ratings as a result of interactivity.
4.1. Introduction

This study further investigates the principle of interaction and how it may influence user experience by focusing on specific interactive design features. It aims to extend the findings from Study One reported in the last chapter, which demonstrated the importance of interactivity in positively influencing user experience. Both the quantitative and qualitative results supported the hypothesis that websites with interactive features prompted positive quality ratings (affect, hedonics, and overall impression) and generated a higher volume of positive comments. Clear individual differences in website preference became evident in a polarisation of opinion in relation to varying interactive features. Although Study One highlighted users' divergent attitudes towards avatars, the three websites were selected from the same domain (art gallery) with similar characteristics, however the content and design was not controlled. Subsequently, a second more controlled experiment was developed in order to test specific interactive features (avatars and the use of video media), through testing two identical specially designed websites.

This second study was conducted to verify the importance of interactivity in determining users' preferences and perceptions of experience, and therefore adopts the same hypothesis used in Study One:

H1: *Websites with Interactive Features will provide a more Positive User Experience than Standard Websites containing no Interactive Features*

Based on the results from the Study One, two further hypotheses are also proposed:

H2: *Participants’ User Experience ratings of a website will increase after Interaction*

H3: *Participants’ User Experience ratings of a website that contains Interactive Features will be more Positive than a Standard Website (without Interactive Features)*

A controlled experiment design was chosen in order to verify the hypothesis (H1), by isolating the effect of one variable (interaction), while controlling all the other confounding variables (e.g., aesthetics, content, context). The effect of interactivity is evaluated before (initial exposure to homepage), and after interaction (activity tasks), while user engagement (video) and presence (avatar) were tested using a ‘with and without avatar and video’ design. The influence of the initial affective responses, website qualities (hedonics), usability (pragmatics), beauty and goodness and individual user preferences and attitudes to the websites will also be investigated in relation to user experience.
4.2. Materials

Two bespoke websites were developed based on a University of Manchester website guide for new students. Both sites followed the same structure which consisted of 5 pages (including the homepage), and a further 8 sub pages (see Figure 36).

The two sites were aesthetically identical (same design, layout, colour, fonts etc.), but differed as the enhanced interactive site contained an avatar guide, interactive links and embedded video, while the standard site did not contain these interactive features (see Figure 37). Content was controlled as far as possible, i.e., the avatar only spoke text that was available on the standard site, while the video only extended the view of the static images presented in the standard design.

Figure 36. The structure of the two bespoke websites (interactive & standard)

Figure 37. The two bespoke website homepages: with avatar guide & videos (right) and without (left)
4.2.1. Avatar and Videos

The avatar was a photorealistic Microsoft Agent (MS Agent) called ‘Vanessa’ and was developed using the 3D Guile Studio\(^1\) and was used within the interactive site to act as a virtual assistant guiding the user through the website. The avatar appeared on every page and varied in size and posture (see Figure 38), and was scripted to appear when a user clicked on a page and gave a short automatic verbal introduction to that page. The avatar’s voice was replaced with a pre-recorded human voice, which synchronised with the avatars expression and movements. Participants could not interact with the avatar, but they were told in advance that they could control the avatar by right clicking on Vanessa, where they had options to either ‘mute’ the avatar, or make her disappear from the page. This action only lasted for that page visit, although when participants navigated to a different page the avatar would appear again. Participants were given limited control to avoid them turning the avatar off at the start of the experiment, to ensure that they experienced the effect of the avatar.

\(^1\) https://guile3d.com/en/

Figure 38. The two different sizes and positions of the avatar: reading (left), and pointing (right)

The interactive site also included short videos that were mixed with static images and distributed throughout the website, with all videos being accessed from the gallery page. Static photographic images used in the standard website, were used to replace the videos, while the gallery consisted only of static images.

4.2.2. Interactive Links and Map

The interactive site contained embedded hyper-links within the text, so when a user hovered over a certain key piece of text it would change colour to indicate a link, and when clicked it took them to another page. In addition, all the images within the gallery were interactive, so if
a participant clicked on an image it would enlarge to fill the screen. The standard site had no text-links and images could not be enlarged.

The campus map used within the interactive site contained hot-links that related to numbered circles within the map (see Figure 39), that represented various buildings around the campus, so when a participant clicked on a circled number on the map an image of that building was shown. The standard site had no interactive hotspots on the map.

Figure 39. The interactive map shows hotlinks (left) when clicked it reveals a picture of the building (right)

### 4.3. Method

To investigate users’ motivations, attitudes and judgements towards the two different websites a mixed methods was adopted that used a triangulation of quantitative questionnaires and qualitative interviews and observation, similar to Study One.

#### 4.3.1. Experimental Design

The two websites used a within-subjects experimental design, where the two websites were counterbalanced across the participant group to avoid order effect. It used a two-way (2x2) repeated measures design for two measurement scales (*Affect & Website Quality*), and a one-way (1x2) repeated measures design for the remaining measurement scales, see Table 27. The two task activities were initial exposure (visual perception) and Task 1 (interactive task), as outlined in the procedure.
Chapter 4 Study Two: Interactivity - Avatars & Media

4.3.1.1. Procedure

The experimental procedure is shown in Figure 40, and was conducted as follows:

1. **Pre Online Questionnaire**: prior to arrival participants were asked to complete an online questionnaire to save time during the experiment, the Power Usage was used to capture individual differences (see Table 29).

2. **Briefing and Familiarisation**: after reading a briefing sheet, completing the consent form and demographic questionnaire, participants were given a short training session in which they were shown a similar university homepage and asked to complete the first two questionnaires (Affect and Website Quality), to familiarise them with the initial-exposure test.

3. **Initial Exposure**: screen shots of the two test websites homepages (see Figure 37) were shown for 0.5 seconds (Lindgaard, Fernandes, Dudek, & Brown, 2006) using PowerPoint, which controlled the exposure time by replacing the image with a mask (grey screen). After completion of each website test, participants completed the same two short questionnaires (Affect and Website Quality).
4. **Interactive Task:** a short navigation task provided a common baseline starting point for all participants. Starting from the homepage, participants visited five pages, which included exploring the campus map and gallery, which exposed participants to the avatar and videos in the enhanced interactive site, or text and static images in the controlled standard design (see Appendix 1). After completing the tasks participants were given up to three minutes to explore the site freely without instruction. After the task and exploration period, participants completed a number of short questionnaires, as shown Figure 40 (see 4.3.2 Measurement Scales for further details).

5. **Interview:** a semi-structured interview was conducted to elicit participants’ preferences and experiences while interacting with the two websites. During the interview participants were verbally asked to complete the criteria judgement questionnaire, and to recall the first remembered features from either of the sites. At the end of the interview they were asked to rank which website they preferred, which completed the study design (see Appendix 8 for semi-structured interview questions).

Each experiment was run using a desktop PC running Microsoft XP, and took place in a university lab. The duration of each experiment took approximately 30 minutes and participants received a £10 Amazon gift voucher in exchange for their time. Quantitative data was captured using a mix of paper-based questionnaires and online-surveys, while qualitative interviews were recorded using a digital recorder.

### 4.3.1.2. Participants

Forty participants (23 female) participated in the study, with age ranging from 18-25 (38%), 26-35 (50%), 36-45 (10%) and 46-55 (2%). The majority of participants were students (78%), while the remaining were university staff. All participants were educated to degree level and came from a variety of subject areas including business (80%), humanities (8%), business computing (5%), and arts (5%) and computing (2%).

### 4.3.2. Measurement Scales

Eleven questionnaire scales (adapted from Study One) were used to capture participants’ quality ratings, as shown in Figure 40. All the scales employed a 7-point Likert scale (apart from Demographics and PANAS). Nine of the measurement scale items are outlined in Table 28, and are summarised as follows:

- **Demographics:** General participant information was collected, e.g., age, gender, education etc., followed by rating awareness of brand quality (Manchester University).
• **Affect:** An 9-item bipolar scale (employed in Study One) designed to capture value-charged affect immediately after each task (Karapanos, Martens, & Hassenzahl, 2010). The scale items were drawn from several sources on arousal, hedonics and emotion, (see Table 3 in Chapter 3 for further information on scale sources). This scale was completed twice after initial exposure and interactive task (see Figure 40).

• **Website Quality:** An 13-item scale composed of a reduced version from the AttracDiff2 scale (Hassenzahl, 2004), including 6 hedonic items, 3 for perceived hedonic quality-stimulation (HQS), and 3 for hedonic quality-identification (HQI); and 4 pragmatic quality (PQ) items, and two global evaluation constructs (EC), goodness and beauty, plus one additional item added to measure participants level of attention (Lindgaard et al., 2006), see Table 28. The two additional measures used in Study One (Perceptual-Design Quality, and Spatial-Design Quality) were omitted to keep the scale short and focus on the hedonic and pragmatic measures only, with only one measure (attention) retained. This scale was completed twice (see Figure 40).

• **PANAS:** An 14-item bipolar scale consisting of a reduced version of the PANAS (Positive and Negative Affect Schedule) scale derived from (Watson, Clark, & Tellegen, 1988), which is split between 7 items measuring positive affect (PA) and 7 items measuring negative affect (NA). Only suitable measures were selected (that applied to website qualities). This scale was included to explore its potential in measuring both positive and negative affective responses. The scale was completed once post-interaction and used as a comparison to the shorter affect scale.

• **Immersion:** An 10-item bi-polar scale designed to capture the level of immersion was developed based on Study One findings, where the scale used simple word ratings to aid fast participant response that minimalized cognitive processing. The items were based on a variety of literature sources (Csikszentmihalyi, 1990; Jennett et al., 2008; Witmer & Singer, Michael, 1998), see Table 28.

• **Usability:** An 5-item scale that captured users perceptions of the functionality, navigation, learnability, convenience and utility of the websites after interaction, taken from (Lavie & Tractinsky, 2004; Tractinsky & Zmiri, 2006). This slightly extended scale from Study One, used a 7-point (agree/disagree) design and was administered once after interaction.

• **Content:** An 3-item scale that captured participants’ ratings on the quality, relevance and the quantity of the websites content, based upon a reduced version of the Bernier Instructional Design Scale (Bernier, 1996), derived from (De Angeli, Hartmann, & Sutcliffe, 2009).
• **Service Quality**: An 3-item scale derived from Lavie & Tractinsky, (2004) measured the reliability, trustworthiness and quality of the websites.

• **Overall Experience**: An 3-item scale asked if participants would ‘visit the site again’, ‘recommend the site to friends’, and to rate how ‘stimulating the overall experience of using the site was’, based on items used by O’Brien, (2010).

• **Criteria Importance**: Participants were verbally asked to rate the importance of five criteria (aesthetics, ease-of-use, content, engaging interaction and brand), in influencing their overall judgement of the quality of the website to gauge if any of these criterion influenced their overall preference. This was scale was completed twice (see Figure 40).

<table>
<thead>
<tr>
<th>Affect</th>
<th>Website Quality</th>
<th>PANAS</th>
<th>Immersion</th>
<th>Usability</th>
<th>Content</th>
<th>Service Quality</th>
<th>Overall Experience</th>
<th>Criteria Importance</th>
</tr>
</thead>
</table>


**Table 28.** Summary of all measurement scales used within Study Two (See Appendix 2 to 5 for Affect, Website Quality, Usability, Content, Quality & Immersion Scales)

A 10-item **Power User** scale was also employed to determine participants’ attitudes towards technology use (Marathe, Sundar, Bijvank, van Vugt, & Veldhuis, 2007), which used a 9-point Likert scale (see Table 29).

**Table 29.** The Power User Scale from Marathe et al., (2007)
4.4. Quantitative Analysis and Results

The analysis is split between the statistical analysis (this section), conducted on the results from the survey questionnaires; and the qualitative analysis (Section 4.5), which used the data drawn from the participant interviews and observation while interacting with the websites.

The quantitative results were analysed using SPSS to identify the main patterns within the main questionnaire data.

4.4.1. Reliability

Questionnaire scores (Affect, Website Quality, Immersion, Usability, PANAS and Overall Experience) were aggregated (averaging individual item scores), since the individual scales all produced high levels of internal reliability with Cronbach alpha (α) ranging from α=0.75 to 0.96, as shown in Table 30. Remaining scales Service Quality and Content showed worse reliability, and so were merged (Service Quality & Content) to product an acceptable α=0.81/0.75 (Nunnally, 1978).

<table>
<thead>
<tr>
<th></th>
<th>Interactive</th>
<th>Standard</th>
<th>Interactive</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect (with Mood)</td>
<td>α=0.92</td>
<td>α=0.93</td>
<td>α=0.95</td>
<td>α=0.95</td>
</tr>
<tr>
<td>Hedonics (HQ) (HQ5 &amp; HQL)</td>
<td>α=0.89</td>
<td>α=0.92</td>
<td>α=0.92</td>
<td>α=0.94</td>
</tr>
<tr>
<td>Pragmatics (PQ)</td>
<td>α=0.78</td>
<td>α=0.86</td>
<td>α=0.77</td>
<td>α=0.79</td>
</tr>
<tr>
<td>Immersion</td>
<td>-</td>
<td>-</td>
<td>α=0.78</td>
<td>α=0.83</td>
</tr>
<tr>
<td>Usability</td>
<td>-</td>
<td>-</td>
<td>α=0.94</td>
<td>α=0.90</td>
</tr>
<tr>
<td>Overall Experience</td>
<td>-</td>
<td>-</td>
<td>α=0.094</td>
<td>α=0.89</td>
</tr>
<tr>
<td>PANAS</td>
<td>-</td>
<td>-</td>
<td>α=0.76</td>
<td>α=0.76</td>
</tr>
<tr>
<td>*Quality-Content</td>
<td>-</td>
<td>-</td>
<td>α=0.81</td>
<td>α=0.75</td>
</tr>
<tr>
<td>Service Quality</td>
<td>-</td>
<td>-</td>
<td>α=0.76</td>
<td>α=0.60</td>
</tr>
<tr>
<td>Content</td>
<td>-</td>
<td>-</td>
<td>α=0.50</td>
<td>α=0.60</td>
</tr>
</tbody>
</table>

Table 30. Cronbach Alpha (α) results for all the main scales

4.4.2. Website Design and Task Differences

A two-way ANOVA was conducted on the Affect scale and sub-scale Website Quality scales items (HQ, PQ, Beauty & Goodness), using website (2) and task (2) as within subject factors. Mauchly’s test indicated that the assumption of sphericity had been violated therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. No significant effect was found when Gender or Age was added as a between-subject factor, or when Brand (Manchester University) was added as a covariate, and this was consistent for all the variables tested (Affect, HQ, PQ, PANAS, Immersion, Usability, Content-Quality and Overall Experience).
4.4.2.1. Affect Scale

Affect produced a significant main effect of the task; $F(1, 39) = 16.11, p < .001, \eta^2 = 0.29$, the website; $F(1, 39) = 20.54, p < .001, \eta^2 = 0.35$, but no interaction. Participants’ affective responses were significantly higher for the enhanced interactivity design website (M=4.4), than for the standard design site (M=3.7), see Figure 41. Also, participants’ affective ratings significantly increased ($p < .001$) from initial exposure (M=3.6) to post-interaction (M=4.4) for both sites, indicating that affect increases as a result of interaction.

![Figure 41. Affect Mean ratings for website (x2) and task (x2)](image)

The individual items within the Affect scale followed the same pattern as the aggregated Affect scale. A comparison between Initial Exposure and the Interactive Task aggregated mean results (all three websites) is shown in Figure 42.

![Figure 42. Aggregated means for single item affect responses comparing the 2 sites](image)
Participants initially found both sites *Curious* and *Interesting*, and this became more evident post-interaction, with all ratings clearly showing an increase as a result of interaction, with *engaging* and *fun* ratings showing the greatest increase.

### 4.4.2.2. Website Quality Scale

A two-way ANOVA was carried out on the *Hedonics (HQ)* and *Pragmatic Scale (PQ)* scale using website (2) and task (2) as within subject factors. The results for *Hedonics (HQ)* revealed a significant main effect of website; $F(1, 39) = 17, p < .001, \eta^2 = 0.3$, task $F(1, 39) = 4.36, p < .05, \eta^2 = 0.1$, and interaction (*website x task*) $F(1, 39) = 8.31, p < .01, \eta^2 = 0.18$. Results for *Pragmatics (PQ)* showed a significant main effect of website; $F(1, 39) = 16.5, p < .001, \eta^2 = 0.3$, task $F(1, 39) = 25.3, p < .001, \eta^2 = 0.4$, and interaction (*website x task*) $F(1, 39) = 9.43, p < .01, \eta^2 = 0.2$.

![Figure 43. Hedonic Scale (HQI & HQS) mean ratings for website (x2) and task (x2)](image)  
![Figure 44. Pragmatic Scale (PQ) mean ratings for website (x2) and task (x2)](image)

Participants gave the same *Hedonic* ratings (M=3.4) for both sites on initial exposure, but after interaction ratings increased more for the enhanced interactive site (M=4.4) than the standard site (M=3.4), see Figure 43. Following a different pattern, PQ ratings for both sites were similar on initial exposure (M=4.6 to 4.7), but after interaction increased more for the standard site (M=5.8) than the interactive site (M=4.9), see Figure 44. The enhanced interactive site was more attractive, but fared worse for PQ (usability) after interaction, although the perception of usability improved for both site designs as a result of interaction.
Participants were asked how much each of the websites held their Attention and a significant difference was found between website: \( F(1, 39) = 9.06, p < .01, \eta^2 = 0.19 \), and task: \( F(1, 39) = 8.67, p < .01, \eta^2 = 0.18 \), but no interaction. Participant attention was held more by the interactive site (M=4.5) than the standard site design (M=3.9), and attention ratings increased after interaction (M=3.8 to M=4.6), see Figure 45.

A two-way ANOVA was carried out on the single items Beauty and Goodness, revealing a significant main effect for Beauty on task \( F(1, 39) = 4.14, p < .05, \eta^2 = 0.1 \), but not for website or interaction. Participants rated both sites more beautiful after interaction (M=3.8 to M=4.2), but no difference was found between sites, indicating an interaction general effect, see Figure 46. No significant effects were found for Goodness.

4.4.3. Analysis of Post-Interaction Experience

A one-way repeated measures ANOVA was conducted on the PANAS (+/-), Immersion, Usability, Quality-Content and Overall Experience Scales (that were captured post-interaction) to identify any differences between the two website designs (interactive and standard).

4.4.3.1. PANAS Scale

The 14-item PANAS scale (see Table 28) was administered once post-interaction. A significant difference was found on Positive Affect (+PA) between website; \( F(1, 39) = 15.31, p < .001, \eta^2 = 0.28 \), with the enhanced interactive site giving higher positive ratings (M=2.9) than the standard website (M=2.3), but no significant difference was found for the aggregated Negative Affect (-NA) scale.
Individual items within the PANAS scales were explored for any interesting patterns. The significant +PA items were; Bold; $F(1, 39) = 14.05, p < .001, \eta^2 = .27$, Surprising; $F(1, 39) = 18.44, p < .001, \eta^2 = .32$, Exciting; $F(1, 39) = 10.12, p < .01, \eta^2 = .21$, and Enthusiastic; $F(1, 39) = 5.81, p < .05, \eta^2 = .13$, all displaying the same patterns as the aggregated +PA. So participants found the enhanced interactive site more surprising, exciting, bold and enthusiastic than the standard site. This concurs with the Affect scale findings, where participant affective responses were higher for the more enhanced interactive site.

Two individual -NA items showed a significant difference, for Irritating; $F(1, 39) = 6.13, p < .05, \eta^2 = .14$, and Distressed; $F(1, 39) = 4.43, p < .05, \eta^2 = .10$, with both displaying the same patterns, where higher ratings were given for the enhanced interactive site for irritating (M=2.3), compared to the standard site (M=1.6), and also for distressed (M=1.6, M=1.3). So despite the positive affective ratings given for the enhanced interactive site, participants also rated it more irritating and distressing, possible reflecting the poor PQ scores.

4.4.3.2. Immersion, Usability, Quality and Content

No significant difference was found between websites for the aggregated Usability, Immersion, Quality-Content, (and the individual sub-themes Service Quality or Content scales), or for any of the single items that make up the scales. This suggests both sites were equally usable, and were rated similarly for their content and quality.

4.4.3.3. Overall Experience

The 3-item Experience scale was aggregated and showed no significant differences, however the single item Stimulating Experience revealed significant differences between sites; $F(1, 39) = 4.33, p < .05, \eta^2 = .14$, with participants rating the interactive site more stimulating (M=4.5) than standard site (M=3.7).

4.4.4. Individual User Experience

The data was further explored to identify any individual differences in user experience. The Power User scale (Marathe et al., 2007) was used to determine whether the participants predisposition towards technology may affect their judgement of the quality of the websites. Cluster Analysis (CA) of the Power User ratings was used to investigate possible sub-groups of users to identify differing attitudes towards technology, and if these groups influenced their overall satisfaction and user experience.

The same two-step CA method reported in Study One (see Chapter Three 3.6) was used to determine the optimal number of clusters. The agglomeration schedule generated for the first Hierarchical Cluster Analysis plots the change in the coefficients per cluster (mean distances
from the combining clusters), and clearly showed a break at the elbow at 4, as shown in in Figure 4.4.1, which suggest that the optimal number of clusters should be three.

![Scree plot showing the coefficient change against the number of clusters](image)

A K-means cluster analysis was performed based on the three-cluster centroids (mean cluster centers) determined by the Hierarchal Cluster analysis. A comparison between the Hierarchal and K-Means Cluster showed little change in cluster groups, with only one item (participant) being assigned to a different cluster. The final three-cluster group is shown in Table 31, which gives the Power User names: IT Positive (High), IT Neutral (Mid) and IT Negative (Low); and the participant number per cluster, cluster centers and standard deviation.

<table>
<thead>
<tr>
<th>Three User Groups</th>
<th>N</th>
<th>Cluster Centres</th>
<th>Std. D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. IT Positive (High Cluster)</td>
<td>16</td>
<td>2.3</td>
<td>1.26</td>
</tr>
<tr>
<td>2. IT Neutral (Mid Cluster)</td>
<td>17</td>
<td>3.1</td>
<td>1.44</td>
</tr>
<tr>
<td>3. IT Negative (Low Cluster)</td>
<td>7</td>
<td>5.4</td>
<td>1.63</td>
</tr>
</tbody>
</table>

N=Number of Cluster Members
Cluster Centres= Mean for each cluster (based on 1-9 Likert Scale: 1=very well disposed to IT, and 9= very adverse to technology)

Table 31. Three cluster groups from cluster analysis of Power User groups

4.4.1. The Effect of Predisposition on Affect & Website Quality
The three User groups were then used as a factor to investigate the effect of prior attitude on users judgement on the quality constructs affect and website quality.

A mixed 3-way between group repeated measures ANOVA was conducted on the Affect and Website Quality scales using task (2) and website (2) as independent variables, and users (x3) as between-group factors, see Table 32. Levene's test did not show a violation of homogeneity of variance for any of the significant between-factor results reported. Since the
main effect for task and website were the same as those reported previously in the 2-way analysis, only the effect of user group differences are reported in this section.

Table 32. Repeated measures (2x2x3) mixed design for Affect and Website Quality

<table>
<thead>
<tr>
<th>Affect Scale</th>
<th>Website Quality Scale</th>
<th>Task (x2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site (x2)</td>
<td>1. Interactive Site</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>2. Standard Site</td>
<td>X</td>
</tr>
<tr>
<td>Users (x3)</td>
<td>1. IT Positive</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>2. IT Neutral</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>3. IT Negative</td>
<td>X</td>
</tr>
</tbody>
</table>

No significant main effect or interaction was found between users for Affect, and only one single item within the scale showed significant main effect between users, which was Attractive; $F(2, 37) = 3.69, p < .05, \eta^2 = .17$, but no interaction was found. Significant differences ($p < .05$) were found between IT Positive (M=3.7) and IT Negative (M=4.4), see Figure 48, where the IT Negative users gave higher ratings for attractive than the IT positive users for the standard site.

Figure 48. User group analysis for Attractive (from Affect scale)

No significant main effects were found in any of the aggregated Website Quality scales including Hedonics (HQI & HQS), Pragmatics (PQ) and Goodness or Beauty. A significant interaction between users was found for Pragmatics (PQ); $F(2, 37) = 5.08, p < .05, \eta^2 = .22$, which revealed that both the IT Negative (standard website M=4.0, interactive site M=5.1), and IT Neutral users (standard M=4.6, interactive M=5.3), gave lower ratings for the interactive site than the standard design; while the IT Positive users gave similar ratings for both sites (M=5.2, M=5.3). Therefore the IT Negative and IT Neutral users found the enhanced interactive site less usable than the standard site, while the IT positive users found both sites equally usable, as shown in Figure 49.
No significant main effects or interactions between users were found on the aggregated +PA (Positive Affect), or -NA (Negative Affect) scales. However, significant differences between users were found on two of the +PA individual items; Enthusiastic; \( F(2, 37) = 3.45, p < .05, \eta^2 = .16 \), and Happy; \( F(2, 37) = 3.24, p < .05, \eta^2 = .15 \). The IT Positive users gave higher ratings for Enthusiasm for the interactive site (M=3.5) compared to the IT Neutral (M=2.8) and IT Negative (M=2.4), and also a similar pattern was found for Happy; with higher ratings given by the IT Positive users (M=3.6), compared to IT Neutral (M=2.7) and IT Negative (M=2.9), although post hoc tests for both items were non-significant, see Figures 50. Overall ratings for the standard site were lower than the enhanced interactive site, with the IT Positive users giving higher ratings for enthusiasm and happy for both sites, which increased more for the enhanced interactive site.

**Figure 49.** User group analysis for PQ (Pragmatic Scale)

**Figure 50.** User group analysis for Enthusiasm (left) and Happy (right) from Positive Affect (PANAS)
No significant main effects between Users or interactions were found on the aggregated Usability, Immersion, Content-Quality and Overall Experience Scales, or any of the individual items within these scales.

4.4.5. Criteria Importance

Participants were verbally asked to rate the importance of five different criteria, which mapped to the questionnaire scales (Aesthetics, Ease-of-Use, Content, Interaction and Brand), to investigate how criterion importance may have influenced their user judgement. Importance of each criterion was rated independently on a 7-point scale where 7=extremely important and 1= not important; mean scores were calculated for each of the criteria ratings, see Table 33. There were no significant difference between sites (one-way repeated measures ANOVA), and all criteria was judged to be approximately equally important by the participants, although brand appeared to be slightly lower, and therefore the least influential.

<table>
<thead>
<tr>
<th></th>
<th>Interactive</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease-of-Use</td>
<td>5.8</td>
<td>5.8</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>5.8</td>
<td>5.7</td>
</tr>
<tr>
<td>Content</td>
<td>5.6</td>
<td>5.6</td>
</tr>
<tr>
<td>Interaction</td>
<td>5.3</td>
<td>5.3</td>
</tr>
<tr>
<td>Brand</td>
<td>4.7</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Table 33. Mean ratings for Criteria Importance items for both sites

4.4.6. Overall Site Preference

At the end of the interview participants were asked to rank their favorite website in preference order. The enhanced interactive website was rated the highest (63% preferred), than the standard site (37%). After the main task participants were given up to 3 minutes to explore each site, and participants explored the interactive site (M=2.5) for a slightly longer time compared to the standard site (M=2.2).

4.4.7. Inter-variable Relationships (Regressions)

Five variables showed inter-site and inter-task differences; Affect, HQ, PQ, Beauty and Positive PANAS. As in Study One, the influence of participants’ affective emotional responses and quality ratings on their overall experience is investigated using a multiple regression analysis to test the hypothesis:

Website quality (HQ & PQ), Affect and Positive Emotion (PANAS) can influence users’ evaluations of Goodness, Beauty and Overall Experience.

A Multiple Regression analysis was used to assess the influence of four independent (predictor) variables on three dependent variables for all both sites (interactive and
standard), across pre and post interaction. Predictors used were: Hedonics (HQ), Pragmatics (PQ), Affect and +PANAS (applied only for post interaction), on the Dependent Variables (1) Beauty, and (2) Goodness and (3) Overall Experience (for post interaction only). Preliminary analysis was conducted to ensure there was no violation of the assumptions of normality, linearity and multicollinearity, and all values were within recommended tolerance (> 10: 0.226 to 0.975), and VIF (< 10: 1.03 to 4.43), (Tabachnick & Fidell, 2007).

For both sites pre and post interaction HQ predicted both Beauty and Goodness. PQ predicted Goodness in both sites pre interaction, however post interaction there was only a weak effect for the Standard site. Affect and HQ were the main predictors for Overall Experience (post interaction) for both sites, and no significant results were apparent for +PANAS. The results are summarized in Table 34.

<table>
<thead>
<tr>
<th>Dependent Variables (DV)</th>
<th>SITES</th>
<th>Predictor (IV)</th>
<th>HQ</th>
<th>PQ</th>
<th>AFFECT</th>
<th>+PANAS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BEAUTY</strong></td>
<td></td>
<td>Pre: (R² = .615)</td>
<td>Post: (R² = .709)</td>
<td>Pre: (R² = .707)</td>
<td>Post: (R² = .740)</td>
<td>Post: (R² = .780)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIG</td>
<td>Β</td>
<td>SIG</td>
<td>Β</td>
<td>SIG</td>
</tr>
<tr>
<td>Enhanced Interactive Site</td>
<td></td>
<td>HQ</td>
<td>***</td>
<td>.702</td>
<td>***</td>
<td>.699</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PQ</td>
<td>ns</td>
<td>.175</td>
<td>ns</td>
<td>-.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AFFECT</td>
<td>ns</td>
<td>.175</td>
<td>ns</td>
<td>.183</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+PANAS</td>
<td>ns</td>
<td>-.113</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GOODNESS</strong></td>
<td></td>
<td>Pre: (R² = .684)</td>
<td>Post: (R² = .732)</td>
<td>Pre: (R² = .696)</td>
<td>Post: (R² = .732)</td>
<td>Post: (R² = .761)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SIG</td>
<td>Β</td>
<td>SIG</td>
<td>Β</td>
<td>SIG</td>
</tr>
<tr>
<td>Standard Site</td>
<td></td>
<td>HQ</td>
<td>***</td>
<td>.636</td>
<td>***</td>
<td>.801</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PQ</td>
<td>ns</td>
<td>.072</td>
<td>ns</td>
<td>.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AFFECT</td>
<td>ns</td>
<td>.210</td>
<td>ns</td>
<td>.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+PANAS</td>
<td>ns</td>
<td>.047</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Table 34. Summary of the regression analysis results for the three sites (pre & post interaction).

Following the same procedure as adopted in Study One, this analysis is compared with the inference model (Hassenzahl & Monk, 2010; Hassenzahl, 2004) to assess how judgement of product qualities might be inferred from general perceptions of Goodness and Beauty. Therefore, further regressions were run to investigate the inference model hypothesis, using Goodness and Beauty as predictors for HQ and PQ.

The results are consistent with the inference model in that Goodness and Beauty predict HQ for both sites, for pre and post interaction (see Table 35). The diminishing PQ effect post interaction indicates a weak PQ-Goodness association (p<.05) in line with the hypothesis. However, the PQ tests showed poor results with a low R² accounting for less than 34% of variance. In contrast, when using PQ for Global constructs, it produced much higher values in the range of 70% with higher significant values (p<.001). Goodness was the main predictor for PQ, although this weakened after interaction, with the standard site showing the greatest difference.
Chapter 4  Study Two: Interactivity - Avatars & Media

<table>
<thead>
<tr>
<th>Predictor (IV)</th>
<th>HEDONICS (HQ)</th>
<th>PRAGMATICS (PQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td><strong>Interactive Site</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goodness</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>Beauty</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td><strong>Standard Site</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goodness</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Beauty</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

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

Table 35. Summary of the general to specific (inference) model regression analysis results for 3 sites

The models generated from the two regression tests are summarised in Figure 51. As reported in Study One (Chapter 3, Section 3.4.8.), HQ shows a consistent relationship with Beauty and Goodness in both the specific to general and general to specific (inference model). Goodness shows a much stronger association with PQ pre interaction in both models. Overall experience is predicted by Hedonics and Affect in the specific to general model, with Hedonics appearing much stronger after interaction.

Figure 51. Summary of the regression results from the specific to general, and general to specific (inference) model perspectives.

4.4.8. Summary of Quantitative Results

A summary of the overall significant effects found in the aggregated measures and single items are shown in Table 36, along with rank of the favored site. The significant main effects found in the quantitative results verify the original main hypotheses set out in Study One, (H1), that websites with interactive features provided more positive UX than standard website designs (menu-link navigation). Participant UX ratings of the interactive website were more positive than standard websites (without interactive features), as reported on all measures (H3), apart from PQ. All UX ratings showed an increase after interaction (from initial exposure to interactive task), indicating that interaction makes a difference (H2).
The enhanced interactive design produced a powerful effect on UX as measured by positive Affect, HQ, Attention, Beauty and Emotion (+PA), as demonstrated by the significant levels. In contrast the standard site (menu-link navigation) was rated higher for PQ. The task effect was considerable for Affect and PQ, while smaller effects were seen for HQ, Beauty and Attention, indicating that interaction generated positive UX.

Clearly the inclusion of enhanced interactive features was considered more visually stimulating (HQ), surprising and exciting (+PA) and increased participants attention, but negative PQ ratings (confusing, complicated, unpredictable) suggest they impacted on usability. However, no differences were found between sites for the usability scale, indicating a difference in the scale measures (PQ-Usability). Despite no inter-site usability differences found, the more interactive site received higher ratings for irritating and distressed (-PA), further indicating the interactive features (e.g., avatar) could be contributing to poor usability, as measured by the PQ scale. Further insight into feature level experience will be explored in the qualitative interview data. No differences were found between sites for content, service quality, indicating both sites was considered similar. No main effects were found on the immersion scale.

The inter-variable relationships were consistent with Study One (Chapter 3). HQ showed a consistent relationship with Beauty and Goodness in both specific-to-general and general-to-specific (inference model). Goodness showed a stronger relationship with PQ pre-interaction in both models, and Affect and HQ were the main predictors for Overall Experience in the specific-to-general model.

Cluster analysis revealed three different user groups (IT positive, IT negative, and IT neutral) based on their predisposition towards technology. The IT positive group rated both sites
equally usable, while the IT negative and IT neutral rated the interactive site less usable than the standard site. Furthermore, the IT positive users rated the interactive site higher for enthusiasm and happy, while the IT negative rated the standard site more attractive. This could explain why some of the interactive features generated some negative affect (irritation and distressed), although it remains unclear which interactive feature these may relate to.

In summary, the general experience of interaction showed an increase in measures (from pre to post interaction). The addition of enhanced interactive features has a strong effect on UX measures and overall preference, where the enhanced interactive site was ranked first over the standard site. The less technically dispersed users rated the interactive site less usable, (and were less happy/enthusiastic), and rated the standard site more attractive. A conflict in measurements between PQ and Usability measures were identified, which will be explored further in the qualitative data. The next section reports on the qualitative analysis, and will investigate users’ opinions towards the different interactive features and how this may influence UX

4.5. Qualitative Analysis and Results

Qualitative data from the participant interviews and observation was analysed using the same three-step method used in Study One (see 3.5.1). The frequency of coded comments were assigned a net valency (NV) and used to compare main themes across the two websites.

4.5.1. Memory Recall

At the start of the interview participants were asked ‘what they remembered the most from either of the two websites’, and the first item participants recalled was recorded as either positive or negative. From a total of 40 most remembered items, 75% of participants (N=30) recalled the avatar, with 57% (N=17) recalling it positively, and 43% negatively. Other items most remembered related to the website content, usability problems, and the interactive links.

4.5.2. Interview Comments

The interview comments were coded into three groups: Interactive site, Standard site and Non-specific. A total of 447 comments were sub-coded using 7 different sub-themes:

- **Attractive**: General (negative or positive) reaction to visual appeal.
- **Content**: General website information quality.
- **Ease-of-Use**: How user-friendly the website is to use.
- **Usability**: Specific website problems (e.g., navigation or operational).
Aesthetics: Interface Design qualities (style, layout, colour, fonts, graphics etc.).

Engaging: General impressions relating the interactive experience (flow/presence/immersion).

Interactive Features: Opinions related to specific features (e.g., avatar, video etc.).

The total percentage frequency and net valency (NV) score was recorded for each theme, as shown see Table 37.

<table>
<thead>
<tr>
<th>Main Website Feature Themes</th>
<th>Interactive Site Freq</th>
<th>Standard Site Freq</th>
<th>Non-Specific Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attractive</td>
<td>7% 28+</td>
<td>2% 11+</td>
<td>1% 4</td>
</tr>
<tr>
<td>Content</td>
<td>5% 17+</td>
<td>3% 1-</td>
<td>10% -4</td>
</tr>
<tr>
<td>Ease of use</td>
<td>2% 9+</td>
<td>2% 7+</td>
<td>6% +19</td>
</tr>
<tr>
<td>Usability</td>
<td>8% 37-</td>
<td>2% 7-</td>
<td>1% -3</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>4% 4-</td>
<td>2% 2-</td>
<td>8% -13</td>
</tr>
<tr>
<td>Engaging</td>
<td>8% 34+</td>
<td>3% 6+</td>
<td>- -</td>
</tr>
<tr>
<td>Interactive Features</td>
<td>22% 32+</td>
<td>5% 21-</td>
<td>- -</td>
</tr>
<tr>
<td>TOTAL</td>
<td>56% 79+</td>
<td>18% 6+</td>
<td>26% -5</td>
</tr>
</tbody>
</table>

Table 37. NV’s for the Main Website Feature themes by site

The enhanced interactive site received a highest frequency of comments (56%) and a far greater positive NV (79+), compared to the standard site (18%, NV=6+), although 26% of the total comments were general without any site-specific reference. Most of the comments for the enhanced interactive site referred to the interactive features (22% of comments), followed by usability and engaging (both 8%), and then attractive (7%). Although the interactive features (NV=32+), engaging (NV=34+) and attractive (NV=28+) received high positive NV scores, usability (NV=37-) in particular, and aesthetics (NV=4-) were adverse. The frequency distribution of comments for the standard site design was similar for all themes; however, its interactive features received many adverse comments (NV=21-), even though few actual usability problems occurred, with comments referring to its lack of interactivity.

The non-specific comments consisted mainly of general comments relating to the content (10%), where some participants “felt disappointed because there wasn’t much information” [P2], while other felt both sites “were informative”[P15], and “interesting to read”[P2]. Both sites were considered to be “simple and very easy to use” (+19), although the aesthetics for both sites were criticized for looking “really simple and not really professional” [P6], probably due to the simple condensed design of the two test sites.

Although the content of the two sites was exactly the same, the enhanced interactivity version was perceived to have more content, (5%, NV=17+) compared to the standard site (3%,...)
NV = 1-), possibly due to the inclusion of videos; “The one (site) with the videos obviously gives more information” [P28]; and the avatar, “The lady was showing us where we can find some relevant information” [P19]. Whereas the standard website was considered to have too much content; “There was too much information...like you are trying to read it and its so boring, you just skip over it” [P29], making it appear far more text-based, “it had more content in terms of description... which is boring to the user to look at” [P7].

Many of the negative comments for the standard design website referred to the lack of interactivity; “Everything was too static... I found it boring basically” [P30], and, “It wasn’t fantastically wow, its pretty basic” [P2]. The lack of interactive features most commented upon were the absence of interactive links; “Its basic...there’s hardly anything to click on” [P40]; no hyperlinks on images, “When I clicked on pictures, I tried to show them full size, but I couldn’t” [P37]; along with the absence of videos; “There were just pictures in the gallery... no videos” [P2]. Overall, the participants’ comments indicated that they favored the more interactive site, despite reporting usability problems. To understand why the interactive site was favored above the standard site, the comments across the 7 themes were further divided into sub-groups to compare the differences at the feature level, see Table 38.

<table>
<thead>
<tr>
<th>Comment Category</th>
<th>Freq</th>
<th>%</th>
<th>NV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactive Features</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avatar</td>
<td>59</td>
<td>26.9%</td>
<td>-5</td>
</tr>
<tr>
<td>Video</td>
<td>22</td>
<td>10%</td>
<td>+22</td>
</tr>
<tr>
<td>Interactive Links</td>
<td>9</td>
<td>4.1%</td>
<td>+9</td>
</tr>
<tr>
<td>Interactive Map</td>
<td>8</td>
<td>3.7%</td>
<td>+6</td>
</tr>
<tr>
<td>TOTAL</td>
<td>98</td>
<td>44.7%</td>
<td>+32</td>
</tr>
<tr>
<td>Attractive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avatar</td>
<td>15</td>
<td>6.8%</td>
<td>+15</td>
</tr>
<tr>
<td>Video</td>
<td>6</td>
<td>2.7%</td>
<td>+6</td>
</tr>
<tr>
<td>General</td>
<td>9</td>
<td>4.1%</td>
<td>+7</td>
</tr>
<tr>
<td>TOTAL</td>
<td>30</td>
<td>13.7%</td>
<td>+28</td>
</tr>
<tr>
<td>Aesthetics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avatar</td>
<td>9</td>
<td>4.1%</td>
<td>-9</td>
</tr>
<tr>
<td>Structure</td>
<td>4</td>
<td>1.8%</td>
<td>+2</td>
</tr>
<tr>
<td>General</td>
<td>5</td>
<td>2.3%</td>
<td>+3</td>
</tr>
<tr>
<td>TOTAL</td>
<td>18</td>
<td>6.2%</td>
<td>-4</td>
</tr>
<tr>
<td>Engaging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avatar</td>
<td>16</td>
<td>7.5%</td>
<td>+14</td>
</tr>
<tr>
<td>Video</td>
<td>5</td>
<td>2.3%</td>
<td>+5</td>
</tr>
<tr>
<td>General</td>
<td>15</td>
<td>6.8%</td>
<td>+15</td>
</tr>
<tr>
<td>TOTAL</td>
<td>36</td>
<td>16.4%</td>
<td>34+</td>
</tr>
<tr>
<td>Usability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avatar Mute Control</td>
<td>20</td>
<td>9.1%</td>
<td>-20</td>
</tr>
<tr>
<td>Avatar Pop-Up</td>
<td>6</td>
<td>2.7%</td>
<td>-6</td>
</tr>
<tr>
<td>Avatar Pace</td>
<td>6</td>
<td>2.7%</td>
<td>-6</td>
</tr>
<tr>
<td>Interactive Map</td>
<td>5</td>
<td>2.3%</td>
<td>-5</td>
</tr>
<tr>
<td>TOTAL</td>
<td>37</td>
<td>16.9%</td>
<td>-37</td>
</tr>
</tbody>
</table>

Table 38. Frequency and net valency of comments for the Interactive Features for the interactive site

For the interactive site, the avatar attracted most of the comments made about the interactive features (26.9%), although overall there was a slight negative NV (-5), which reflects the polarization of opinions towards the avatar. As some participants found the avatar “intriguing” [P10], and “innovative, as it adds something new and fresh” [P17], which “adds to
the experience” making participants feel more “involved”[P30]; while others found the avatar “quite distracting”[P2], “very annoying”[P35], and “very irritating while you are trying to read the website”[P4], because it “made my attention divert from the actual content” [15], causing participants to feel they were “getting lost” and “confused”[P7]. Although the videos received fewer comments (10%), these were all positive (+22), as participants found the inclusion of videos “absorbing”[P30] as they bring “you closer to the event”[P38] and “kept my interest longer than just reading content”[P34]. However, a few participants (2%) reported not noticing the videos at all. The interactive links (4.1%, +9) and interactive map (3.7%, +6) shared a similar frequency of comments and NV ratings, with the links being considered “quite helpful”[P3], as they added “some movement”[P30], which “made it more stimulating to actually explore”[P23]. Although the map was also thought to be a “good idea”[P27], which was “very useful”[13] and “worked very well”[P29], it did receive a few poor usability comments (2.3%, NPV -5) due to the map not being fully interactive; “I tried to click, but it did not really work... sometimes you have to click twice to make it (photo) appear”[P6], and sometimes “you click on the map and some of it has photos and then some hasn’t... it can be quite a frustrating experience”[P22].

Attractiveness comments referring to the avatar (6.8%) were nearly all positive (+15), as were engaging avatar-related comments (+14, 7.3%), where the avatar “added another dimension”[P3], because she “jumps right out at you”[P3], which “stimulates your aural senses”[P27]; especially when “she talks, its more engaging because she’s telling you about it...so it feels nice”[P16]. In contrast the avatar-related comments for aesthetics (4.1%) and usability (14.6%) were negative (-9, -32). Negative aesthetics comments referred to the size and realism of the avatar, such as; “she’s too big compared to the size of the page... she’s taking up too much space”[P6], and her appearance, “I found her a bit tacky... somehow it didn’t fit with the website”[P23]. Quite a few participants found her “kind of scary”[P1], and “quite frightening”[P6], as “it felt very artificial”[21] and “really computerised”, with a few participants preferring a “cartoon or maybe something a little more real”[P27].

The video feature were rated more favourable for attractiveness (+6) and engagement (+5), with no adverse comments, with the same pattern apparent for the other general aspects related to the site, which have been grouped under ‘general’ in Table 38. Nearly all the usability problems pertained to the avatar, with the lack of mute control being the most frequent (9.1%). The following excerpts illustrate the problems, many of which could be attributed to poor usability design: “When I clicked on a new page she popped up again, which was annoying”[P9], as “she’s speaking a lot of the time, even though I put her on mute”[P38]; not being able to control her talking; “she begins to talk automatically even though I didn’t click anything”[P20], especially when listening to videos, “I clicked into the videos so I could
hear the content, at the same time she was speaking” [P11]; and confusion about the pop-up photos that the avatar was referring to when she spoke (2.7%), “these photos would just start coming out and you couldn’t figure out the logic of what makes them come out…. it was kind of irritating” [P1], as it caused disruption and “it was really confusing… as it was disturbing a lot of my reading” [P7]; lastly was the pace of the avatar (2.7%), “because her pace and my pace is not syncing, so its difficult to follow her” [P7], as “there were some delays on her speaking” [P17]. Generally participants felt they did not have enough control over the avatar; “When you turn off the assistant there's no way of getting her back” [P11]; and they were unable to, “adjust the volume” [P14], or alter the speed at which she was speaking, “the pace at which she was speaking was very slow” [P21]. The interactive map also sustained a few adverse comments, but these were minor in comparison to the avatar comments.

4.5.3. Qualitative Results Summary

In summary, the qualitative results confirm the quantitative findings that the enhanced interactive design produced more favourable user experience than the standard design. The qualitative analysis provided greater insights into the perceived differences between the two website designs. All the interactive features were generally well received, while the standard design website was considered boring and rated less favourable due to its absence of interactive features. Interestingly, this negative view appeared to influence participants’ perception of content, despite both sites having identical content, the standard site appeared to have more text-based, due to the lack of engaging interactive feature, thus creating a ‘negative halo’ effect on content. This was supported when content comments were examined according to the presentational order, where more negative comments were given when participants had seen the interactive site prior to the standard site, than vice versa.

The interactive features generated the most positive comments, which showed some interesting differences. The avatar had the most prominent effect, with conflicting comments, with some considering it attractive and engaging, while other criticizing it for being irritating, artificial and unsuitable for the context of the website. It also received many negative usability and aesthetic comments, which could be due to the limited option to interact with the avatar directly resulting in feeling a lack of control. This highlights the importance for more intuitive ‘intelligent interactive’ avatars that are well-designed and interactive (Qiu & Benbasat, 2005). The other interactive features (video, interactive map and links) were well received and generally found more attractive. It appears that the interactive character, media and active links can lead to enhanced user engagement, where users overcome considerable usability problems.
4.6. Chapter Summary

This chapter first verified the hypothesis originally set out in Study One in two ways, first interaction enhances UX, and second, websites with enhanced interactive features provides more positive UX than standard websites without interactive features, as supported by both the quantitative and qualitative results.

Adopting a controlled experiment, the effect of interactivity was evaluated by controlling confounding variables (aesthetics, content, context) that may influence UX. Interaction clearly made a difference, with affect and hedonic ratings increasing after interaction (pre-post interaction), and especially for the interactive site. Clearly the interactive features had a strong effect on user engagement, where the interactive site was ranked first over the standard site for overall preference.

However, the interactive website generated worse PQ ratings, where the standard design outshone, although both sites were rated equally for usability, indicating a measurement conflict. Despite positive affect ratings, negative ratings for irritation and distressed can be seen as a result of the interactive features. A contributory factor could be the differing attitudes of technology, where user disposition towards technology appeared to influence their judgement of the websites; where IT positive users were more tolerant of usability problems then IT negative or neutral users. However, overall the interactive site was the preferred choice, indicating that poor usability did not adversely influence users overall choice.

Inter-variable relationships (regressions) were consistent with Study One, which partly supports the inference model, where HQ was associated with Goodness and Beauty, although the PQ-Goodness link was less consistent post-interaction; while Affect and HQ were the main predictors for Overall Experience. This supports the hypothesis that users' judgement of product quality (as measured by HQ & PQ), influences their overall judgement of global qualities (Goodness & Beauty), impacting on overall experience.

The qualitative data provided insights into the reasons behind the difference between the two sites at the feature level. All interactive features were well liked, with the standard website design being considered dull and boring due to its lack of interaction. Despite the avatar having the most positive effect, it received many negative usability and aesthetic comments, which highlights the need for well-designed interactive features that are implemented seamlessly allowing users to have adequate control over their experience.
Chapter 5  Study Three: Longitudinal Study (Quantitative)

This chapter reports on a six-month ecological longitudinal study that examines the temporal patterns of UX when using new tablet technology. In contrast to previous longitudinal studies in HCI, which captured retrospective accounts of UX that may be subject to a reconstructed memory bias (Karapanos, Martens, & Hassenzahl, 2009b), this study took place in the field by investigating UX using a variety of concurrent experience sampling methods. Adopting the same mixed methods approach used within the previous two studies, this study not only explores how UX changes over time, but also examines how the complexity of applications and device impacts on product satisfaction.

In this chapter the main study design and quantitative findings will be presented, Chapter 6 reports on the quantitative findings and overall discussion.
5.1. Introduction

Past UX research has focused on the early stages of the process where users first encounter IT products and make initial judgements about its quality (Hassenzahl & Monk, 2010; Lavie & Tractinsky, 2004; Lindgaard et al., 2006). More recent calls for longitudinal UX studies and ecological studies 'in the wild' (Hassenzahl & Tractinsky, 2006) have proposed methods for assessing experience over time, based on graphical reconstruction of feelings (Karapanos et al., 2009) and experience sampling (van Schaik, Hassenzahl, & Ling, 2012). The few longitudinal studies that have investigated UX used relatively small samples (Wright & McCarthy, 2010) due to the cost for both the researcher conducting such studies and the users participating within them (Menard, 1991).

The previous two chapters (Study 1 and 2) identified how specific interactive design features enhance positive affect while using websites. Clearly, interactivity is an important component of UX (Teo et al., 2003), yet how much does interactivity influence long-term product satisfaction remains unclear. The advent of tablet computers such as the iPad has freed the user from the constraints of a keyboard and mouse and created new ways to consume information in a digital form. The device provides interactivity through a multi-touch interface, combined with the wide array of interactive applications that capitalize on the touchscreen technology providing visual, auditory and kinaesthetic experiences. Understanding the long-term use of new interactive technologies is an important area of UX in order to improve design and evaluation (Kujala, Vogel, Pohlmeyer, & Obrist, 2013).

The patterns of UX are investigated in the next two chapters (5 & 6), which reports on a 6-month longitudinal study of new users of tablet technology. The quality judgement model of UX (Hartman et al., 2008) is further expanded by investigating the decision-making process over long-term, in order to understand what key constructs of UX may influence user satisfaction and product adoption. User diversity is also considered, where individual differences in attitudes and experience may be related to group level attitudes and judgements, and how these may change over time.

Four research questions were used to motivate this study:

- How does UX change over time and product (iPad) usage?
- What relationship exists between UX and user satisfaction/technology adoption?
- How heterogeneous are users’ experiences and attitudes to adoption, and how do individual differences contribute to group level judgements?
- How do user predispositions towards technology impact upon their experience and adoption attitudes?
The study findings are split between two Chapters (5 & 6), where Chapter 5 first outlines the study design and methods, and then reports on the main quantitative results from the various questionnaire methods used to capture UX over the 6-month time frame; whereas Chapter 6 reports on the main qualitative analysis and findings drawn from the interview data, and further explores the contributory factors that may lead to product non-adoption.

5.2. Study Design

A six-month longitudinal, mixed methods study was designed to capture the changing experiences of medical students using iPads to support their learning in hospital environments. The majority of students were novice users, who were given iPads to support their clinical studies during their first year of hospital-based training, and during their project group-work learning. However iPad usage was discretionary, so no set tasks were required as the study aimed to capture participants ‘natural’ experiences while using the device, where both work and leisure use was captured.

5.2.1. The iPad

Participants received a new iPad 2 16GB Wi-Fi only device at the beginning of their 3rd academic year to support hospital-based learning. The iPads were not 3G enabled, so students relied on wireless areas around hospital placements and within the university. Every iPad came preinstalled with various applications (apps) provided by the university for students to use while on hospital placements, these included Blackboard VLE (Virtual Learning Environment) and Numbers (spreadsheet) app.

5.2.2. Participants

Fifty-one participants (65% females) aged between 18-25 (N=45) and 26-35 (N=6) took part in the study, with 4 (female) participants dropping out halfway through the study, leaving a total of 47 to complete the final study. All participants were recruited from Manchester Medical School in the UK. Participants volunteered to take part in the study through an online questionnaire and received incentives of £25 after completing the first 3 months of the study, and a further £25 upon completing the full 6-months of the study. In addition two free applications Keynote (Apple version of PowerPoint) and GoodReader (PDF reader/annotator) were given at the start of the study to every participant in order to obtain a consistent base line of UX data for the study. All participants owned a mobile phone with 88% being a smartphone (see section 5.3.1.), and a further 7 previously owned an iPad.
5.2.3. **Method**

A mixed method approach was adopted that triangulated data from *Questionnaires* (x3), *Semi-Structured Interviews* (x3), and a *Weekly Diary Study* (x24), over the 6-month time frame, as shown in Figure 52.

**Figure 52. Plan of the study methods over six months**

5.2.3.1. *Questionnaires*

A mix of online and paper based questionnaires was used to collect data on demographics, activities, apps, affect and various quality ratings, (see measures section 5.2.4). Questionnaires were applied three times over the 6-month study (Early-Study, Mid-Study and Final-Study), as shown in Figure 52.

5.2.3.2. *Semi-Structured Interviews*

Each participant was interviewed three times after they had completed the questionnaires (*Early, Mid & Final-Study*) see Figure 53. At the beginning of the *Early-Study* interview participants were asked to think back to before they received their iPad and to retrospectively verbally rate their previous expectations on eight different affect and global items (see Verbal Interview Ratings 5.2.4.3.), which they then repeated at the end of the interview, with ratings based on their current experiences. For the *Mid/Final-Study* interviews participants were first asked to verbally rate the same eight items to elicit how they currently felt about using their iPad. Participants were then interviewed using semi-structured questions, where they were asked to describe their main activities, preferences, barriers and motivations of use. The *Final-Study* interview was extended to enable participants to retrospectively reflect on their iPad usage over the past 6-months using participants' individual time-lines extracted from the diary study that acted as prompts for
discussion during the interview (Karapanos, Martens, & Hassenzahl, 2010; Kujala et al., 2013). At the end of the Final-Study interview participants were asked what apps they most recalled (see Appendix 8 for summary of the semi-structured interview questions used).

Figure 5.3. An overview of the structure for the three interviews (Early, Mid, Final).

5.2.3.3. Diary Study
Weekly online mini-questionnaires were sent to every participant via email to record their weekly activities, affective reflections and experiences. Response rates were closely monitored every week and reminder prompts were sent to avoid non-responses, thus ensuring consistent data throughout the 6-month study. Participants were also prompted to report up to four good or bad experiences they had using their iPad that week using a free-form format entry. The weekly mini-questionnaire activities and affective experiences for each participant was captured and displayed in individual time-line graphs that were used during the final interview as prompts.

5.2.4. Measures
A mix of interview, diary study mini-questionnaires, short interview verbal ratings and application ratings were employed (see Table 39), along with an individual difference scale, as now reported:
5.2.4.1. Questionnaire Scales

Seven questionnaire scales were used three times during the 6-month study (Early, Mid & Final-Study). Participant demographics were collected during the Early-Study, (e.g., age, gender, hospital location, brand rating and current technology use). All scales employed 7-point Likert scale, which were amended from the previous two studies (reported in Chapters 3 & 4), as follows:

• Affect: To capture negative affect two additional items (frustration and boredom) were added to the original scale (used in Chapter 3 & 4) resulting in an 11-item bi-polar scale that captured positive affect (9-items), and negative affect (2-items).

• Design Quality (DQ): An 14-item bi-polar scale consisting of a reduced version of only AttracDiff2 scale items that include 4-items from pragmatic quality (PQ), hedonic quality-identification (HQI), hedonic quality-stimulation (HQS), and two global constructs (GC), (Hassenzahl, 2004).

• Usefulness: An 8-item scale comprising of a reduced version of the Technology Acceptance Model (TAM) scale, which capture participants’ perceived usefulness (Davis, 1993).

• Overall Experience: An 4-item scale with general evaluation rating questions adapted from the previous studies (Chapter 3 & 4) based on (Cyr, Head, Larios, & Pan, 2009).

• App Ratings: A 7-item app rating scale was used to capture participants’ judgement of a number of different apps used throughout the study.

In addition participants iPad Activity (work & leisure) was captured over the 6-month study.

<table>
<thead>
<tr>
<th>Affect</th>
<th>Design Quality</th>
<th>Usefulness</th>
<th>Overall Exp.</th>
<th>Apps</th>
<th>Activities</th>
<th>Affect</th>
<th>Verbal</th>
</tr>
</thead>
<tbody>
<tr>
<td>(PA) Affect</td>
<td>Hedonics (HQS)</td>
<td>Utility (PU)</td>
<td>Motivated</td>
<td>Often Used</td>
<td>Learning*</td>
<td>(PA) Affect</td>
<td>(PA) Affect</td>
</tr>
<tr>
<td>Fun+</td>
<td>Cheap (HQS)</td>
<td>Performance</td>
<td>Recommend</td>
<td>Utility</td>
<td>Internet</td>
<td>Enjoy+</td>
<td>Enjoy</td>
</tr>
<tr>
<td>Attractive+</td>
<td>Tacky (HQS)</td>
<td>Productivity</td>
<td>Enjoyable</td>
<td>Usability</td>
<td>Email</td>
<td>Engage+</td>
<td>Engage</td>
</tr>
<tr>
<td>Curious+</td>
<td>Amateurish (HQI)</td>
<td>Effectiveness</td>
<td>Satisfied</td>
<td>Content</td>
<td>Notes</td>
<td>(NA) Affect</td>
<td>(NA) Affect</td>
</tr>
<tr>
<td>Interesting+</td>
<td>Challenging (HQS)</td>
<td>Quality</td>
<td>Learn</td>
<td>Importance</td>
<td>VLE</td>
<td>Boring -</td>
<td>Boring</td>
</tr>
<tr>
<td>Pleasurable+</td>
<td>Dull (HQS)</td>
<td>Time</td>
<td>O. Experience</td>
<td></td>
<td>Reading Groups</td>
<td>Frustrating-</td>
<td>Satisfaction (GC)</td>
</tr>
<tr>
<td>Absorbing+</td>
<td>Conservative (HQS)</td>
<td>Easier to Use</td>
<td></td>
<td></td>
<td>Learning Apps</td>
<td>Leisure *</td>
<td>Leisure</td>
</tr>
<tr>
<td>Exciting+</td>
<td>Commonplace (HQS)</td>
<td>Useful</td>
<td></td>
<td></td>
<td>Internet</td>
<td>SN</td>
<td>Ease-of-Use</td>
</tr>
<tr>
<td>Engaging+</td>
<td>Pragmatics (PQ)</td>
<td>Effort</td>
<td></td>
<td></td>
<td>Video</td>
<td>Groups</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>Enjoyable+</td>
<td>Confusing (PQ)</td>
<td></td>
<td></td>
<td></td>
<td>Photo</td>
<td>Music</td>
<td>Attractive</td>
</tr>
<tr>
<td>(NA) Affect</td>
<td>Complicated (PQ)</td>
<td></td>
<td></td>
<td></td>
<td>Apps</td>
<td>Leisure Apps</td>
<td>Attractive (PA)</td>
</tr>
<tr>
<td>Frustration-Boring-</td>
<td>Unpredictable (PQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beauty (EC)</td>
<td>Impractical (PQ)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goodness (EC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Table 39. Summary of all measurement scales used within the longitudinal study three
(See Appendix 2, 3, 6 & 7 for examples of Affect, Design Quality, Usefulness and Diary Study Scales)
5.2.4.2. Diary Study Mini-Questionnaires

Three mini-questionnaire scales were sent every week via email over 24 weeks, as follows:

- **Learning Activity:** An 7-item scale asked participants if they had performed any learning based activities on their iPad over the last week (see Table 39).

- **Leisure Activity:** A 7-item scale captured participants leisure based iPad activities over the last week (see Table 39). All activities were recorded on a 1-5 Likert scale (1= no activity that week, 5=several times a day).

- **Affective Experience:** A 5-item scale (based on the 11-item main Affect questionnaire scale was used to capture participants’ weekly affective responses. These consisted of two positive affect (enjoyable and engaging), and two negative affect (frustrating and boring) items, along with one global measure, overall satisfaction. Responses were recorded on a 1-7 Likert scale (1=low, 7=high), (see Appendix 7 for the diary-study mini questionnaire).

5.2.4.3. Verbal Interview Ratings

Seven items were verbally rated at the start of each interview using a 1-7 rating score, with positive affect (enjoyable and engaging), negative affect (frustrating and boring) and three global measures overall satisfaction, attractive and ease-of-use.

5.2.4.4. User Predispositions

A reduced 10-item version of the same Power User (PU) scale employed in Study 2 was employed in the last month of the study to determine participants’ attitudes towards technology (Marathe et al., 2007) which used a 7-point Likert scale.

5.3. General Results

The general trends in media usage (iPad, computer and mobiles); the changes in iPad activity over time; and an overview of the favored applications usage and their ratings are now reported.

5.3.1. Device Usage

At the start of the study 67% (33) of the participants used a PC computer or laptop, 29% (14) used a Mac, while the remaining 4% (2) used both platforms (see Figure 54). All participants owned a mobile phone at the start of the study, split between the iPhone (43%), Android (35%), Blackberry (10%), and the remaining (12%) owning older type phones (Samsung or Nokia), as shown in Figure 54.
Participants (N=40) rated the Apple Brand slightly higher (M=2.5) compared to the Android Brand (M=3.2). A total of 14% (7) participants had owned an iPad (Apple iOS) prior to taking part within the study, and the length of ownership time varied from 1-3 months (4%), 3-6 months (2%) and longer than 6 months (8%). The effect of prior experience was analyzed and no impact on the results was found (see 5.5.2.3 Gender, Location and Pre-Owned Checks).

### 5.4. Activity Changes Over Time

Participant activities and affective responses were captured across the 24 weekly mini-questionnaires to explore any changes over time. All activity (work and leisure) showed a gradual decline over time (apart from watching videos), with checking emails, browsing the Internet and social networking being the most frequent activities occurring daily. Overall satisfaction and positive affect remained high throughout the study, while negative affect showed a gradual decline, (see Appendix 9 for participants aggregated diary activity and affect timeline graphs).

A variety of different work and leisure activities were captured in the main questionnaires (Early, Mid and Final-Study), according to three main devices used (computer/laptop, mobile phone and iPad). Nearly all learning and leisure activities were shared between the iPad and computer/laptop, with the iPad being the preferred device for checking emails, reading, accessing medical apps, use within groups and playing games. The mobile phone was favoured over the computer for checking emails, social networking, playing games and taking photos. Little difference in activity was shown over time, with a decline in using the iPad to play games, and a drop in using the computer for reading. An increase in accessing medical apps was found for both the computer and phone, which could be due to favourite apps that were first used on the iPad being accessed later on alternative devices (see Appendix 10 for activity usage graphs and summary table).
5.5. Quality Variables Analysis Over Time

The statistical results report first on analysis using a One-Way ANOVA on the data captured from the three questionnaires (main questionnaires, diary-mini-questionnaires and verbal interview ratings), to explore changes over time.

5.5.1. Questionnaire Results

Repeated measures one-way ANOVA was conducted using SPSS on the (1) Main Questionnaires, (2) Diary Mini-Questionnaires and the (3) Verbal Scale results (see Table 41 for Mean Scores and SD results). Where Mauchly's test indicted that the assumption of sphericity had been violated the Greenhouse-Geisser correction was used.

All questionnaire scores were aggregated (averaging individual items scores) and all scales proved good levels of internal reliability using Cronbach’s alpha coefficient ranging from .61 to .93 (see Table 40). No missing data was found as forced choice questionnaires were used.

<table>
<thead>
<tr>
<th></th>
<th>Early</th>
<th>Mid</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect (Aggregated)</td>
<td>0.85</td>
<td>0.92</td>
<td>0.88</td>
</tr>
<tr>
<td>Hedonics (HQI &amp; HQS)</td>
<td>0.61</td>
<td>0.73</td>
<td>0.72</td>
</tr>
<tr>
<td>Pragmatic Quality (PQ)</td>
<td>0.71</td>
<td>0.69</td>
<td>0.68</td>
</tr>
<tr>
<td>Usefulness (TAM)</td>
<td>0.86</td>
<td>0.91</td>
<td>0.86</td>
</tr>
<tr>
<td>Overall Experience</td>
<td>0.88</td>
<td>0.91</td>
<td>0.93</td>
</tr>
</tbody>
</table>

Table 40. Reliability results for all scales using Cronbach Alpha coefficient

5.5.1.1. Affect Scale

A one-way repeated measure ANOVA was conducted on the 9-item Positive Affect aggregated items (fun, enjoyable, attractive, curious, interesting, pleasurable, absorbing, exciting and engaging), and the 2-item Negative Affect aggregated items (frustrating and boring) to see if there were any significant differences over the three time frames. A significant difference was found on the Positive Affect scale ($F(1.7, 77.5) = 4.49, p < .05, \eta^2 = .09$). Post hoc tests revealed a significant decrease ($p < .05$) between Early-Study (M=5.8) and Mid-Study (M=5.5), see Figure 55.

No significant difference was found on the Negative Affect items, but the single item Frustrating was analyzed to explore any relationship with pragmatic quality (usability) and significant difference was found ($F(2, 92) = 4.77, p < .05, \eta^2 = .09$), where post hoc tests reveal significant decrease ($p < .05$) between Early (M=2.9) and Mid-Study (M=2.3) ratings, see Figure 55.
5.5.1.2. Design Quality Scale

A one-way repeated measure ANOVA was carried out on the Design Quality sub-scales, and a significant difference was found for Pragmatic Quality (PQ): $F(2, 92) = 6.19, p < .01$, $\eta^2 = .12$, Hedonics (HQS & HQI): $F(2, 92) = 5.65, p < .01$, $\eta^2 = .11$, and the Global Construct (GC) Beauty-Ugly: $F(2, 92) = 3.73, p < .05$, $\eta^2 = .08$. Post hoc test revealed a significant increase ($p < .001$) for PQ from Early-Study ($M=5.4$) to Final-Study ($M=5.8$), and for Beauty ($p < .05$) from Mid-Study ($M=5.7$) to Final-Study ($M=6.0$), see Figure 56. Hedonics post hoc tests revealed a significant decrease ($p < .05$) from Early-Study ($M=5.8$) to Mid-Study ($M=5.5$), and then increased slightly (Final: $M=5.7$) in the last 3 months, (although not significant), indicating a mid-study dip, as shown in Figure 56, (see Table 41 for mean scores). No significant effect was found for the single Global Construct (GC) Goodness.

![Figure 55](image1.png)

**Figure. 55.** Mean scores for Positive Affect, and the single item Frustrating (from Negative Affect)

![Figure 56](image2.png)

**Figure. 56.** Mean ratings for Pragmatic Quality (PQ), Hedonics (HQI & HQS) and Beauty-Ugly (GC)
5.5.1.3. Usefulness and Overall Experience Scales

A significant difference was found on the Usefulness scale: $F(2, 96) = 3.53$ $p < .05$, $\eta^2 = .07$, with post hoc test revealing a significant increase ($p < .05$) from Early-Study ($M=5.5$) to Final-Study ($M=5.7$), as shown in Figure 57.

![Figure 57. Mean ratings for Usefulness, and single item Motivated](image)

No significant difference was found on the aggregated Overall Experience scale, however the individual item ‘Motivation’ was analyzed to determine if this was a predictor of user satisfaction (Teo et al., 2003; Zhang & Von Dran, 2000), and a significant difference was found for Motivated: $F(1.7, 76.6) = 4.82$ $p < .05$, $\eta^2 = .10$. Post hoc test showed a significant increase ($p < .05$) between Early-Study ($M=5.9$) and Final-Study ($M=6.3$), with the sharpest increase occurring within first 3 months of usage, see Figure 57.

An overview of the mean patterns, standard deviations and post hoc results for all the significant main questionnaire scales are shown in Table 41.

<table>
<thead>
<tr>
<th></th>
<th>Early-Study</th>
<th>Mid-Study</th>
<th>Final-Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><em><em>p&lt;0.05</em>, p&lt;0.001</em>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Affect*</td>
<td>5.8</td>
<td>6.7</td>
<td>5.6</td>
</tr>
<tr>
<td>Frustrating* (Negative Affect)</td>
<td>2.9</td>
<td>1.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Pragmatic Quality (PQ)**</td>
<td>5.4</td>
<td>0.95</td>
<td>5.6</td>
</tr>
<tr>
<td>Hedonics (HQI &amp; HQS)**</td>
<td>5.8</td>
<td>0.56</td>
<td>5.5</td>
</tr>
<tr>
<td>Beautiful/Ugly*</td>
<td>5.9</td>
<td>0.90</td>
<td>5.7</td>
</tr>
<tr>
<td>Usefulness*</td>
<td>5.5</td>
<td>0.90</td>
<td>5.6</td>
</tr>
<tr>
<td>Motivated* (Experience Scale)</td>
<td>5.9</td>
<td>1.11</td>
<td>6.2</td>
</tr>
</tbody>
</table>

*Bold Items = Aggregated Scales, Not Bold Items = Single Scale Items*

Table 41. Results for the significant scale items from the main questionnaire across each 3 time frames
5.5.1.4. Summary of Main Questionnaire Results

Positive Affect (fun, enjoyment and curiosity) dropped over time, especially in the first 3 months of usage, but increased slightly thereafter, indicating a familiarisation or novelty effect. Frustration showed a similar pattern as initial frustrations were either overcome or just accepted. Pragmatic Quality (PQ) and Usefulness increased over time as participants found using their iPads clearer, simpler, more predictable, the more useful the iPad became. Little difference was found in Hedonics (HQ) over time, apart from a mid-study dip, where the iPad was considered more commonplace and less challenging over time. Participants rated the iPad more Beautiful in the last 3 months of usage, indicating a possible halo effect from PQ (Usability). No significant difference was found on the aggregated and overall experience scale, but motivation increased significantly over time, especially in the first 3 months of usage, indicating an initial novelty or learning phase, where motivation was as it highest.

5.5.2. Diary Mini-Questionnaires

To capture any changes over the 6-month diary study a one-way repeated measure ANOVA 1x5 (Week 1, Week 6, Week 12, Week 18 and Week 24) was performed on all three Diary Mini-Questionnaires, Activity (Learning & Leisure) and Affective Experiences (Positive & Negative). All scale items showed good internal reliability with Cronbach’s alpha coefficient, with an average ranging from .64 to .87.

5.5.2.1. Activities

A significant difference was found on the aggregated Learning Activity scale \( F(3.27, 150.19) = 15.71, p < .001, \eta^2 = .26 \), and Leisure Activity Scale \( F(2.87, 131.92) = 4.96, p < .01, \eta^2 = .10 \). Post hoc tests revealed significant decrease in Learning Activity \( p < .001 \) between Week 1-12, Week 1-18, Week 1-24, and Week 6-18, with the steepest drop occurring after 3 months of usage. Post hoc results for Leisure Activity revealed a weaker significant decrease between Week 1-18 \( p < .05 \), Week 6-12 \( p < .01 \), and Week 18-24 \( p < .05 \), which occurred in the first 4 months, and then increased slightly thereafter. Overall participants reported using their iPads more frequency for Learning Activities \( M=3.2 \), then Leisure Activities \( M=3.4 \), as shown in the timeline graph shown in Figure 58.
All the individual items within the Learning Activity scale showed a significant difference ($p < .001$ to $p < .05$), apart from Reading, with the most significant decreases in activity between weeks 1-18, occurring in the first 3 to 4 months of usage. Only two Leisure Activity items showed significant differences ($p < .001$ to $p < .01$) over time, with an increase for Videos (from week 18-24), and a decrease for Photos (from week 1-18 and 6-18).

5.5.2.2. Affective Experience

The affective ratings were aggregated into two groups: Positive Affect (Enjoyable and Engaging), Negative Affect (Frustrating and Boring), along with the single global item Overall Satisfaction. All affect ratings were captured on a 1-7 Likert scale (1=low, 7=high) in order to gain greater granularity of their experience.

A significant difference was found on the aggregated Negative Affect scale; ($F(1.9, 88.86) = 5.68$, $p < .01$, $\eta^2 = .11$). The post hoc tests showed a significant decrease ($p < .05$) for Negative Affect between week 1 and 24. No significant difference was found on the aggregated Positive Affect and Overall Satisfaction, with both ratings remaining consistently high over the 6-month time frame, see Figure 59.
As reported in the main questionnaire results, a significant difference was found on the single item Frustration: $F(2.38, 109.53) = 6.54, p < .001, \eta^2 = .12$, and post hoc tests showed a significant decrease of Frustration ($p < .01$) from week 1-24, and also ($p < .05$) from week 6-24, see Figure 59.

5.5.2.3. Gender, Location and Pre-Owned Checks

A repeated measure ANOVA was conducted on both the Main Questionnaires and the Diary Mini-Questionnaires using Gender and Hospital Groups (four different hospital locations) as a covariant (as a between-subject factor), and no significant differences were found.

The aggregated mean scores from 5 main questionnaire data results (Affect, Hedonics (HQ), Pragmatics (PQ), Usefulness, and Satisfaction) from the 7 participants, who owned an iPad prior to the study were compared with the remaining participant non-owners, and no significant difference was found between the groups. In summary, hospital locations, pre-owned iPad users and gender had little impact on the results.

5.5.2.4. Summary of Diary Mini-Prompts

All activities declined over time, with a slight mid-study dip after 3 months (week 18), indicating either a familiarity or novelty effect. Checking emails, browsing the Internet and social networking were the most frequent activities. The most significant increase in activities (which occurred after week 18) was watching videos and making notes, while the most significant decline was taking photos, (which could be attributed to the iPad’s poor camera quality, as reported in the qualitative findings, see Chapter 6).
Participants gave consistently high positive affective ratings, indicating they found the iPad enjoyable, exciting and satisfying throughout the study. Negative affect ratings showed a gradual decline over time, where frustration was the main contributor, suggesting that initial irritations or annoyances were either overcome or forgiven.

5.5.3. **Verbal Interview Ratings**

The verbal interview ratings were aggregated into two groups: Positive Affect (Enjoyable and Engaging), Negative Affect (Frustrating and Boring), along with the single global items Overall Satisfaction, Attractive and Ease-of-Use. A one-way repeated ANOVA (1x4) was performed on the aggregated groups positive and negative affect, across the four different time-scale ratings: (Expected, Early, Mid and Final-Interview), and no significant differences were found.

5.5.3.1. **Ease-of-Use, Attractive and Satisfaction**

The three single global measures showed significant differences: Ease-of-Use; $F(2, 92.8) = 16.85 \ p < .001, \eta^2 = .26$; Attractive: $F(3, 138) = 5.71 \ p < .001, \eta^2 = .11$, and Overall Satisfaction, $F(3, 138) = 3.25 \ p < .05, \eta^2 = .07$. Post hoc test using Bonferroni correction showed significant increase ($p < .001$) for Ease-of-Use from Expected (M=5.4) to Early (M=6.1), Mid-Study (M=6.2), and Final-Study (M=6.3). A significant increase was found between Expected-Use and Early-Study for Overall Satisfaction: ($p < .05$, M=5.8 to 6.3), and a significant decrease ($p < .01$) for Attractive from Expected (M=6.4) to Final-Study (M=5.9), see Figure 60, (indicating a possible early brand effect).

![Figure 60. Mean ratings for the 3 significant global Verbal Interview Scale items (dotted line= single items)](image-url)
5.5.3.2. Summary of Verbal Interview Items

Participants found the iPad more easier to use and were more satisfied than they had expected, with ease of use increasing the most over time, which concurs with the main PQ questionnaire results (see 5.5.1.2), indicating a possible learning effect. However attractive verbal ratings (5.5.3.1) decreased over time (indicating a possible high brand expectation effect), which contrasts to the increasing beauty ratings reported in the main questionnaires (5.5.1.2), suggesting that the participants judged the quality attribute attractive differently from beauty.

5.5.4. Inter-variable Relationships (Regressions)

In the previous two chapters (3 & 4) Affect and Hedonic ratings increased after interaction, providing a strong indicator of UX. These datasets were further investigated for inter-variable relationships, with both studies showing Hedonics to be a consistent predictor of Goodness and Beauty; however the Pragmatics-Goodness relationship was less consistent (post interaction); and Affect and Hedonics were the main predictors of Overall Experience. The same investigation is conducted using the longitudinal data from Study Three, to explore the influence of participants’ affective emotional responses and quality ratings on their overall experience over time. A Multiple Regression was conducted to test the hypothesis:

Product (iPad) quality (HQ & PQ) and Affect can influence users’ evaluations of Goodness, Beauty, Overall Experience and Usefulness

A Multiple Regression analysis was used to investigate the relative influence of three independent (predictor) variables on four dependent variables across all three time frames. Predictors used were: Affect, Hedonics (HQ), and Pragmatics (PQ), on the Dependent Variables (1) Beauty, and (2) Goodness, (3) Overall Experience and (4) Usefulness. A further regression analysis was conducted to explore the influence of PQ and HQ (independent variables) on Affect (dependant). These five tests were performed across the three-time frames (Early-Mid-Final) to explore any changes over time.

Preliminary analysis was conducted to ensure there was no violation of the assumptions of normality, linearity and multicollinearity, and all values were within recommended tolerance (> 10: 0.482 to 0.799), and VIF (< 10: 1.25 to 2.07), (Tabachnick & Fidell, 2007). Results for all three time frames are shown in Table 42.
### Table 42. Multiple Regression results for 5 Dependant Variables analysis across three time frames

**Hedonics** showed to be the main predictor of **Affect**, which strengthened over time (as indicated by the R-square value: Early-Study: 30% to Final-Study: 42% of variance). **Affect** was a strong predictor for both **Goodness** and **Overall Experience**, which strengthening over time for **Goodness** (Early-Study: 42% to Final-Study: 58%), but remained constant for **Overall Experience** (42%). **Affect** also became the main predictor for **Usefulness** (Final-Study only), and **Beauty** over time, whilst **PQ** initially was found the main predictor for **Beauty** (Early-Study), but was later replaced by **Affect**.

Following the same procedure as adopted in the previous two studies (Chapter 3 & 4), these findings are compared with the inference model (Hassenzahl & Monk, 2010; Hassenzahl, 2004) to assess how judgement of product qualities might be inferred from general perceptions of **Goodness** and **Beauty**. Further regressions were run to investigate the inference model hypothesis, using **Goodness** and **Beauty** as predictors for **HQ** and **PQ**.

The results are slightly less consistent than Study 1 & 2, with **Goodness** showing be the main consistent for **HQ**, and **Beauty** only showing an influence in Final-Study, see Table 43).

Initially **Beauty** was the main predictor for **PQ**, but this changed over time, as **Goodness** was found to be the main predictor, although the **PQ-Goodness** link association is still weak (*p<.05*), as reflected by the main hypothesis. The **PQ** tests also showed poor results, with a low R² accounting for less than 32% of variance, as reflected in the global construct values.
### Table 43. Summary of the general to specific (inference) model regression analysis results for 3 sites

The models generated from the two regression tests are summarised in Figure 61, and show the changes in relationships over time. The specific-to-general model clearly shows **Hedonics** was the main predictor of **Affect**, suggesting **Hedonic** perceptions were mediated by **Affective** responses. Over time, **Affect** became the main predictor of all four product-quality variables (**Goodness, Beauty, Useful** and **Overall Experience**), suggesting that **Affect** became more influential as time progressed. **Hedonics** is more likely to be device-influenced (early-use), whereas over time **Affect** dominated, possibly due to the combination of app/device experiences, which may also be reflected in the **Beauty-HQ** (final-study) relationship. Early perceptions of **Beauty** were influenced by **PQ** ratings (usability), when participants were more focused on device use, as confirmed in the general-to-specific model. Interestingly, **PQ** had little influence in spite of adverse comments, so may have been discounted. **Affect** was the main predictor for **Overall Experience**, as reflected in the previous two experimental studies (reported in Chapter 3 & 4). **Goodness** showed a strong relationship with **HQ** in the general to specific, as reflected by the **Affect-Goodness** link in the specific to general, further indicating that **Affect** was being mediated by **Hedonics**, thus supporting to hypothesis.

### Specific to General

**Hedonics** > **Affect** > **Beauty** > **Useful** > **Overall Experience**

**PQ**

### General to Specific (Inference)

**Goodness** > **HQ** > **PQ**

**Beauty** > **HQ** > **PQ**

**Figure 61. Summary of the regression models: Specific to General (top), & General to Specific (inference) model (bottom) perspectives, showing changes over time.**
5.5.5. **Overall Summary**

A summary of the three questionnaire results (main, diary, & verbal) is shown in Table 44.

<table>
<thead>
<tr>
<th>Scale Items</th>
<th>1-3 months</th>
<th>4-6 months</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Questionnaires</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Affect*</td>
<td>↓</td>
<td>~</td>
</tr>
<tr>
<td>Frustrating (Single Item)*</td>
<td>↓</td>
<td>~</td>
</tr>
<tr>
<td>Pragmatic Quality (PQ)**</td>
<td>~↑</td>
<td>~↑</td>
</tr>
<tr>
<td>Beauty-Ugly*</td>
<td>~</td>
<td>↑</td>
</tr>
<tr>
<td>Hedonics (HQ)**</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Usefulness*</td>
<td>~↑</td>
<td>~↑</td>
</tr>
<tr>
<td>Motivation (Single Item)*</td>
<td>~↑</td>
<td>~↑</td>
</tr>
<tr>
<td><strong>Diary Mini-Questionnaires</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning Activity***</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Leisure Activity**</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Negative Affect***</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td>Frustrating (Single Item)***</td>
<td>↓</td>
<td>↓</td>
</tr>
<tr>
<td><strong>Verbal Interview Items</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease-of-Use***</td>
<td>~↑(P)</td>
<td>~↑</td>
</tr>
<tr>
<td>Attractive***</td>
<td>~↓(P)</td>
<td>~↓</td>
</tr>
<tr>
<td>Overall Satisfaction*</td>
<td>~↑(P)</td>
<td>~</td>
</tr>
</tbody>
</table>

*Key: (P)=Pre-Use, ↑=Non Sig ↑=p < 0.05, ~↑= p < 0.01 to .001, ~=No change
(Arrows show either decrease or increase over time)*

**Table 44. Summary of the sig. results across all scales showing change over time**

The iPad was initially used more for **Learning** than **Leisure**, with most **Activities** decreasing over time, especially within the first 3 months of usage. Checking emails, browsing the Internet and social networking were the most common activities, whilst playing games, listening to music and taking photos were the least common. Reading remained consistent and watching videos increased in the last few months suggesting the iPad was being used as dual-device, supporting both learning and leisure use.

**Positive Affect** showed a mid-study dip but remained high, with **Motivation** ratings increasing over time, indicating participants felt happy using the iPad. **Negative Affect** showed a gradual decline, with **Frustration** ratings decreasing in the first 3 months of usage, as they were either overcome or just accepted. **PQ** and **Usefulness** gradually increased over time as participants found their iPads increasingly **Easier-to-use**, and therefore more useful. **Hedonics** showed a significant dip in the first 3 months of usage, possibly due to a familiarity effect. Although participants rated the iPad more **Beautiful** in the last 3 months of usage (main questionnaire), verbal ratings for **Attractive** showed a gradual decrease over time, suggesting attractiveness maybe be judged differently to beauty, and could be subject to a brand effect. Participants verbally rated the iPad **Easier-to-use** and were generally more **Satisfied** than they had expected prior to use, and this remained unchanged over time, apart from a gradual increase in ease-of-use ratings, as users became more familiar with using the device (as reflected in the PQ ratings).
Inter-variable relationships showed that over time Affect was the main predictor for all variables, where Hedonics was the main mediator for Affect. Initially PQ was the main predictor of Beauty in both specific-to-general and general-to-specific models, as participants were more device-focused, but over time PQ showed little influence. Affect was the main predictor for Overall Experience (as reflected Chapter 3 & 4 results), which strengthened over time, possible due to the app/device experiences. The consistent Goodness-Hedonic link in the general-to-specific model partly supports the inference perspective with the Beauty-Hedonic being less consistent.

5.6. User Characteristics

Further analysis was conducted to investigate if user predispositions towards IT influenced the way the participants rated the iPad's quality. Data derived from the 10-item Power Usage scale (adapted from Study Two) was employed using exploratory Cluster Analysis to identify different user groups by analysing participants' attitudes and motivation, efficacy and expertise towards their use of technology (Marathe et al., 2007), (see Chapter 4, Table 29 for measurement scale).

The questionnaire scale showed very good internal reliability using Cronbach's alpha coefficient reported as .94. All items (apart from two) correlated using Pearson's correlation analysis and no issue of collinearity among the items was found (highly correlated items: > 0.90), indicating that all items make a equal valid contribution.

5.6.1. Cluster Analysis

The same 2-step cluster analysis method used in Chapter 4 was applied (see 4.4.4). The agglomeration schedule generated from the first Hierarchical Cluster Analysis showed a distinct break between the coefficients just before cluster 4, suggesting that the optimal number of clusters should be 3. A K-means cluster analysis was performed based on the 3 cluster groups determined by the hierarchal analysis. Instead of allowing SPSS to choose the cluster centres, the same cluster centroids (which is the mean values of each cluster) found within the hierarchical analysis were used.

A comparison between the Hierarchal (step 1) and K-Means (Step 2) cluster groups showed little change between the two methods, with only two items (participants) being assigned to a different cluster. The K-Means ANOVA indicate the three power user clusters were significantly different \((p>.001)\) and split between High \((N=21)\), Medium \((N=17)\) and Low \((N=9)\) Power Users, see Table 45 for mean range, standard deviation and cluster centres (mean
value of each cluster) for each Power User group, (with the range only showing a slight overlap). Participants who pre-owned an iPad prior to the study were evenly split between clusters indicating pre-ownership had little impact.

1. **High Power Users**
   - N: 21
   - Mean: 6.1
   - Range: 5.2 – 6.5
   - Std. D: 0.78

2. **Medium Power Users**
   - N: 17
   - Mean: 4.6
   - Range: 3.5 – 5.1
   - Std. D: 1.17

3. **Low Power Users**
   - N: 9
   - Mean: 2.8
   - Range: 1.8 – 4.9
   - Std. D: 1.28

**Table 45. K-Means for the three Power Users clusters**

The three power user groups were then used as a factor to explore how the influence of prior experience and technology disposition impacted on users’ judgement across the various quality scales.

### 5.7. Between Group Differences

A mixed between-group repeated-measures ANOVA 3x3 design with Users as between-group factors (High, Medium, Low), and Time as within-group factors (Early, Mid & Final-Study) was conducted on the aggregated main questionnaires, diary mini-questionnaires and verbal interview items. This was used to explore if there were any significant main effects between users, or interaction (users x time), for each of the quality variable ratings.

#### 5.7.1. Main Questionnaire Results (Between Groups)

Significant results for the main questionnaires between-groups repeated measures ANOVA are now reported. Main Time effects are not reported, as they were disclosed in the one-way ANOVA analysis in Section 5.5, so this section focuses upon between-group differences. When the assumption of sphericity was violated in the ANOVA, the Greenhouse-Geisser correction was used.

**5.7.1.1. Affect**

No significant main effects or interaction was found on the aggregated Positive or Negative Affect scales. The single item Frustration was analysed to see if there were any inter-group differences, and a main effect for users: \( F(2, 44) = 3.51, p < .05, \eta^2 = .14, \text{Wilks' Lambda} = .84 \) was found. Post hoc tests using Bonferroni correction found that the Low Power Users \( (p < .05, \text{M}=3.4) \) were more likely to feel frustrated than High Power Users \( (\text{M}=2.4) \), and both Low and Medium Power Users experienced the greatest decrease in frustration (especially in the first 3 months of usage), see Figure 62.
5.7.1.2. Design Quality

The aggregated **Hedonic Quality (HQ)** subscale revealed a significant main effect between users: \( F(2, 44) = 4.34, p < .05, \eta^2 = .17, \text{Wilks' Lambda} = .79 \), but no significant interaction \((\text{time} \times \text{users})\). No significant post hoc difference were found between users, although mean scores indicate the **High Power Users** gave higher Hedonic ratings than the other two groups, see Figure 62. No significant main effects or interaction were found for **Beauty** or **Goodness**.

![Figures. 62. Mean ratings for single Affect Scale item Frustration (left), and Hedonics (HQ) (right)](image)

The aggregated **Pragmatic Quality (PQ)** scale showed a significant main effect for users: \( F(2, 44) = 8.22, p < .001, \eta^2 = .27, \text{Wilks' Lambda} = .73 \), but no significant interaction. Post hoc tests revealed the **High Power Users**: (M=6.0) gave significantly higher ratings \((p < .05)\) than **Medium Power Users** (M=5.4), and \((p < .01)\) **Low Power Users** (M=5.1), indicating they found the device less confusing and were more tolerant of usability problems than the other two power user groups, see Figure 63. Initially the **Low Power Users** gave the lowest PQ ratings, but over time they showed the greatest increase, suggesting initially both groups found the iPad more complicated, confusing and unpredictable than **High Power Users**, however the gap narrowed over time and prolonged use.
5.7.1.3. Usefulness and Overall Satisfaction

No significant difference between users was found for Usefulness, although a significant interaction (users x time) effect was found: \( F(4, 88) = 3.3, p < .05, \eta^2 = .13 \). Post hoc tests showed the Low Power Users experienced the greatest increase in Usefulness over time, see Figure 63.

5.7.2. Diary Mini-Questionnaires

The Diary Mini-Questionnaire results was analysed across 5 time frames using a mixed between group repeated ANOVA 3x5 design, with user groups (High, Medium, Low) and Time (1 Week, 6 Weeks, 12 Weeks, 18 Weeks and 24 Weeks), to see if there were any inter-group differences. (Main effects for time were reported earlier in Section 5.5.2). No significant differences were found between users or interaction for Learning Activity, Leisure Activity, or for Positive of Negative Affective experience.

5.7.3. Interview Verbal Scale Items

The 7-item verbal quality ratings asked at the start of each interview were analysed across 4 time frames using a mixed between group repeated measures ANOVA design, with 3x Users and 4x time-periods (Expected, Early, Mid, Final), to explore inter-group differences, (time effects were reported earlier in section 5.5.3).

A significant effect between users was found for Ease-of-Use \( F(2, 44) = 14.30, p < .001, \eta^2 = .39, \) Wilks’ Lambda = .57. Post hoc tests revealed the High Power Users gave higher ratings \( p < .001 \) than Low Power Users, and the Medium Poser Users \( p < .01 \), see Figure 64, which is
consistent with the main questionnaire PQ results reported earlier (5.9.1.2). No significant differences between users were found for the Positive Affect (Enjoy & Engagement), Negative Affect (Frustrating or Boring), or the global items Satisfaction or Attractive.

**Figure. 64.** Mean ratings for Interview Verbal Global Scale Item Ease-Of-Use

### 5.7.4. Between Group Summary

A summary of all the between user group analysis results is shown in Table 46. The significant user group differences are ranked in order (highest group ratings) and the significant changes over time are shown across all the quality variables.

<table>
<thead>
<tr>
<th>Scale Items</th>
<th>Time Trends</th>
<th>Rank Order of Power Users</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Questionnaires</strong></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Frustrating*</td>
<td>↓</td>
<td>2</td>
</tr>
<tr>
<td>Hedonics (HQ)*</td>
<td>↓ ↑</td>
<td>1</td>
</tr>
<tr>
<td>Pragmatic Quality***</td>
<td>↑ ↑</td>
<td>1*</td>
</tr>
<tr>
<td>Usefulness*</td>
<td>↑ ↑</td>
<td>2</td>
</tr>
<tr>
<td><strong>Verbal-Interview</strong></td>
<td>↑</td>
<td>1***</td>
</tr>
</tbody>
</table>

Key: Arrows = Time Change: (P)=Pre-Use, ↑=Non Sig ↑=p <.05, ↑= p <.01 to .001, ~=No change
Sig: ***=p <.001, **=p <.01, *=p <.05, **Bold items**=Aggregated Scales, Not Bold= Single Items

**Table. 46.** Summary of the between groups (PU) differences and changes over time

Significant differences were found between the Power User Groups indicating prior experience influences user judgement. High Power Users gave higher ratings (than the other two groups) for Hedonics, Pragmatic Quality and the Ease-of-Use. So the High Power Users rated the iPad more predictable, simpler, clearer and overall more usable, and found the device more appealing, pleasurable and stimulating to use (Hassenzahl, 2004). Pragmatic Quality (PQ) and Usefulness increased over time with Low Power Users benefiting more, as device problems were resolved. Unsurprisingly, the Low Power Users felt more frustrated than the other two groups, and both Low and Medium Power Users experienced more marked
Chapter 5  Study Three: Longitudinal Study (Quantitative)

decreases over time (especially in the first three months of usage), contributing to the overall decrease in frustrating ratings. A mid-point dip in Hedonic ratings for both Low and Medium Power Users may have been when usability was more difficult. Most measures between groups converged over time apart from PQ and Frustrating, which suggests a prior experience effect on UX.

5.7.5. Application Analysis

So far the analysis has focused on the iPad as a combined app and device tool. This section takes a closer look at some of the key applications that were used by the participants in order to begin to unravel the influence individual apps may have had on participant’s quality judgements.

5.7.5.1. Application Audit

Before analysing the application data an audit of the main learning apps was conducted, which were split between the General Apps, Productivity Apps (see Table 47), and special Clinical/Medical (related) Apps (see Table 48).

<table>
<thead>
<tr>
<th>APPLICATION TYPE</th>
<th>Interactive Features</th>
<th>Interactive Rating</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD=Productivity Apps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GN= General Apps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ = Rival Apps</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notability (Note Taker) PD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edit ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Draw ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annotate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Media ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organise ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reader ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calculate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graphs ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Animate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desk-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloud-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive Rating 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost £1.49</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Evernote (Doc Organiser) PD

| Edit ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Draw ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Annotate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Record ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| +Media ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Organise ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Reader ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Search ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Calculate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Graphs ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Animate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Desk-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Cloud-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Interactive Rating 8 |
| Cost Free (Limited) |

GoodReader (PDF Reader) PD

| Edit ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Draw ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Annotate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Record ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| +Media ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Organise ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Reader ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Search ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Calculate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Graphs ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Animate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Desk-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Cloud-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Interactive Rating 8 |
| Cost **Free (£2.99)** |

iBooks (Library Books) GN

| Edit ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Draw ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Annotate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Record ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| +Media ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Organise ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Reader ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Search ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Calculate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Graphs ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Animate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Desk-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Cloud-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Interactive Rating 5 |
| Cost Pre-Loaded |

Numbers (Spreadsheet) PD

| Edit ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Draw ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Annotate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Record ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| +Media ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Organise ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Reader ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Search ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Calculate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Graphs ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Animate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Desk-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Cloud-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Interactive Rating 6 |
| Cost *Free (£6.99)** |

Keynote (Presentations) PD

| Edit ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Draw ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Annotate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Record ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| +Media ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Organise ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Reader ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Search ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Calculate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Graphs ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Animate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Desk-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Cloud-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Interactive Rating 5 |
| Cost **Free (£6.99)** |

DropBox (File Sharing) GN

| Edit ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Draw ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Annotate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Record ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| +Media ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Organise ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Reader ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Search ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Calculate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Graphs ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Animate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Desk-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Cloud-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Interactive Rating 5 |
| Cost Free |

Blackboard (Mobile VLE)*** GN

| Edit ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Draw ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Annotate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Record ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| +Media ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Organise ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Reader ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Search ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Calculate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Graphs ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Animate ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Desk-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Cloud-Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Sync ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ |
| Interactive Rating 5 |
| Cost Free |

* Given to Participants by * Medical School ** Given by Researcher, *** University Virtual Learning Environment

Table 47. Summary of the key applications used prior to conducting the study.

The apps were analysed according to the interactive features they offered, which was based on specific functions that enabled activities such as reading (repository), note taking (annotating, drawing, recording etc.), organising (search), and reporting or presenting (calculating, creating graphs & animating). Thirteen interactive features identified across all apps were used to determine an Interactive Rating for each app, (i.e., the total interactive feature score each app provided). So the productivity apps Notability scored the highest rating, as it offered 10 different interactive features. Rival apps often shared similar features yet differed in interactive ratings, e.g., Notability (10) Evernote (8), and Goodreader (8) iBooks (5).
An Interactive Rating was also calculated for the most commonly used Medical Apps (taken from the Mid-Study audit), as shown in Table 48. Most apps shared similar features (e.g., read, search, website options) with few additions, where the game-based app Prognosis scored the highest Interactive Rating (5), due to its quiz-based design that enabled question and answer feedback.

<table>
<thead>
<tr>
<th>Clinical/Medical Applications</th>
<th>Used For</th>
<th>Connection</th>
<th>Reader</th>
<th>Search</th>
<th>Calculate</th>
<th>Annotate</th>
<th>+Interact</th>
<th>+Share</th>
<th>Website</th>
<th>Interactive Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNF Nice Guidelines</td>
<td>Prescribing &amp; Diagnosis</td>
<td>Off-line</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>BMJ Best Practice</td>
<td>Aid Clinical Decisions</td>
<td>On-line</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Medscape</td>
<td>Medical Reference</td>
<td>Off-line</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Prognosis Diagnosis</td>
<td>Interactive Diagnosis Cases</td>
<td>Both</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Almost a Doctor</td>
<td>Medical Encyclopaedia</td>
<td>On-line</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Patient.co.uk</td>
<td>UK Patient Medical Leaflets</td>
<td>On-line</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Prep4Finals</td>
<td>Multiple Choice Questions</td>
<td>On-line</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Table 48. Summary of the most prominent Clinical/Medical applications

5.7.5.2. Application Usage

At the start of the study participants were asked if they had downloaded and used two apps: Blackboard and Numbers (as these were mandatory, and therefore the most likely two apps students would first use). The Blackboard App was downloaded by 65% (33) and Numbers was downloaded by 86% (44) of the participants. A free Numbers app voucher was given to all students when they received their iPads, as this was a mandatory application (used to record skills and patient logs), and Blackboard VLE (Virtual Learning Environment), also mandatory (used to access learning materials), was pre-installed as a link, (so users still had to download both the Numbers and Blackboard apps in order to use them).

In the Mid-Study participant’s were asked if they had downloaded and used eight of the most popular applications (these were obtained from the qualitative data taken from the Pre-Study and Dairy Mini-Questionnaires). The results showed that 94% had downloaded iBooks, 88% Numbers, 86% Facebook, 80% Notability, 76% Keynote, 73% Blackboard, 65% GoodReader and 57% Evernote. Participants were asked if they still used the same apps in the Final-Study, and the % difference is shown in Figure 65 (in brackets next to each apps). All ratings dropped apart from Facebook (-1%), which remained the same, and Blackboard (+7), which showed a slight increase, with Numbers and Keynote (-18%) having the largest drop in usage.
5.7.5.3. Frequency of Downloads

Participants downloading of new apps decreased over the 6-months time frame (Early: M=3.8 to Final: M=4.8), from once or several times a week, to only once or twice a month, although an increase in standard deviation (SD=1.8) was shown, suggesting a wide variation in download activity. Participant’s estimated the number of apps they had downloaded at the start of the study was between 20-30 applications (M=2.9), which increased to 30-40 apps in 3-months (M=3.4), and showed a slight increase (M=3.6) in 6-months of usage, suggesting app downloads were more frequent in the first 3 months of usage as participants experimented with different apps, after which participants app download activity decreased.

The frequency of how often the 8 apps were used, was captured twice (Mid/Final-Study) over the last 3 months of the study, see Figure 66.

![Figure 65. Key Applications (%) that were downloaded and used (at least once)](image)

![Figure 66. How often the 8 main Applications were used (comparing Mid to Final results)](image)
The most used apps were Facebook, (used daily), then iBooks, Notability and GoodReader, (several times a week), with little change over the 3-month time frame. iBooks was marginally favoured over the rival app GoodReader, although iBook usage dropped slightly over time, while GoodReader increased. Blackboard, Numbers, Evernote and Keynote were used least frequently (once or several times a month). Blackboard usage increased over the 3-months time frame, as reflected by the increase in download and usage +7%, (see Figure 65). The rival app Notability was favoured over Evernote, despite it being more expensive, although there was a slight drop over time, as reflected in the usage results (-12%), possibly indicating a drop in note taking. Numbers was used infrequently and dropped slightly over time (despite it being a mandatory app), with Keynote showed the greatest drop in usage.

Three more apps were rated during Final-Study, (since they became salient in the Mid-Study interviews), with 85% reported to have downloaded YouTube, 81% Dropbox and 55% Skype.

5.7.5.4. Application Ratings

All eleven apps were then rated on 6 different quality items: functionality (utility), ease of use (usability), information quality (content), effort to learn, importance and overall experience.

Two mandatory apps (Numbers and Blackboard) were rated across three time frames (Early, Mid and Final-Study), as shown in Table 49. A one-way repeated ANOVA was conducted to see if there were any changes in the ratings over time. Blackboard was the only app to show significant differences (p < .05 - p < .01) for Utility, Usability, Importance and Overall Experience, with post hoc tests for Usability showing the greatest increase, especially in the first 3 months of usage (Early: M=4.4 to Mid; M=5.1 to Final: M=5.6).

<table>
<thead>
<tr>
<th></th>
<th>Utility</th>
<th>Usability</th>
<th>Content</th>
<th>Effort to Learn</th>
<th>Importance</th>
<th>Overall Experience</th>
<th>Total Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-2-3</td>
<td>1-2-3</td>
<td>1-2-3</td>
<td>1-2-3</td>
<td>1-2-3</td>
<td>1-2-3</td>
<td>1-2-3</td>
</tr>
<tr>
<td>Numbers</td>
<td>5.3 5.6 4.9</td>
<td>5.3 5.3 4.4</td>
<td>6.5 6.3 5.6</td>
<td>5.1 5.1 5.1</td>
<td>5.1 5.1 5.1</td>
<td>6.5 6.5 6.5</td>
<td>5.3 5.3 5.3</td>
</tr>
<tr>
<td>Blackboard</td>
<td>4.1 4.5 5.0*</td>
<td>4.4 5.1 5.6**</td>
<td>6.5 6.3 5.6</td>
<td>5.1 5.1 5.1</td>
<td>5.1 5.1 5.1</td>
<td>6.5 6.5 6.5</td>
<td>4.1 4.5 5.0</td>
</tr>
<tr>
<td>iBooks</td>
<td>- 6.0 6.0</td>
<td>- 6.2 6.3</td>
<td>- 5.6 5.7</td>
<td>- 6.0 6.3</td>
<td>- 5.7 5.7</td>
<td>- 6.1 6.3</td>
<td>- 5.9 6.1</td>
</tr>
<tr>
<td>Notability</td>
<td>- 6.1 6.2</td>
<td>- 6.1 6.2</td>
<td>- 5.7 5.6</td>
<td>- 5.2 5.5</td>
<td>- 5.8 5.8</td>
<td>- 6.2 6.3</td>
<td>- 5.8 5.9</td>
</tr>
<tr>
<td>Facebook</td>
<td>- 5.8 5.7</td>
<td>- 6.5 6.3</td>
<td>- 5.2 5.2</td>
<td>- 6.3 6.2</td>
<td>- 4.5 4.7</td>
<td>- 6.1 6.0</td>
<td>- 5.7 5.7</td>
</tr>
<tr>
<td>GoodReader</td>
<td>- 5.5 5.6</td>
<td>- 5.3 5.2</td>
<td>- 5.2 5.3</td>
<td>- 4.8 4.6</td>
<td>- 4.8 4.6</td>
<td>- 5.5 5.6</td>
<td>- 5.2 5.2</td>
</tr>
<tr>
<td>Keynote</td>
<td>- 5.1 5.3</td>
<td>- 5.1 5.3</td>
<td>- 4.8 5.0</td>
<td>- 4.6 4.8</td>
<td>- 4.0 4.4</td>
<td>- 5.0 5.3</td>
<td>- 4.8 5.0</td>
</tr>
<tr>
<td>Evernote</td>
<td>- 5.0 4.9</td>
<td>- 5.1 5.2</td>
<td>- 4.9 5.1</td>
<td>- 4.6 5.0</td>
<td>- 4.3 4.4</td>
<td>- 5.1 5.0</td>
<td>- 4.8 4.9</td>
</tr>
<tr>
<td>Dropbox</td>
<td>- - 5.0</td>
<td>- - 5.2</td>
<td>- - 5.5</td>
<td>- - 5.6</td>
<td>- - 5.8</td>
<td>- - 5.8</td>
<td>- - 5.5</td>
</tr>
<tr>
<td>Skype</td>
<td>- - 5.0</td>
<td>- - 5.2</td>
<td>- - 5.5</td>
<td>- - 5.6</td>
<td>- - 5.8</td>
<td>- - 5.8</td>
<td>- - 5.5</td>
</tr>
<tr>
<td>YouTube</td>
<td>- - 4.4</td>
<td>- - 5.4</td>
<td>- - 5.7</td>
<td>- - 5.8</td>
<td>- - 6.1</td>
<td>- - 6.1</td>
<td>- - 5.6</td>
</tr>
</tbody>
</table>

1 = Early-Study  2 = Mid-Study  3 = Final-Study

Results are based on a Likert Scale 1-7 (1=Very Bad, 2=Bad, 3=Neutral, 4=Poor, 5=Fair, 6=Good, 7=Very Good)

Table. 49. Means for 6 Quality Ratings for 11 key applications
A further six apps were rated twice (Mid and Final Study), see Table 49, where mean scores indicate that the most favoured apps were iBooks, Notability and Facebook scoring high on all items, (although Facebook was rated less for content and importance). GoodReader came fourth and scored low for effort to learn and overall experience. The lowest rated, and least used apps were Numbers, Blackboard, Keynote and Evernote.

A Wilcoxon Matched-Pairs test showed no significant difference between the two time frame ratings for any of the apps. Mean ratings show a slight increase for all apps from Mid to Final Study (apart from Facebook and GoodReader), possibly because as participants became more familiar with them, their ratings improved, see Figure 67. Interestingly the mandatory app Blackboard showed the greatest increase (+5.2%).

An additional three apps, (YouTube, Skype and Dropbox) were rated once at the end of the study (Final Study), scoring higher then some of the work based apps, due to high ratings for overall experience, effort to learn and importance (see Table 49), indicating a mix of work/leisure use.

All work-based apps (W) received lower ratings apart from Notability, which was the favoured note-taking app and rated higher than Evernote (its rival) for utility, usability and overall experience. Dual purpose apps (DP) such as iBooks, YouTube and Dropbox received high ratings, as they enabled both work and leisure use; while Facebook scoring high as it offered a specific purpose (S) for social-networking, or Skype for its free communication function. Dual-purpose apps enabled work and leisure use, for example, Dropbox was used 20% for social use; Facebook 80%, and Skype 97%. Dual-usage of apps could account for why some apps such as iBooks and YouTube were highly rated.
**Specialty Medical Applications**

Participants were asked *Mid-Study* to state their five most used medical-related applications. A total of 44 applications were reported, of which the top 15 apps and percentage popularity is shown in Figure 68. Six medical apps were rated more popular than *iBooks* and *Notability* (after 3 months of usage), suggesting the iPad was being used primarily as a clinical support tool.

The same 6 most popular medical apps (as shown in blue in Figure 68), were rated 3 months later (*Final-Study*), along with the newly developed app from Manchester Medical School, called *Manchester eForms (UoM eForms)*, which enabled the students to submit assessment forms directly from their iPad. Quality ratings were the same as the previous apps (*most used, utility, usability content, effort to learn, importance and overall experience*).

![Bar Chart: The 15 most Popular Medical Apps (out of 44) after 3 months (Mid-Study) (The Top 6 apps - shown in blue - were rated in the Final Study – see Fig11).](image)

How often the 7 apps were used is shown in Figure 69, with *BMJ Best Practice (British Medical Journal)*, *Medscape* and *Almost a Doctor* being used the most often (from weekly to a few times a week). Although *BNF Nice Guidelines* was the most popular app *Mid-Study* (ranked 1), it was one of the least used apps in Final-Study (ranked 5), along with *Prognosis Diagnosis* and *UoM eForms* (reported to be used either monthly or seldom).
The quality ratings for each of the 7 medical apps are shown in Table 50. A Pearson’s correlation between the frequency of use (shown in Fig 67) and quality ratings for BMJ, BNF, Medscape and Patient.co were significant: \((p<.001\) to \(.05\)), see Table 50. The most frequency used app (BMJ and Medscape) scored the highest mean ratings, and were most used, although Medscape was only ranked 5 in popularity Mid-Study. Both apps scored high on content, as they provided instant access to a comprehensive range of medical information from either a trustworthy site (Medscape), or highly reputed authority (BMJ). The middle scoring apps (Almost-a-doc, Patient.co and Prognosis) scored well on all items apart from importance, and were popular apps used as additional reference support. Prognosis scored slightly less (for content and importance), and was not as popular, as this was a more specialist app, providing a game-based tool for revision. The worst rated app (ranked 6 and least used) was BNF scoring low on usability and effort to learn, despite being ranked 1 for popularity and scoring well on content and importance; while UoM eForms was the least used app and scored low on all items indicating it was not a well liked application.

<table>
<thead>
<tr>
<th></th>
<th>UTILITY</th>
<th>USABILITY</th>
<th>CONTENT</th>
<th>EFFORT TO LEARN</th>
<th>IMPORTANCE</th>
<th>OVERALL EXPERIENCE</th>
<th>TOTAL MEAN</th>
<th>Pearson’s r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medscape</td>
<td>6.0</td>
<td>5.9</td>
<td>6.1</td>
<td>5.8</td>
<td>5.6</td>
<td>5.9</td>
<td>5.9</td>
<td>-.416*</td>
</tr>
<tr>
<td>BMJ Best Practice</td>
<td>5.9</td>
<td>5.9</td>
<td>6.1</td>
<td>5.7</td>
<td>5.8</td>
<td>5.9</td>
<td>5.9</td>
<td>-.707***</td>
</tr>
<tr>
<td>Almost-a-Doctor</td>
<td>5.8</td>
<td>5.8</td>
<td>5.7</td>
<td>6.2</td>
<td>5.4</td>
<td>5.8</td>
<td>5.8</td>
<td>-.089</td>
</tr>
<tr>
<td>Patient.co.uk</td>
<td>5.7</td>
<td>5.8</td>
<td>5.7</td>
<td>6.1</td>
<td>5.4</td>
<td>5.7</td>
<td>5.8</td>
<td>-.489**</td>
</tr>
<tr>
<td>Prognosis Diagnosis</td>
<td>5.6</td>
<td>6.1</td>
<td>5.5</td>
<td>6.1</td>
<td>5.0</td>
<td>6.0</td>
<td>5.7</td>
<td>-.230</td>
</tr>
<tr>
<td>BNF Nice Guidelines</td>
<td>5.3</td>
<td>4.9</td>
<td>5.9</td>
<td>5.1</td>
<td>5.7</td>
<td>5.1</td>
<td>5.3</td>
<td>-.348*</td>
</tr>
<tr>
<td>UoM eForms</td>
<td>4.5</td>
<td>4.7</td>
<td>4.3</td>
<td>4.6</td>
<td>4.8</td>
<td>4.7</td>
<td>4.6</td>
<td>-.201</td>
</tr>
</tbody>
</table>

Results are based on a Likert Scale 1-7 (1=Very Bad, 2=Bad, 3=Poor, 4=Neutral, 5=Fair, 6=Good, 7=Very Good)

\(***=p<.001\), \(**=p<.01\), \(*=p<.05\)

Table 50. Mean ratings for 7 most used Medical Apps after 5 months of usage

5.7.5.5. App Memory Recall

During the Mid and Final-Study interviews participants were asked to recall (from memory) their top 3 apps they had used the most since owning their iPad. A total of 75 different apps
were recalled during the *Mid-Study* interview, and a total of 63 apps from the *Final-Study* interview (with participants recalling between 1-2 apps). The most popular 11 apps (apart from email, safari and calendar), that were mentioned at least 5 times or more (in either *Mid* or *Final Study*) is shown in Figure 70.

![Figure 70. The 11 most recalled apps during the Mid & Final Interviews](Image)

The most recalled apps include a mix of social, work and medical-based apps, with *Notability* being recalled the most after 3 months (*Mid-Study*), but showed the greatest drop after 6 months (*Final-Study*). Although *Notability* was rated high it became less salient in participants’ minds, reflecting a drop in note-taking activity, whereas the social networking app *Facebook* was recalled more frequently over time. Interestingly, the video player apps *iPlayer* and *YouTube* showed the greatest increase in memory recall, suggesting the iPad was being used for more leisure based activities over time. Furthermore, the most recalled apps were all self-selected, while mandatory apps (*Blackboard* and *Numbers*) were in the tail of the poorly recalled apps, probably due to their poor usability and effort to learn.

### 5.7.5.6. Summary of Application Findings

The most popular apps are listed in Table 51 according to the highest % frequency mentioned, (where general apps and medical apps were collected at *Pre, Early, and Mid-Study*). These are compared with mean ratings of how often the apps were used, the quality mean ratings, and the most recalled of all apps from memory (collected at the final-interview). An overall mean score from these three ratings was calculated for each app (see Table 51).
memory (recall) as the other top apps (e.g., frequently Dual based), which may not have appealed to all.

Prognosis favour of usage ratings, popular due to their brand name medical apps (quality rating The interactive features out of all the apps Notability its unique functionality and ability to support learning the highest scoring of all apps (most used apps, and were consistently in the top positions (Total Mean=2/3). Notability was the highest scoring of all apps (Mean=2.3), and the only high scoring work-app, rated high for its unique functionality and ability to support learning (e.g., note-taking). Interestingly, Notability scored the highest Interactive Rating (see Table 47), as it offered the most interactive features out of all the general apps.

The medical apps (BMJ, Medscape) scored the next highest (Total Means= 4/6), receiving high quality rating, and being recalled the most, although they were not used as often. Other medical apps (Medscape, Almost-a-Doctor) scored well on all measures (Mean=6/8), and were popular due to their brand name. Interestingly BNF Nice Guidelines was rated the most popular app (Pre-Study), and was the third most recalled app, but received low quality and usage ratings, suggesting that poor quality (e.g., usability, functionality) were forgiven in favour of the apps trusted content and brand name. Less popular medical apps (such as Prognosis) were least used and recalled, probably due to their unique functionality (game-based), which may not have appealed to all.

Dual-purpose (DP) apps (work or leisure), (e.g., iBooks or YouTube), were highly popular, frequently used and received good quality ratings, despite being not as salient in participants memory (recall) as the other top apps (e.g., Notability, Facebook). While the popular specific
purpose (S) app Facebook was used most often and recalled the most, it received a lower quality rating, suggesting participants forgave poor quality in favour of the specific function it provided. The popular dual-purpose app Dropbox (file exchange), receiving average ratings, as it had no other strong rivals. In functional equivalent apps (Evernote/Keynote), the more popular apps (Notability/Prezi), scored higher for most measures indicating they were the preferred choice. Notability scored a higher interactive rating (10), compared to its rival Evernote (7), which could have influenced user choice. Similarly, iBooks was the highest scoring app, despite the low interactive score (5), compared to its rival GoodReader (8), due to its high quality rating.

Mandatory work apps (such as Numbers, UoM Forms) scored low on all ratings, where poor quality ratings (usability, effort to learn) may have impacted on overall usage. Although Blackboard (also a mandatory app) showed an increase in usage and quality ratings over time, and was used fairly frequently, but remains unclear if the ratings were for the app or the website (as some participants reported accessing Blackboard through Safari).

5.8. Chapter Summary

An overview of the key findings from this chapter is shown in Table 52, which compares the three questionnaire results (Main, Diary, Verbal-Interview) used in the longitudinal study, and the main power user group differences.
The results show a consistent picture across the three measures, apart from two differences. No affect differences were found within the verbal questionnaires, which could be due to the scaled down version of the scale. However, the diary mini-questionnaire captured negative affect and frustrating on a similar scale, indicating a difference in verbal to questionnaire type ratings. The second difference was in overall satisfaction, which was captured by the interview-verbal items, and not through the paper/online questionnaires, as participants are maybe more likely to give more positive verbal ratings during a face-to-face interview to please the interviewer, where the presence of the interviewer may have a positive influence on the user judgement ratings.

The results across all three questionnaires show significant differences over time and between user groups, as now summarised:

*Activities:* Overall all iPad activities declined over time, (apart from reading and watching videos), with the most prominent activities (checking emails, browsing the Internet and social networking) taking place across different devices (iPad, computer and mobiles) creating competition between devices.

*Affect:* Participants generally felt happy using the iPad, with positive affect (fun and enjoyment) ratings remaining high, despite a mid-study dip, while the negative affective ratings dropped over time, as the problems causing participants’ initial frustrations were either accepted, overcome, or just avoided.

*Usability and Utility:* PQ and Usefulness ratings increased over time as participants found their iPads increasingly easier to use, so therefore more useful.

*Hedonics:* HQ ratings showed a drop in the first 3 months, indicating a familiarity effect, yet participants rated the iPad more beautiful in the last 3 months of usage, suggesting a possible halo effect with PQ. However, verbal ratings for attractive showed a gradual decline over time, suggesting attractiveness maybe judged differently to beauty, which maybe subject to a brand effect.

*Overall Experience:* Participants did not expect the iPad to be as easy to use prior to receiving it, and so they felt more satisfied in the first few weeks of ownership. Motivation to use the device increased over time, in line with the increase in ease-of-use and usefulness.

*User Differences (Prior Experience):* High Power Users (PU) who have a stronger positive predisposition towards technology (Marathe et al., 2007), gave higher ratings (than Med or Low PU’s) for hedonics, pragmatic quality, and ease-of-use. So they rated the iPad more usable, predictable, simpler, clearer, and generally more easier-to-use;
Furthermore, they found the device more appealing (aesthetically) and pleasurable. Both High and Medium PU’s gave satisfactory hedonic ratings although a mid-point dip may have been when usability or learning was more difficult. PQ ratings increased over time with Low PU’s benefiting more, as device problems were resolved, whereas High PU’s gave higher PQ ratings (than the other two groups), as they experienced fewer problems. Frustation ratings decreased over time for all groups, especially in the first three months, with the High and Medium PU’s experiencing more marked decreases. Low PU’s found the iPad more useful than the other two groups, with all three groups ratings converging over time.

Influences on Judgement: Regression analysis showed that Hedonics was the main predictor for Positive Affect, and over time participants Affective responses increased, strongly influencing overall quality ratings. Early perceptions of Beauty were influenced by PQ ratings, when participants were more focused on device use (familiarisation), as time progressed this was replaced with Positive Affect, possibly due to the influence of app-based judgements, where device usability (PQ), was adequate. In summary, Hedonic perceptions were a precursor to Affective responses, which increased overtime, positively influencing product quality judgements.

Apps: Successful apps were rated high for their unique functionality (e.g., enhanced interactive features) and ease of use. They provided either dual-use (work or leisure), or provided a specific function, such as social networking or communicating. Speciality medical apps were chosen for their trusted content and ease of access, where usability problems were often forgiven. Unsuccessful apps (often mandatory) scored low for usability and effort to learn, so were less favoured over similar rival apps.

In summary, the main changes over time was an increase in PQ where prior experience mattered, as the more technically disposed users (High PU) gave consistently higher ratings, were less susceptible to usability problems and rated the device more appealing and pleasurable. Initial frustrations for all groups decreased over time as users either resolved or tolerated usability problems. Despite a drop in activity, usefulness ratings increased as over time, as ease-of-use improved, which enhanced motivation, (possibly due to the productivity benefits from the apps). There was little change in user satisfaction (apart from expected use), as all participants felt good about using the device giving high hedonic and attractiveness verbal ratings, although these declined over time, possibly because of a familiarity effect. In contrast, beauty ratings increased after 3 months of usage, suggesting a possible halo effect with PQ.
This chapter gives a consistent picture, however questionnaire measures provide holistic ratings that aggregate both the device and apps, so it remains unclear how much the apps alone contribute to the overall ratings. The app analysis provided some insights into the reasons behind the most favoured apps, and revealed how usability and functionality can impede or improve UX. The most successful apps showed consistently high ratings over time, indicating they contributed to participants' motivation, usefulness and overall satisfaction with the iPad.

The next chapter aims to separate the experiences related to apps from device, and will use the qualitative data from interviews and diary studies to answer questions that have arisen from this chapter:

- The apps showed considerable differences, but did only a few apps influence the positive ratings, and were poor apps just ignored or tolerated?
- How does the increasing usefulness (utility) ratings link to the apps, and/or the device?
- PQ clearly improved over time, yet what evidence is there about why this occurred?
- What reasons were there for the High/Low Power User's ratings?
- Do participants' judgements change (instant or reflective) depending on the method employed?

The final question refers to the mixture of quantitative methods used within in this chapter, where the main questionnaire findings give a much clearer picture, compared to the weekly mini-questionnaires and verbal interview items which provided less defined results. Chapter 6 will use the qualitative data to provide answers to these questions in order to extend the findings gained from this chapter.
Chapter 6  Study Three: Longitudinal Study (Qualitative)

This chapter aims to expand the findings reported from the last chapter (Chapter 5), through the analysis of the qualitative data captured during the longitudinal study. First the qualitative analysis method will be outlined, followed by the main findings that combine both the interviews and open questionnaire diary findings. The chapter then presents a non-adopter follow-up study, and then concludes with an overall discussion of the qualitative results.
6.1. Introduction

The quantitative analysis in Chapter 5 found that pragmatics and utility improved over time and usage, while hedonics showed a mid-study dip with overall ratings remaining high showing little change over long-term use. The quantitative results captured participants’ composite ratings that combined both device and app. However the app analysis revealed that successful apps were due to their unique functionality, that either enabled dual-purpose activities (work/leisure), or provided a specific purpose (communicating), with specialist medical apps being favored for their trusted content. However, how much influence individual apps have upon UX still remains unclear. The qualitative data captured from participant interviews and weekly self-reporting diary responses will be used to further understand the factors that contributed or hindered user satisfaction. Building from Chapter 5, this chapter aims to tease apart the complexity of apps from device, and determine which apps provided positive UX and why, and how much they may impact on usability, utility and usefulness. Further insights into product adoption will be explored through a follow-up study to reveal reasons why some participants disengaged with using the iPad.

6.2. Qualitative Analysis

The qualitative data collected from participant interviews and the weekly diary prompts were analysed to explore changes over time.

6.2.1. Interviews

The interview data was analysed across the three time frames; Early-Study (2-3 weeks of usage), Mid-Study (3 months of usage) and Final Study (6-months of usage), as shown in Figure 71.

Figure 71. An overview of the three longitudinal interviews
Semi-structured interviews took place in participants’ placement hospital locations and lasted between 20-40 minutes in duration. Interviews began with a verbal questionnaire (analysed in Chapter 5), then the first two interviews (Early & Mid) followed the same open question format, but allowed flexibility to explore any interesting themes that emerged. The final interview used each participant’s individual time-line graph (generated from their 24 weekly mini-questionnaire responses, see Figure 72), to prompt retrospective memory of the iPad usage over the 6-month duration and aid discussion.

![Figure 72. An individual participant’s timeline graph used as a prompt in the final-interview (Shows the Contextual Influences over the 24 weeks)](image)

Participants responded well to reviewing their individual time-lines as it enabled them to consider their past activity, and it aided disclosure of less salient affective responses that occurred over the last 6 months. Comparing the 47 individual graphs a few common patterns were shown, such as contextual fluctuations (shown in Figure 72), where ratings were influenced by participants activities (e.g., exams, hospital placements, holidays etc.), a typical familiarisation phase, were initial frustrations impacted on positive affect, but were soon overcome as participants became more familiar with the device (see Figure 73), and work-around problems, where some users may have experienced usability problems, impacting on positive affect, but eventually found work-around solutions to resolve them (Figure 73). The patterns were fairly evenly spread; familiarisation (30%) contextual (21%), work around (21%), with some time-lines showing no general pattern (28%); although there were considerable pattern variation within each group.
Chapter 6   Study Three: Longitudinal Study (Qualitative)

6.2.2. Diary Prompts

Participants were given the option to describe up to four activities (2 good and 2 bad) they had experienced while using their iPad each week, along with how they felt about each experience, and their responses were entered via a free-from format entry text box (see Diary Mini-Questionnaire Appendix 7). The data was grouped into six weekly time frames resulting in four groups: T1=(1-6 weeks), T2=(7-12 weeks), T3=(13-18 weeks), T4=(19-24 weeks). A total of 1687 self reported weekly text entries were coded using the same interview method in the previous studies (reported in Chapter 3 & 4), as follows.

6.2.3. Method

All quantitative data was coded and analysed using the same iterative process, as shown in Figure 74. This began first with a directed content analysis method (top down) that used an initial coding framework generated from the questionnaire constructs (*e.g.*, aesthetics, usability, ease-of-use), then a thematic open-coded method was used that allowed for additional themes to emerge (bottom-up), along with a key word/phrase search (*e.g.*, enjoyable, engaging, frustrating, useful etc.).

![Figure 73. Two common patterns in the individual participants time-line graphs: Initial familiarisation phase (left), Work-around usability problem (left)](diagram)

![Figure 74. The qualitative data analysis method for both interview and diary-prompt data](diagram)
Subsequently initial codes were grouped (axial coding) through an continuous process so frequency counts could be used to direct the overall analysis and develop the sub-themes (Hsieh & Shannon, 2005). Comment frequency for each coded theme was counted, or enumerated (Goetz & LeCompte, 1981), and Net Values (NV) were calculated for each theme, (by subtracting the negative from the positive frequencies), to access the valency for each theme (Cassell & Symon, 2004).

6.3. Qualitative Findings

An aggregated total of 6786 coded comments were split between the interview data (46%) and diary prompts (54%), as shown in Table 53.

<table>
<thead>
<tr>
<th></th>
<th>Interview NV</th>
<th></th>
<th>Diary NV</th>
<th></th>
<th>Total NV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td></td>
<td>%</td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>Quality Ratings</td>
<td>+176 (827)</td>
<td>41%</td>
<td>+327 (1197)</td>
<td>59%</td>
<td>+503 (2024)</td>
</tr>
<tr>
<td>Motivations &amp; Barriers</td>
<td>-46 (954)</td>
<td>48%</td>
<td>+386 (1036)</td>
<td>52%</td>
<td>+340 (1990)</td>
</tr>
<tr>
<td></td>
<td>-140 (288)</td>
<td>56%</td>
<td>+2 (230)</td>
<td>44%</td>
<td>-138 (518)</td>
</tr>
<tr>
<td>Interactivity</td>
<td>+162 (316)</td>
<td>60%</td>
<td>+122 (210)</td>
<td>40%</td>
<td>+284 (526)</td>
</tr>
<tr>
<td>Activities</td>
<td>+356 (730)</td>
<td>42%</td>
<td>+894 (1008)</td>
<td>58%</td>
<td>+1250 (1728)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>+508 (3105)</td>
<td>46%</td>
<td>+1731 (3681)</td>
<td>54%</td>
<td>+2239 (6786)</td>
</tr>
</tbody>
</table>

Table 53. Total frequency of comments comparing Interview and Diary-Prompts (% split), and the total frequency across the five themes (% Split)

A total of 3105 interview comments were coded across each of the 3 time frames with an overall positive NV (+508), with the most comments collected during the final interview (Early=31%, Mid=24%, and 45%), as these were more in-depth. The diary prompts generated a total of 3681 comments that were extracted from the (1687) weekly text entry and showed an overall positive NV (+1731), with the most comments reported at the start of the study (T1=42%, T2=28%, T3=19% and T4=11%). The drop in self-reporting comments can be attributed to a number of factors; a familiarity of using the device (so fewer comments were reported), or diary study fatigue as participants became used to completing the diary-prompt. These trends were common across all qualitative themes causing a method bias, so any time changes reported within the sub-themes will allow for this bias. There was an equal spread across each of the four hospital groups, so location had little influence.

Both interview and diary-prompt comments were coded across each of the five coding themes: Quality Ratings (Device & Apps) (30%), Motivations and Barriers (29%), Positive or Negative Affect (8%), Interactivity (8%), and Learning or Leisure Activities (25%), with a total positive NV (+2239); so overall participants gave more favourable comments about the iPad.
Chapter 6  Study Three: Longitudinal Study (Qualitative)

The findings from each of the five themes are now reported along with examples of participant quotes.

6.3.1. **Quality Ratings (Device and Applications)**

In order to tease apart attitudes towards the apps (contained on the iPad), and the device as a whole, *Quality* comments were split between *Device* and the *Applications*. A total of 827 interview (quality) comments were evenly spread between device (51%) and apps (49%); and a total of 1197 diary prompt (quality) comments were collected, with apps generating the highest frequency (66% of quality comments).

<table>
<thead>
<tr>
<th>Q. Ratings</th>
<th>Early</th>
<th>Mid</th>
<th>Final</th>
<th>TOTAL</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERVIEW</td>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
</tr>
<tr>
<td>Device</td>
<td>+3</td>
<td>26%</td>
<td>+5</td>
<td>19%</td>
<td>+23</td>
</tr>
<tr>
<td>Application</td>
<td>+25</td>
<td>22%</td>
<td>+71</td>
<td>29%</td>
<td>+49</td>
</tr>
<tr>
<td>DIARY</td>
<td>T1: 1-6 weeks</td>
<td>NV</td>
<td>T2: 7-12 weeks</td>
<td>NV</td>
<td>T3: 13-18 weeks</td>
</tr>
<tr>
<td>Device</td>
<td>-19</td>
<td>33%</td>
<td>-30</td>
<td>28%</td>
<td>+5</td>
</tr>
<tr>
<td>Application</td>
<td>+134</td>
<td>49%</td>
<td>+142</td>
<td>29%</td>
<td>+52</td>
</tr>
</tbody>
</table>

*Table 54. The NV and % frequency of Quality (Device & Apps) comments for interviews and diary-prompts*

The apps received a much higher positive NV for both the Interview: (+145) and Prompts: (+355), with both showing a mid-study peak in app NV scores: (Interview: Early=+25, Mid=+71, Final=+49), (Prompts: T1=+134, T2=+142, T3=+52), see Table 54. The interview device NV score (NV=+31) was much lower than the apps, and the prompts showed a negative NV score (-28), although both scores improved over time (Interview: Early=+3 to Final=+23, Prompt: T1=-19 to T4=+16). So initially app comments were very positive, but declined over time, while the device received very negative comments, which improved slightly over time.

The *Quality* comments were further divided across the five sub (quality rating) themes:

- **Utility/Content**: Utility refers to the device quality and suitability of use. While content refers to the suitability of information contained in the apps.
- **Functionality**: The performance and efficiency of the device features, and the enabling functionality provided by the app (e.g., recording, note-taking etc.).
- **Usability**: The device/app faults and defects that impact on usability.
- **Usefulness**: How beneficial the device/app is to the user.
- **Ease-of-Use**: General comments on how user-friendly the device/apps were to use.
- **Aesthetics**: The visual appeal of the device/app (colour, layout, design etc.)
The apps receiving an overall positive NV score (+500) compared to device (+3), with negative device comments being attributed to Functionality, Utility, and Usability; whereas all the app themes received positive ratings apart from Usability, see Table 55.

<table>
<thead>
<tr>
<th>QUALITY</th>
<th>INTERVIEWS</th>
<th>PROMPTS</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DEVICE</td>
<td>APP</td>
<td>DEVICE</td>
</tr>
<tr>
<td></td>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
</tr>
<tr>
<td>Utility/Content</td>
<td>-27</td>
<td>19%</td>
<td>+31</td>
</tr>
<tr>
<td>Functionality</td>
<td>-51</td>
<td>30%</td>
<td>+72</td>
</tr>
<tr>
<td>Usability</td>
<td>-19</td>
<td>10%</td>
<td>-102</td>
</tr>
<tr>
<td>Usefulness</td>
<td>+43</td>
<td>16%</td>
<td>+63</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>+55</td>
<td>13%</td>
<td>+55</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>+30</td>
<td>11%</td>
<td>+26</td>
</tr>
<tr>
<td>TOTALS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 55. The NV and % frequency of the 6 Quality themes split between the Device & Prompts

6.3.1.1. Utility and Content

All device Utility comments (19% of device) received a negative NV score (-62) compared to apps (12%), which were positive (NV +123). Initial device utility comments related to the poor camera quality; “I don’t really like the camera, I think its got very poor resolution, so I don’t use it for photos” [P 20], and poor battery life; “its quite unpredictable how quickly the iPad battery is going to run out…” [P29, T1], which declined as users became familiar with how frequent the device need charging; “the battery died again in the middle of my revision in the library, it is annoying, but I think it is more about me not preparing beforehand” [P10, T2], while the lack of memory space; “I am rapidly running out of memory on my iPad, which is very frustrating” [P47, T2], was reported after a few months of usage. Positive app comments increased in the first 3 months of usage (Early=-2, Mid=+19), as participants explored the “the diversity of the apps” [P16]. Comments related to specific app content; “there is an app for everything, so there is a massive amount of content” [P19], especially the medical related apps that; “include the essential and important information” [P6], that aided learning; “I am able to access information on the go, its very useful and helpful, it saves time and prevents the risk of forgetting to look something up later” [P20].

6.3.1.2. Functionality

The total device functionality comments received a high frequency (30% of device comments) and a negative NV (-80), while app functionality gained a slightly higher frequency (39% of app comments), but a strong positive NV (+289). Negative device functionality comments related to a combination of reasons; lack of a central filing system, “Things like the central file storage issue where you can’t file everything in one place” [P30], not being able to multi-task; “What I would really like is if you could scroll between screens” [P1], as “you cannot have two windows open at the same time” [P40, T2], incompatibility

176
Chapter 6   Study Three: Longitudinal Study (Qualitative)

between devices; “I’m not entirely sure how to sync everything, so that’s a bit frustrating” [P25], and the lack of support for Adobe Flash player; “it doesn’t play flash, which is annoying” [P7]. Initial work-around solutions were reported such as using Dropbox; “I’ve pretty much got everything revolving around Dropbox now” [P23], but raised further problems later; “You can’t edit in Dropbox, so you have to open it into something else, so you can’t multi task between different types with the same format, it’s quite frustrating” [P30].

Positive app functionality comments related to specific app features such as Notability; “I like the fact you can basically do all your recordings, all your notes, all your filing, so I like that app” [P8], and GoodReader: “I could read my PDF books and highlight the text” [P7, T1], with the mid-study peak (Early=+8, Mid=+42, Final=+22) being attributed to more apps being discovered, such as exam revision apps; “I’m using a lot of revision apps as well and they only came out quite recently… I use them loads during exam time” [P30], which coincided with half-yearly exams which enabled students to “go over past exam questions” [P21] or “produce good mind maps to use for revision” [P41, T1].

6.3.1.3. Usability

All app ratings were positive, apart from usability, which generated a high frequency of negative NV comments (20%, -234), compared to device usability NV (12%, -80). App usability comments often referred to the mandatory apps such as Numbers; “its awful, I just hate it... its not straightforward, none of it is logical” [P19], and Manchester eForms app; “which isn’t great because the boxes are tiny, as you start filling out your responses you can’t see half of what you’ve written... its really awkward... but we have to use that one” [P40], and the VLE Blackboard app; “its just not very user friendly” [P37]. However, app usability was somehow forgiven due to its unique functionality; “the user interface of Numbers is difficult to use... I am motivated as it is the only spreadsheet app I am aware off, although it took longer then it should have editing” [P16], whereas popular apps (e.g., iBooks, Facebook etc.) received fewer usability comments. Device usability comments related mainly to software problems; “the screen has locked a couple of times this week, so I had to turn it off and on again which is quite frustrating” [P40, T2], and the keyboard control; “you cant turn off the annoying features like auto-suggest words and keep useful ones like spell check” [P42, T3].

6.3.1.4. Usefulness

A similar frequency of Usefulness comments was reported, with apps (17%, +178) gaining a higher positive NV then device (23%, +110). Device usefulness comments remained constant (T1=+19 to T4=+18) with comments reflecting its increasing usefulness; “the more you know how to use it [iPad] and the more useful it becomes” [P46]; while app usefulness comments showed a drop over time (T1=+47 to T4=+5), with the highest frequency being reported in
the first 3 months (Early=+27, Mid=+17, Final=+19) (reflecting an initial investigation phase) where participants determined what apps are the most useful; “certain applications have to be useful” [P29] and “you need to have the right stuff on the iPad for it to be useful” [P4].

6.3.1.5. Ease of Use
Both the device (+85) and apps (+118) were reported easy-to-use, despite the high usability comments reported, with general ease-of-use comments; “I love the way it's so easy and simple to use and much quicker to set up than the laptop” [P47], and to specific apps being “quick and easy to use” [P11, T1]. Comments increased over time for both device and app (Early +24 to Final +56), as participants became more familiar with using iPad. Also, due to the wide variety of apps, with many offering an array of similar functionality; “I like that you can get an app for anything, if you want it to do something it probably will do it” [P51], participants had the flexibility to easily replace unusable apps, “I download it and if it doesn’t do what I need it to do, I just delete it” [P47].

6.3.1.6. Aesthetics
No Aesthetics comments were reported in the diary-prompts, and the interviews received a low frequency of comments for both device (6%, NV +30) and app (2%, NV +26) showing little difference in NV scores; with comments relating to the appearance of the device; “its very sleek and professional looking”[P32], or the app icons; “if an app doesn’t look good on my home screen, I delete it!” [P4], which impacted on app choice: “if it looks new, if it looks modern, you’re more likely to choose it than something that’s a bit just all shoddily done” [27].

6.3.2. Motivation and Barriers
To investigate the factors that may impact on user satisfaction and long term use of the iPad, the qualitative data was split two themes; aspects that motivated product usage, and conversely, aspects that provided barriers to use, which could ultimately hinder product adoption.

A total of 1990 interview and diary prompt comments were evenly spread between motivators (52%) and barriers (48%). The interview comments were evenly divided (50%), while the diary prompts receiving a higher frequency of motivator comments (69%). The motivator diary study comments gained a higher positive NV score and greater frequency (+596, 69%) than the interview (+259, 50%), while conversely, the interviews gained the highest frequency of comments (-305, 50%), than the diary prompts (-201, 31%). Overall, motivator gained a higher positive NV (+855) compared to barriers (-515), and more motivator related comments were given compared to barrier comments.
Further analysis was conducted to separate the app related comments from the general device comments in each of the Motivator and Barrier eight sub-themes, see Table 56. From the total number of motivator and barrier comments (1990), the device generated the highest frequency of comments (69%) compared to apps (31%), where device generating 80% of barrier comments, and 62% of motivator comments.

<table>
<thead>
<tr>
<th>MOTIVATORS</th>
<th>Device</th>
<th>App</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
</tr>
<tr>
<td>Accessibility</td>
<td>+281</td>
<td>67%</td>
<td>+119</td>
</tr>
<tr>
<td>Portability</td>
<td>+137</td>
<td>67%</td>
<td>+67</td>
</tr>
<tr>
<td>Repository</td>
<td>+31</td>
<td>23%</td>
<td>+111</td>
</tr>
<tr>
<td>Perceived Utility</td>
<td>-9</td>
<td>67%</td>
<td>+118</td>
</tr>
<tr>
<td>TOTALS</td>
<td>+440</td>
<td>67%</td>
<td>+118</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BARRIERS</th>
<th>Device</th>
<th>App</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
</tr>
<tr>
<td>Context of Use</td>
<td>-194</td>
<td>99%</td>
<td>+1</td>
</tr>
<tr>
<td>Connectivity</td>
<td>-183</td>
<td>99%</td>
<td>+2</td>
</tr>
<tr>
<td>Reluctance to Change</td>
<td>-44</td>
<td>47%</td>
<td>-3</td>
</tr>
<tr>
<td>Ordinarily</td>
<td>-77</td>
<td>84%</td>
<td>-15</td>
</tr>
<tr>
<td>TOTALS</td>
<td>-500</td>
<td>-15</td>
<td>-515</td>
</tr>
</tbody>
</table>

Table 56. The Motivator and Barrier sub-themes results split between Device and Apps

6.3.2.1. Motivators

The motivator comments were divided into four sub-themes: Accessibility (immediacy-of-use), Portability (lightweight & small), Repository (information archive), and Perceived Utility (appropriateness of use), see Table 56.

Accessibility of the iPad was the strongest motivator of use, gaining the highest NV (+400) overall, and a high frequency of (motivator) comments (36%), with comments referring to immediacy; “its always on... its instant rather than having to fire up the computer” [P37]; saving time; “its so quick, I can be online in 10 seconds - have found my information and off again in under a minute” [P40, T3]; filling in time (30%); “during the day there tends to be really big gaps where I have nothing to do, so the iPad encourages me to do more work” [P10, T2]; and just-in-time learning (20%), “to look up drugs in the oncology clinic to help my understanding... I was able to do this in the limited time between patients and this helped me make more of my time” [P13, T1].

The second strongest motivator (+204) was the portability of the iPad, as “its compact, easy to take anywhere” [P21], receiving much lower comments (17%) than accessibility. Both Accessibility (+281) and Portability (+137) scored the highest positive NV device-related comments, and generating a greater frequency of comments (67%) compared to apps (+119/+67), see Table 56.

In third position was Repository (NV +142, 12%) which refers to the ability to store multiple books; “I've got six books in my bag at the moment, one iPad that I can fit into by handbag, its everything in one”[P7], documents; “the iPad has made carrying documents, textbooks... so much easier...books and notes have been condensed to the featherweight of an easily portable
"iPad" [P13, T4], and notes; "I do like carrying all my notes in one place" [P10], as "it's a lot lighter than carrying massive pads of paper around and various pens" [P 19]. Repository comments were more app-related, scoring a higher NV and a greater frequency (77%, +111), compared to device (+31), with comments relating to specific apps that enable the iPad to be used as storage device; "I like the fact that all my notes are in one place, so if I go to Notability everything would be there, and like iBooks, instead of having to carry books around with you. I think that's really, really convenient" [P10].

Perceived Utility gained a high frequency of motivator comments (35%), but the lowest NV (+109), due to a mix of positive and negative comments, which were mainly device-centric (67%, NV-9), compared to the positive app-related comments (+118). Negative device-comments referred to the applicability of using the device to support learning activities, such as using the touch screen to type; "I don’t particularly like typing on it... I just would rather write, I find it quicker" [P50], which was often due to poor functionality “the autocorrect on the iPad makes writing notes very frustrating and off putting” [P10, T1], with some users resorted back to using their laptops; “when having to type large amounts of text it becomes very tedious, its much quicker to type on my laptop” [P47, T1]. Positive app-comments referring to the benefit of using the learning apps that support learning; “to quickly look up drug types in clinic helped me follow what was going on better” [P3, T3], aid revision; “the iPad was very helpful in practising questions for the progress test” [P36, T20], and enable reflective learning; “recording myself to do clinical skills was very useful to watch myself back to improve my learning” [P13, T3].

6.3.2.2. Barriers

There was little difference in the frequency of barrier comments between Context of Use (31%), Connectivity (24%) and Reluctance to Change (31%), although Context received the highest negative NV score (-193), marginally higher than Connectivity (-181), both of which are device-centric (99%), with Connectivity (-183) and Context of Use (-194) scoring the highest NV and frequency compared to apps, which generated very few comments (1-2%), see Table 66.

Context of use refers to the impracticalities of using the iPad within clinical settings, due to the size; “its just big and clunky to carry around with you” [P30], and "quite difficult to find a good place to put it if you have to be hands on with a patient” [P21], as “its difficult to walk round and type at the same time” [P23]; the hospital culture as “some of the consultants are quite old school so they don’t really like it” [25]; along with participants choosing not to use the iPad in front of patients as “it creates a physical barrier between patients and student” [P21, T1], which can “hamper your interaction with some patients” [22]. Connectivity relates to
participants “frustrating due to the lack of internet access” [P37, T3], as “without its connection to the internet... there is not much that I really use it for” [P27], thus creating a barrier to use which “limits my access to learning material” [P2, T1], as “a lot of medical apps rely on it” [P40, T1], making “the iPad slightly useless for me this week!” [P43, T1]. Although work-around strategies were reported; “I’m careful to download stuff onto it when I do have Wi-Fi access, so I know to plan ahead”[P3]. Connectivity was a consistent barrier to use; “it’s a lot less functional if you don’t have the Internet” [P23], where often participants reverted back to using their other devices; “I recently just got an iPhone, so I have used my iPad less as I am able to use all my apps all the time on my phone as I have constant Internet on my phone” [P25, T2].

Reluctance to Change received the lowest NV (-49), with comments split between device/apps (47/53%), but generating a higher negative device NV (-44), reflecting participants disposition, with some reporting a lack of willingness to change their behaviour towards using new technology (the iPad). This is linked closely with participants preferred learning styles; “Its just not my style...its hard to teach an old dog new tricks, so I went back to what I know, because its what works for me”[51]; a lack of know how; “I’m not very good with technology, so someone always has to teach me how to use it” [P25]; and poor advice; “the university had a lecture on how to use the iPad and they went so fast and I was just like I really don’t know what you are talking about” [P38]. Comments in this theme were investigated to see if there were any similarities between the participants who gave negative comments to the Power User groups identified in Chapter 5 (Section 5.6.1). Interestingly, 59% of the negative comments collected were generated from the Low Power User group, (three of which have provided quotes above), 17% were from the Mid Power User group, and 24% from the High Power Group.

Ordinary, or over familiarity received the lowest frequency of comments and NV score (13%, -92), with the interview comments reporting the most comments (21%, -84) compared to prompts (8%, -8). Most comments were device-centric (84%, -77), compared to a much lower app NV score (-15), with the highest frequency reported during the final interview (-76, 37%) indicating a waning novelty effect; “the novelty’s worn off, its just I don’t use it as much as I should or could” [P50], and a drop in aesthetic appreciation; “I see it just as something normal now instead of very stylish” [P21]. However this category was probably only a mild barrier (if any) to use, demonstrating the typical novelty effect (Karapanos, Martins, & Hassenzahl, 2009b) where over time participants are less inclined to spent time exploring the iPad as they become more familiar with how to use it; “at the beginning it was a novelty, then I think now I’ve got a little bit busier and its kind of just there” [P32] and ”the novelty has worn off a little bit, so its not like I’m trying to find loads to do on it” [P41].
6.3.3. Affect

The affective comments were coded across positive and negative themes, and then split between app and device related comments. All the interview NV scores (Early, Mid and Final) were normalised against interview times to allow for variation in interview times.

A total of 518 affect related comments were split equally between Positive Affect (46%) and Negative Affect (54%), with the interview comments collecting more negative comments (60%), and the diary prompts more positive comments (69%), as during the interviews participants were prompted more to disclose negative issues. Positive Affect comments showed a sharp decrease in time with the highest NV score reported in the first few weeks for interviews (Early=+57 to Final=-6), while negative affective interview comments more than doubled over time; (Early=-57 to Final=-124), with diary-prompts showed a decline in all affect comments (attributed to method bias).

All affective comments were coded into six sub-themes, derived from key word searches (and derivatives); for Positive Affect: Enjoy (Fun), Satisfying (Pleased), and Exciting (Engaging); and for Negative Affect: Frustrating, Annoying (Irritating), Boring (Tedious). Some affective key words showed no results so were excluded in the analysis (e.g., Pleasure, Curious, Attractive).

Despite a similar positive/negative frequency split (46/54%), Negative Affect received a much higher NV score (-318) than positive (-181), which was equally split between device (51%) and apps (49%). Participants reported more Negative Affect comments, as these may have been more salient in their memory then Positive Affect comments. However, the apps received a much higher Positive Affect NV score (+156) compared to device (+25), and a much lower Negative Affect score (-113) than device (-206), see Table 57.

<table>
<thead>
<tr>
<th>POSITIVE AFFECT</th>
<th>NEGATIVE AFFECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>App</td>
</tr>
<tr>
<td>NV</td>
<td>% Freq</td>
</tr>
<tr>
<td>Satisfying/Pleased</td>
<td>34</td>
</tr>
<tr>
<td>Enjoyable</td>
<td>-1</td>
</tr>
<tr>
<td>Exciting/Engaging</td>
<td>-8</td>
</tr>
<tr>
<td>TOTALS</td>
<td>25</td>
</tr>
</tbody>
</table>

*Table 57. The Positive and Negative Affect sub-themes results split between Device and Apps*

6.3.3.1. Positive Affect

The greatest frequency of Positive Affect comments related to Satisfaction (or pleased), which accounted for over half all the Positive Affect comments (56%, NV=+119), which was more app-related (+85/70%) with comments relating to favourite apps; “I love the Skype app, that’s
brilliant, its so much easier to use than your laptop” [P25], and apps that supported learning; “I downloaded BMJ... I feel that it has facilitated my learning and feel satisfied and pleased I downloaded this app” [P14]. While device-related comments referred to general usefulness; “the iPad itself is fantastic, its made my life so much easier and I have seen my grades go up, which I think is brilliant” [P38].

Enjoyable generated a much lower frequency of comments (29%, NV=+58), with the majority being app-related (+58/65%), with most comments referring to leisure use; “I’m still getting great enjoyment from the apps I have downloaded for leisure use”, [P19], particularly gaming apps; “I played Sudoku this week, I enjoyed it because it was fun and kept my brain engaged” [P11], and general leisure apps; “shopping on the iPad both online and on apps is great! Yet another thing I can enjoy on my iPad” [P19, T2]. Although later comments became more negative “some of the novelty has worn off, so its not so enjoyable”[P41] due to a familiarisation of using the device over time.

Excitement & Engaging comments were least reported (15%, NV=+4) with comments relating to app functions; “the new Prognosis app I have been using has been very engaging and a good way to test my knowledge” [P50, T1], and general usage; “I am still discovering new and exciting ways to use my iPad to enhance both my work and leisure activities” [19, T4]; while later more negative device comments related to the lack of excitement due to familiarisation; “I was quite exciting about it and then you sort of tail off a bit...because I don’t really engage with it”[P7] and; “you use it for a while and then the novelty rubs off, its not as exciting...its not a new toy” [P29].

6.3.3.2. Negative Affect
The main contributor for negative affect was frustration (NV=-165, 52%), with more device-centric comments (-101, 65%); closely followed by annoying (NV=-136, 39%), also showing more device related comments (-95, 71%), see Table 57.

Frustration device related comments reflect the barrier themes, such as the lack of connectivity; ““the internet access, drives me insane. That’s the most frustrating thing now” [P17], device usability; “I don’t like typing on it, its quite frustrating... and that autocorrect drives me nuts”[P29]; device utility; “I got really frustrated because I kept on running out of space”. Apps received lower negative NV for frustrating (-64), and irritating (-41), with comments relating to app usability; “a lot of the frustration is to do with the apps... a number of my apps were crashing on launch...” [P38]; although workaround solutions were reported; “I’m probably less frustrated with it now because I’ve worked out little tricks...I’ve just found other ways to get round it” [P40].
Chapter 6  Study Three: Longitudinal Study (Qualitative)

Initial (early-study) Annoying comments related to the learning experience; “I still haven’t figured out how to sync my timetable with the iPad calendar and it’s so annoying”[38], whereas later comments reflect similar frustration comments, such as device functionality; “really annoying not to be able to connect to my printer to the iPad so I can’t print directly” [47, T1], and “it doesn’t play Flash, that’s annoying”[P7], the incompatibility; “tried to read a doc. I downloaded, but realised it was incompatible with the iPad, this really annoyed me” [P25, T1], and login problems; “it takes so long to log into… then you leave it for a minute and it goes to sleep, then you do it all over again… its the most annoying thing” [P50].

Boredom related comments were low NV (-17, 10%), with the highest score occurring after 3-6 months of usage (Early=-1, Mid=-1, to Final= -10) as participants had a longer time for boredom to occur, with little app/device difference (-10/-7). Comments related to gaming apps; “I haven’t been playing games for quite some time, it’s just boredom” [P45], or general device familiarity; “I think the novelty’s work off, I don’t use it as much as I used to” [P5].

6.3.4. Interactivity

Interactivity related comments were split across four different sub-themes, Interactive, Communicate, Video, and Games. The first sub-theme Interactive generated comments from a word search for ‘interactive’ words (zoom, touch, swipe, action, flip, flick and flip).

Communicate related to activities that enabled interactive communications, (e.g. Face-time, Skype, Message, Talk). Video related to all comments relating to video media (e.g., watch, view, movie, TV, YouTube, iPlayer, etc.), and Games were comments relating to using the iPad to play games.

A total of 526 comments were coded across the interview and diary-study with the interview comments showing a slightly higher frequency and NV score (60%, +162) compared to diary (40%, +122). Little difference was found over time in both frequency (34%/44%) and NV ratings (Early= +57 to Final= +60) for interview comments. All comments were coded according to device or app, across each of the four themes, as shown in Table 58.

<table>
<thead>
<tr>
<th></th>
<th>INTERVIEWS</th>
<th>PROMPTS</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DEVICE</td>
<td>APP</td>
<td>DEVICE</td>
</tr>
<tr>
<td>INTERVIEW</td>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
</tr>
<tr>
<td>Interactivity</td>
<td>+7</td>
<td>32%</td>
<td>-1</td>
</tr>
<tr>
<td>Communicate</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Video Media</td>
<td>-17</td>
<td>41%</td>
<td>+21</td>
</tr>
<tr>
<td>Games</td>
<td>-11</td>
<td>27%</td>
<td>+21</td>
</tr>
<tr>
<td>TOTALS</td>
<td>-21</td>
<td>(77)</td>
<td>+183</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>(150)</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 58. Interactivity sub-themes results split between Interview/Prompts and Device/App
The apps generated the highest frequency of comments (71%) and a high positive NV (+284), compared to device, which scored a zero NV indicating that interactivity was mainly app related. Video Media generated almost half the frequency of interactivity comments (49%), and scored the highest NV (+142), with apps scoring a much higher NV (+138) than device (+4). Positive app comments related to; “using YouTube videos for interactive teaching was a great way to revise, as I was getting rather bored with my other learning techniques” [P25], as it supported learning; “I guess I’m more of a visual learning... I do prefer learning from watching online lectures as opposed to just reading” [P19], as it provided alternative learning method; “the university lecture videos are quite interactive, so you can use your iPad, as its more interactive you kind of learn outside the book, rather than just reading books, as it gives you something else” [P4]. Further app comments related to watching Video Media for leisure use while on the go; “travelling with it you can watch iPlayer as its far more portable than a laptop” [P33].

Device-related Video Media received a high frequency of comments (60%), which were mixed (positive and negative), resulting in the low NV score (+4). Positive device comments referred to the portability of the iPad; “I have found the iPad very convenient, especially when I can take it to the kitchen and watch lectures while I am cooking” [P29], which enhanced motivation; “watching lectures in places where a laptop would be inconvenient. It helps with my learning and it makes me what to use the iPad more” [P10], as it was “helpful to use the iPad to view videos as it was portable. It was engaging as this was a new experience” [P46]. Also, “the option to record video and then play it back in order to review my technique went well” [P19], which was then used in group learning feedback sessions; “using the iPad with a group to review a patient history taking video was very useful as I picked up areas of improvement that I would not have thought of myself. I felt it was a very good use of the technology” [P3]. Negative Video Media device comments related to device usability issues; “the reflective screen is sometimes difficult to see if you are watching TV” [P40], and “watching videos on the iPad is really inconvenient because if you tilt the screen a little it spins round” [P10], and functionality due to battery life as “it does consume the battery quite quickly” [P29], and “the video quality is not brilliant”, along with having “no flash player to stream live videos which is annoying” [P39].

Games were the second strongest interactive activity gaining (30%) of the comments, showing a positive app-related NV score (+82), due to the variety of gaming apps; “how it stays engaging, is because when you get bored, you can download something new” [P19], and apps that enhance interactivity; “I drew Scooby Doo on Draw Something, it actually looked quite good” [P9]. The negative device score for Games (-10) relates to the preference of using other devices (mobile or laptop) to play games; “I play plenty of games on various things but not just on the iPad” [P6], and “if I’m going to play games I’ll use my laptop” [P29].
Both Communicate (+33) and Interactive (+37) scored an overall similar NV score, with both showing a strong app rating (+33, +31). Communicate related to specific apps that enable communications; “I use FaceTime to interact with friends as opposed to calling them” [P19], and “Skype is very good for keeping in touch with friends” [P27]. Device interactive comments (+6) were not so high as apps (+37), with more negative comments (than positive), with a polarisation of comments relating to the touch screen, where some participants favoured it; “I like the whole touch screen, I find it quite natural...it’s a bit more interactive” [P51], and “even when we are doing work its quite fun to be able to touch things, play with your hands and stuff” [P29]; while others found it problematic; “I think the touch screen, sometimes it can be quite frustrating when your trying to touch something and it doesn’t select the correct thing” [P50].

6.3.5. Activities

A total of 1728 activity related comments were coded across two themes: Work Activity (72%) and Leisure Activity (28%), with work activity receiving the highest frequency of comments and highest NV score (+933) compared to leisure (+317), indicating that the iPad was being used as a learning support tool (see Table 59).

Fewer frequent leisure activities were captured in the diary prompts (20%), than in interviews (40%), indicating leisure activities were less salient in memory, and were only reported once prompted during the interviews. Both activities declined over time, with work showing the sharpest, for interviews. All activity comments were divided into 10 sub-groups, split equally between work and leisure activity themes, see Table 59.

<table>
<thead>
<tr>
<th>WORK ACTIVITY</th>
<th>LEISURE ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviews</td>
<td>Interviews</td>
</tr>
<tr>
<td>NV % Freq</td>
<td>NV % Freq</td>
</tr>
<tr>
<td>Prompts</td>
<td>Prompts</td>
</tr>
<tr>
<td>NV % Freq</td>
<td>NV % Freq</td>
</tr>
<tr>
<td>TOTAL</td>
<td>TOTAL</td>
</tr>
<tr>
<td>Learning Apps</td>
<td>Leisure Apps</td>
</tr>
<tr>
<td>+83 26%</td>
<td>+28 10%</td>
</tr>
<tr>
<td>Notes</td>
<td>Connecting</td>
</tr>
<tr>
<td>+65 43%</td>
<td>+69 28%</td>
</tr>
<tr>
<td>Reading</td>
<td>Video Leisure</td>
</tr>
<tr>
<td>+26 16%</td>
<td>+32 15%</td>
</tr>
<tr>
<td>Video Learning</td>
<td>Games</td>
</tr>
<tr>
<td>+15 6%</td>
<td>-1 31%</td>
</tr>
<tr>
<td>Organising</td>
<td>Music/Photos</td>
</tr>
<tr>
<td>-36 8%</td>
<td>+3 16%</td>
</tr>
<tr>
<td>TOTALS</td>
<td>TOTALS</td>
</tr>
<tr>
<td>+225 100%</td>
<td>+131 100%</td>
</tr>
</tbody>
</table>

Table 59. Work & Leisure Activity themes results split between interview & prompts

6.3.5.1. Work Activity

The diary prompts collected the greatest number of work related comments (76%) than interviews, and received a much higher positive NV (+708), compared to leisure activities (-210), probably due to the frequency of diary-prompt questionnaires, that captured weekly activities more closely. The learning apps (43%, +451) generated the highest frequency and NV score, with the greatest number of comments reported in the first 6 weeks (T1: +119 to
T4: +26), indicating an early learning phase; “I was trying out different apps and things I think earlier on and then perhaps it sort of tailed off a little bit” [P22]. Using the iPad to write, annotate or organise notes received a high NV (+195) and frequency (28%) of work activity comments with the interviews receiving the highest frequency (43%), but a lower NV (+65) compared to the prompts (20%, +130), indicating the interviews captured more negative comments than the self-reported prompts.

Using the iPad to read received a marginally higher score (12%, +103) than using video to support learning (9%, +100), with reading comments relating to individual preferences; “I don’t use the screen to read... I prefer books yes - I need to physically, see it” [P32]; and video learning comments referred to “watching videos for interactive learning was a great way to revise” [P25, T2], or to video record clinical practice, “I used the iPad to video me interviewing a simulated patient, I was then able to upload the clip and reflect on it” [P27, T1]. Organising received the lowest frequency of comments (8%, +84), as participants reported finding new ways to keep organised; “it defiantly helps me be more organised” [P13], “just keeping it all in the same place is really good” [P8], and “using the calendar function to keep organised I found really useful” [P8].

6.3.5.2. Leisure Activity

Similar to work activity, the use of leisure apps (e.g., banking, retail, recipes, or news apps) received a high frequency and NV score (+113, 23%), which also declined over time (Early= +13 to Final= +5), possibly due to familiarity. Using the iPad for “keeping in touch with friends” [P16] via social networking apps (e.g., Facebook or Skype) was a frequent activity (23%) and received a high positive NV (+93), although this was not exclusive to the iPad; “I prefer to do the social things on my phone because I’ve got the internet at all times, so you can’t really do it on your iPad” [P11]. Playing games received the highest frequency of comments (25%), but the lowest NV rating (+27), that decreased over time (Early= +18 to Final= -22), due to novelty effect; “I think obviously at the start, oh my god I can play Angry Birds all the time and I had quite a few...I’m just not interested now, it’s a phase that’s passed now” [25]. Watching videos for leisure (+45, 12%) and listening to music or taking photos (+39, 17%) scored similar frequencies and NV ratings, with videos showing little change over time (Early= +10 to Final= +8), but photos and music showed a decrease (Early= +10 to Final= -12) with more negative comments relating to the limited memory and poor camera; “the quality of photos are not very good” [40, T2].
6.3.6. Application Types

A total of 1039 app related comments were collected with an overall positive NV score (+495). Comments were divided between application type (as listed in Table 60), as follows; Work, Dual Purpose, Mandatory, Social and Medical. Overall the highest frequency of comments related to either medical (24%) or work (23%) related apps, with medical apps gaining the highest NV score (+170) compared to work (+101). Social apps scored the second highest NV score (+133), followed by dual-purpose (+104) apps, with both sharing the same frequency of comments (18%). Mandatory apps generated the lowest frequency of comments (16%), and an overall negative NV score (-13), indicating they were the least preferred apps.

<table>
<thead>
<tr>
<th>App Type</th>
<th>Interview NV</th>
<th>% Freq</th>
<th>Diary NV</th>
<th>% Freq</th>
<th>Grand Total NV</th>
<th>% Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Apps</td>
<td>+55</td>
<td>48%</td>
<td>+46</td>
<td>52%</td>
<td>+101</td>
<td>23%</td>
</tr>
<tr>
<td>Dual Purpose Apps</td>
<td>+54</td>
<td>54%</td>
<td>+50</td>
<td>46%</td>
<td>+104</td>
<td>18%</td>
</tr>
<tr>
<td>Social Apps</td>
<td>+97</td>
<td>72%</td>
<td>+36</td>
<td>28%</td>
<td>+133</td>
<td>18%</td>
</tr>
<tr>
<td>Medical Apps</td>
<td>+95</td>
<td>56%</td>
<td>+75</td>
<td>44%</td>
<td>+170</td>
<td>24%</td>
</tr>
<tr>
<td>Mandatory Apps</td>
<td>-12</td>
<td>48%</td>
<td>-1</td>
<td>52%</td>
<td>-13</td>
<td>16%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>+289</td>
<td>(575)</td>
<td>+206</td>
<td>(464)</td>
<td>+495</td>
<td>(1039)</td>
</tr>
</tbody>
</table>

Table 60. Summary of Apps results grouped into App Types, showing NV and % Frequency Scores

Interview comments gained a slightly higher frequency of comments (55%) than diary-prompts across all themes, with social apps showing the higher frequency (72%), compared to diary (28%), indicating that participants were less likely to self-report social apps in the diary study. Prompts were given during the interviews to capture social-related app usage, since participants were less likely to self-report social app usage, as work related activity appeared more salient in memory.

Over time all app related comments decreased for both interviews: Early= (+157, 46%, to Final= +60, 31%), and prompts: (T1= +89, 50%), to T4= +17, 11%), with no obvious time trends found across the five app types.

6.3.6.1. Individual App Analysis

Table 61 shows the rank rating based upon the aggregated (interview and diary) app NV score for each individual application contained within each app type. The top ranking apps were clinical, social or dual-purpose apps, which were mainly self-selected, while the University mandatory apps were ranked the lowest. Word of mouth and app store ratings appeared to be important influences on choice; “Notability was highly recommended by some of my friends” [P21], and “I tend to go on the App store and have a look through the most popular apps” [P43].
Work Apps

Both Notability and Facebook scored the highest NV result (+57), with Notability gaining a higher frequency of comments (12%) compared to Facebook (8%). Notability was the only high scoring work-based app, being favored over its rival Evernote (+3/4%), ranked the lowest (18) of all work-based app; “I downloaded Evernote to start off with. I didn’t like it. So I now use Notability instead” [P19], mainly due to its limited functionality; “I am not fond of Evernote as a note-taker as you cannot draw… and files can only be attached or tagged the same, its disappointing” [P30]. Notability was preferred for its multi-functionality; “I like the fact you can store all your recordings, all your notes, all your filing” [P8], and chosen by recommendation, “the reason I got Notability, cause one of my friends got it and I like what it looked like” [P8]. However, Notability comments dropped over time (Early= +31 to +8), reflecting a decline of note-taking activity; “I definitely used Notability a lot more at the beginning of the year in group discussions, or in lectures and then I just ended up resorting back to my laptop and notebook if I ever take notes…”[P26].

Similarly, Prezi (ranked 13), the presentational app scored a higher NV (+9) compared to its rival Keynote (+4, ranked 17), as it was the more popular collaborative presentational tool; “when we have to do work together we use Prezi…” [P38], and favoured for its functionality; “its brilliant… it makes presentations look flashy!” [P17], although both these tools were used infrequently; “I haven’t really needed to use it [Keynote] on a scale to produce my own PowerPoint’s” [P35].

Social Apps

Social apps were ranked high with social networking apps (Facebook & Twitter) ranked 2nd, due to the iPads performance; “I use Facebook quite a lot more on my iPad just because it runs a lot quicker” [P30], and the screen size; “I think the bigger screen is the main reason… especially when looking at friends photos…” [P4]. The interview generating the majority of
comments (13% compared to 1% prompts), as participants were questioned about their social usage. Additional favoured social apps were iPlayer (ranked 5); “one of the things I’ve enjoyed is the iPlayer apps, being able to put things on when I’m travelling” [P33] and the communicating app Skype (ranked 7); “It makes it a lot easier to contact people” [P25], which indicate the iPad was being used for pleasure related activities.

Clinical Medical Apps
Two of the speciality clinical medical apps, (BMJ) Best Practice and BNF (Nice Guidelines), received similar NV scores and frequency of comments (8/9%), with both being used as popular reference tool due to their unique content; “BMJ Learning has loads of information on it, you can find most things on there and they’ve got quite a lot of detail - so that’s quite useful if you’re going to look up a condition” [P31], and “the BNF app has been quite useful in terms of quickly finding out about the purpose and side effects of a drug” [P29]. BMJ Best Practice was favoured slightly more (NV=+55) over BNF (+53), where poor usability was forgiven; “the BNF Nice app is not that easy to use...it doesn’t flow very well but because the content is so good I still use it” [P36].

An additional four clinical apps received high NV scores, with Medscape scoring the highest (+31), due to its “extensive source of trusted medical information” which was “very useful on placement, as it doesn’t require Internet” [P24]. Whereas Almost a Doctor received a lower NV score (+18), due to its poor design; “is literary just streams of text in various boxes, but the content is important” [P7], with Patient.co.uk (+6) scoring the lowest as “as it doesn’t have the professional references... and the high level content that you get just by going to the website... so I don’t use that app as much” [P40]. The clinical game-based learning app Prognosis scored slightly better (+7), which “made the start of revision a little more fun!” [P19], by enhancing interactivity, that was “engaging and fun” [P42].

Dual-Purpose Apps
Ranked 6th, iBooks was the main repository app; “I use iBooks pretty much, it’s the main reason I use it [iPad] really, is for carrying around loads of books... its very useful to have lots of books and lots of resources” [P9], scoring a higher NV (+44), much higher than the similar rival app GoodReader (+28), despite the latter offering additional functionally; “Reading journals on GoodReader, you are able to highlight them and annotate, which is really good” [P25], whereas “in iBooks you cannot highlight, copy or edit any of the downloaded text” [P19]. However some participants found “GoodReader was disappointing, with too many features and a bit fiddly”, compared to iBooks, which was "more straightforward and usable" [P42]. In addition, iBooks provided a dual-purpose acting as a quick reference tool for; “reading on the go and for reference... which is very useful” [P35], and also support for leisure use; “I’ve got the clinical
handbook in there and then even personal stuff. I did a pantomime and we put the script into an iBook so everyone had a copy, so I like that." [P17], along with additional free iBook incentives; “I’ve now started to read fiction on it as well... I’ve only just realised you can get free books on it, I thought you had to pay for everything” [P48].

Another popular dual-purpose app was Dropbox (ranked 8th), scoring a high NV (+42), due to its unique functionality; “Dropbox has been great to transfer documents between my laptop at home, and on my iPad on the go” [P19], that acts as and essential a file transfer repository; “I’ve pretty much got everything revolving around Dropbox... I can access any file on my iPad, iPhone or laptop. More importantly, if I edit any file then it immediately updates on all interfaces... its invaluable” [P23]. Although more commonly associated with leisure use, YouTube (ranked 12) provided a dual-purpose app for learning; “using YouTube to watch skills video help me remember what to do when I am in the hospital environment” [P29], which offered an interactive alternative to learning; “some things are easier to learn by watching YouTube videos... I guess I’m more of a visual type learning, I do prefer learning from online lectures, as opposed to reading” [19].

**Mandatory Apps**

All three mandatory apps were ranked poorly. Numbers faired the best (ranked 14), with a low positive NV score (+8), due to its poor usability; “its really not user friendly... we got told Numbers would be a good app... its not straightforward, its not logical...its absolutely awful” [P19], and difficulty to learn; “I have started to get to grips with numbers a little more, and have been able to create a patient log... but I still have a few difficulties using numbers and have found myself accidently losing whole chunks of text” [P39].

The Blackboard app (-10) and UoM Forms (-11) were the least popular apps, with negative NV scores. The Blackboard app received negative ratings due to its poor usability; “Its not that great, I think I’m just better off using my laptop for that kind of stuff... its just not very user friendly” [P37]; with an increase in usage partly due to it containing “some really good videos” [P25], which could be due to participants not accessing the app, but using a web browser; “originally the app didn’t work, but then there’s a new link now...a link that then sends you to Safari” [P26], so the higher rating of Blackboard may not be related to the specific app.

The UoM eForms (used for clinical assessments) received comments related to its poor functionality; “the Manchester forms app... isn’t great because the boxes are tiny, so as soon as you start filling out your responses you can’t see half of what you’ve written, so you have to keep scrolling back up and down, so its really awkward” [P40]; and disliked for its mandatory use; “it was somewhat frustrating that we ‘have’ to use our iPads for this kind of feedback, suddenly I
have to learn how to use the new app" [P29]; and lack of advice; “there have been a lot of problems with eforms, and they had not really been explained to us and what impact it would have on our training” [P25].

6.3.7. Summary of Main Study Qualitative Findings

Overall participants gave more positively related comments (NV=+2239), reporting an equal amount of device and app comments (50% each), with motivators scoring the highest positive NV (+855), despite the negative barrier (-515), usability (-314) and negative affect (-318) comments, which participants seem to overcome, indicating that overall, participants were generally happy with the device, see Table 62.

<table>
<thead>
<tr>
<th>THEMES</th>
<th>DEVICE</th>
<th>APPS</th>
<th>Totals</th>
<th>Time Trends</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
<td>% Freq</td>
</tr>
<tr>
<td>Utility</td>
<td>-62</td>
<td>19%</td>
<td>+123</td>
<td>12%</td>
</tr>
<tr>
<td>Functionality</td>
<td>-80</td>
<td>30%</td>
<td>+289</td>
<td>39%</td>
</tr>
<tr>
<td>Usability</td>
<td>-80</td>
<td>12%</td>
<td>-234</td>
<td>20%</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>+30</td>
<td>6%</td>
<td>+26</td>
<td>2%</td>
</tr>
<tr>
<td>Usefulness</td>
<td>+110</td>
<td>23%</td>
<td>+178</td>
<td>17%</td>
</tr>
<tr>
<td>Ease-of-Use</td>
<td>+85</td>
<td>10%</td>
<td>+118</td>
<td>10%</td>
</tr>
<tr>
<td>Total (Quality)</td>
<td>+3</td>
<td>(830)</td>
<td>+500</td>
<td>(1194)</td>
</tr>
<tr>
<td></td>
<td>+440</td>
<td>62%</td>
<td>+415</td>
<td>80%</td>
</tr>
<tr>
<td></td>
<td>-500</td>
<td>38%</td>
<td>-15</td>
<td>20%</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>+25</td>
<td>35%</td>
<td>+156</td>
<td>65%</td>
</tr>
<tr>
<td>Negative Affect</td>
<td>-206</td>
<td>65%</td>
<td>-113</td>
<td>35%</td>
</tr>
<tr>
<td>Total (Affect)</td>
<td>-181</td>
<td>(265)</td>
<td>+43</td>
<td>(253)</td>
</tr>
<tr>
<td>Interactive</td>
<td>+6</td>
<td>23%</td>
<td>+31</td>
<td>10%</td>
</tr>
<tr>
<td>Communicate</td>
<td>-</td>
<td>1%</td>
<td>+33</td>
<td>10%</td>
</tr>
<tr>
<td>Video Media</td>
<td>+4</td>
<td>60%</td>
<td>+138</td>
<td>45%</td>
</tr>
<tr>
<td>Games</td>
<td>-10</td>
<td>16%</td>
<td>+82</td>
<td>35%</td>
</tr>
<tr>
<td>Total (Interactivity)</td>
<td>0</td>
<td>(150)</td>
<td>+284</td>
<td>(376)</td>
</tr>
<tr>
<td>TOTAL (Device &amp; App)</td>
<td>-238</td>
<td>50% (2621)</td>
<td>+1227</td>
<td>50% (2437)</td>
</tr>
<tr>
<td>Work Activities</td>
<td>+933</td>
<td>72%</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>Leisure Activities</td>
<td>+317</td>
<td>28%</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>Total (Activities)</td>
<td>+1250</td>
<td>(1728)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 62. Summary of the main qualitative findings across all themes*

In total the apps received a high NV score (+1227) compared to an overall negative device score (-238), despite both sharing half the frequency of comments (50%), clearly indicating that it was the apps contained on the device that facilitates positive experience.

All quality app ratings were positive, apart from usability, which received a higher NV score (-234) than device (-80), and higher frequency (12%/20%). This suggests that app usability problems were forgiven in favour for their specific functionality and content, as unusable apps could be easily replaced with similar rivals due to the wide variety of apps available. Negative device utility (-62) and functionality (-80) contrast positive app utility (+123) and functionality (+389), further confirming it was the apps that facilitated positive UX.
Usefulness showed a positive score for both apps (+178) and device (+110), with apps showing the advantage. The apps were considered marginally easier to use (+118), compared to device (+85); and both device and apps receiving similar positive aesthetic related comments.

Similar app (+415) and device (+440) motivator comments were reported, with device-centric comments relating to accessibility and portability, while positive app-related comments referred to repository and perceived utility. Device barrier comments (-500) were much higher than apps (+15), with the main contributors being connectivity (lack of internet) and context of use (hospital use). The users pre-disposition had an effect, where some participant's reluctance to change created a weak barrier to use, (as reflected in the Low Power Users); especially where habitual learning styles were preferred (e.g., reading, note-taking etc.), that was further accentuated by a lack of advice and guidance. Ordinary was more device-centric, relating to the initial novelty of owning the iPad, and only a weak barrier to use. The overall aggregate score (motivator and barrier) was positive (+340), indicating that despite the barriers, participants remained fairly motivated towards using their iPad.

Negative affect scored a higher NV (-318) than positive affect (+181), with the highest frequency of comments collected in the final (in-depth) interview when participants were prompted retrospectively to discuss their experiences (good and bad). The device received a higher NV score (-206) compared to the apps (-113), due to the frustration and annoyance of the continual lack of connectivity, the unsolvable device functionality problems (e.g., no central filling system, multi-tasking) and general usability problems. Positive affect comments were more app-related (+156), than device (+25), showing a decline over time (indicating an initial novelty phase), with comments relating to the enjoyment and satisfaction of using the apps.

Interactivity was more app-related, where specific apps enabled interactive activities such as watching/recording videos, playing games and communicating with friends. Participants had a divided opinion when using the interactive touch screen, with some participants enjoying using it, while others finding it frustrating. Speciality apps (such as YouTube & iPlayer) provided alternative engaging interactive learning experiences that encouraged motivation of use.

Overall all iPad activity comments decreased over time, with note-taking and playing games being the main contributor. Work activity received a much higher frequency of comments and NV score (72%, +933) than leisure (28%, +317), indicating the iPad was considered primarily a work-tool. The highest frequency of work activity comments were reported in
the first 3 months of usage, as users were excited to explore their new device, but over time a natural waning or ‘over familiarity’ occurred, when the device became more ordinary and less novel (as identified as a weak barrier to use). Both work and leisure apps received the highest frequency of activity comments, which further supports, that motivation of use was greatly facilitated by the apps contained on the device.

Successful apps were chosen for their unique functionality, dual-purpose and ease of use. Medical apps received the highest NV score (+170) as they provided reference support for clinical learning, where poor usability was forgiven in favour for their well-known brand name (BMJ and BNF) and trusted content. Social apps received a high NV score (+133), where specific apps provided support to social networking (Facebook), communication (Skype) and leisure use (iPlayer). Dual-purpose apps were successful as they supported both social and work use. Successful work-based apps (such as Notability) provided unique functionality that was unmatched by other apps, with less successful apps (such as Keynote) being least popular due to alternative rivals (Prezi). Mandatory apps (e.g., Numbers) received the only negative NV score (-13) and were rated poorly for usability, poor functionality and effort to learn. As these apps were necessary for clinical learning, participants were unable to use alternatives, which caused continual frustration, impacting on user satisfaction.

6.4. Non-Adopter Study

Both the quantitative and qualitative findings showed that all the participants within the main study were satisfied with using the iPad, and all adopted the new technology to varying levels. The Power User groups (identified in Chapter 5, Section 5.6) found that participants’ attitudes towards technology impacted on their judgement (ratings) of the iPad, yet over time all groups showed similar levels of use. Although the qualitative data revealed some aspect that may hinder iPad usage, (see Barriers 6.3.2.2.), all participants showed no signs of non-adoption and continued to utilise the iPad throughout the 6-month study. So the factors that may induce non-adoption remain unclear.

To investigate the reasons why users may have stopped using the iPad a small study was conducted that involved a new group of non-adopters. The term Non-Adopter refers to a user who has owned an iPad over some time, but has stopped using it.

6.4.1. Method

A total of 10 volunteer participants were recruited via an online advert placed on the Manchester Medical School VLE system (Medley) that requested participants who had owned
Chapter 6  Study Three: Longitudinal Study (Qualitative)

an iPad for over one year, but had stopped using their iPad. Participants were asked to complete a short online questionnaire and take part in an interview lasting around 30-40 minutes, in exchange for a £25 voucher incentive.

At the start and end of the interview participants were asked to verbally rate their experiences (on a 1-7 likert scale) for the same 8 aspects recorded in the longitudinal study (see Chapter 5 Table 39). At the start of the interview ratings were based on participants’ recollection of their expectations of use (just before they receive the iPad), and at the end of the interview, they were asked how they currently felt now (after a year of usage).

6.4.2. Participants

During the interview three participants stated that they were still motivated to use their iPad, so therefore they were excluded from the analysis (as they were not seen as pure non-adopters). The remaining 7 participants (3=female, 4=male) were aged between 18-25 and were in their 4th year of study. All participants had owned their iPad between 12 and 18 months, all owned a mobile phone (iPhone=4, Android=3) and a computer or laptop (PC=4, Mac=3), therefore brand influence was not considered a factor to non-adoption.

6.4.3. Quantitative Results

Ratings from an online questionnaire (based on a 1-7 Likert scale) collected participants’ response on four of the scales, as used in the main longitudinal study: Usability, Usefulness, Quality and Overall Experience (see Measures 5.2.4). An Independent T-test and the non-parametric Mann-Whitney U test were conducted to compare the aggregated Non-Adopter ratings with the aggregated Adopter ratings (from the longitudinal study reported in Chapter 5) for all four scales. Results from both test showed a significant difference for all four scales, see Table 63.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Mean</th>
<th>Standard Dev.</th>
<th>Independent T-Test</th>
<th>Mann-Whitney U-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adopters</td>
<td>Non-Adopters</td>
<td>Adopters</td>
<td>Non-Adopters</td>
</tr>
<tr>
<td>Usability</td>
<td>6.21</td>
<td>5.37</td>
<td>.590</td>
<td>.898</td>
</tr>
<tr>
<td>Usefulness</td>
<td>5.61</td>
<td>3.43</td>
<td>.759</td>
<td>1.106</td>
</tr>
<tr>
<td>Experience</td>
<td>6.26</td>
<td>4.29</td>
<td>.598</td>
<td>1.015</td>
</tr>
<tr>
<td>Quality</td>
<td>6.20</td>
<td>5.05</td>
<td>.503</td>
<td>1.146</td>
</tr>
</tbody>
</table>

Key: *p<0.05, **p<0.01, ***p<0.001

Table 63. Independent T-Test and Mann-Whitney U Test results comparing Adopters with Non-Adopter Ratings for Aggregated Usability, Usefulness, Overall Experience and Quality

The mean results for both groups (Adopter and Non-Adopter) and the percentage differences for each item was calculated, as shown in Figure 75.
The graph clearly shows Non-Adopters gave significantly ($p<.001$) lower Usefulness ratings (M=3.4) than Adopters (M=5.6), with an aggregated difference of 40%, with all items within the scale showing below average or average ratings. Overall Experience ratings also showed a significant ($p<.001$) difference of 32%, the second largest difference, with Non-Adopters ratings (M=4.3) much lower than Adopters (M=6.3). Single items showing the greatest differences were Motivated (36%) and Recommend (37%), suggesting that the Non-Adopters were less satisfied, and motivated to use their iPad. Non-Adopter ratings for Quality were also significantly ($p<.01$) lower (M=5.1), compared to Adopters (M=6.2), with a 24% difference, with Reliability (20%) and Trustworthy (22%) showing the largest differences. However the Adopters (M=6.2), gave significantly ($p<.05$) higher Usability ratings than the Non-Adopters (M=5.4), with single items Clear (21%) and Ease-of-Use (19%) showing the largest differences, which could be due to the Adopters solving usability problems after their continued use.

### 6.4.4. Qualitative Results

The interview data was analysed using the same content analysis method used for the longitudinal study (see 6.2.3). Comments were coded according the same five coding
frameworks, Quality (Device & Apps), Motivation & Barriers, Affect, Interactivity and Activities, with flexibility to allow for additional themes to emerge. Individual app analysis was not conducted, as the main objective was to focus on reasons for non-adoption.

A total of 210 Non-Adopter comments were coded across the five themes, which generated an overall negative NV score (-26), whereas the Adopters gained an overall positive NV score (+2239). To aid comparison between Non-Adopter and Adopters the results were normalised, by first calculating a score per person, and then multiplying this by 100 (so the NV scores become whole numbers), to aid comparison. To allow for the extra time difference for the Adopters (who were interviewed 3 times\(^1\) and contributed to the diary-study data); the Adopter results were divided by 3, so the normalisation calculations were as follows:

Non-Adopter: \((x + 7 \times 100\)\)

Adopters: \((x + 47) \div 3 \times 100\)

The un-normalised results and normalised results across all themes are shown in Table 64, with the normalised results will be used to direct the discussion. Clearly the Non-Adopter gave higher negative comments (NV=429) compared to the high positive Adopters score (+1588), as reflected in the actual results (Non-Adopters=-26, Adopters=+2504).

<table>
<thead>
<tr>
<th>Themes</th>
<th>Non-Adopters</th>
<th>Adopters</th>
<th>Normalised per P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>-10</td>
<td>+9</td>
<td>-143</td>
</tr>
<tr>
<td>Application</td>
<td>+9</td>
<td>+16</td>
<td>+129</td>
</tr>
<tr>
<td>Motivator</td>
<td>+7</td>
<td>+7</td>
<td>+100</td>
</tr>
<tr>
<td>Barrier</td>
<td>-26</td>
<td>-26</td>
<td>-429</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>+4</td>
<td>+3</td>
<td>+57</td>
</tr>
<tr>
<td>Negative Affect</td>
<td>-14</td>
<td>-318</td>
<td>-200</td>
</tr>
<tr>
<td>Interactivity</td>
<td>+5</td>
<td>+248</td>
<td>+71</td>
</tr>
<tr>
<td>Work Activity</td>
<td>-8</td>
<td>+933</td>
<td>-114</td>
</tr>
<tr>
<td>Leisure Activity</td>
<td>+7</td>
<td>+317</td>
<td>+100</td>
</tr>
<tr>
<td>TOTAL</td>
<td>-26</td>
<td>+2239</td>
<td>-429</td>
</tr>
</tbody>
</table>

Table 64. Summary of all Themes comparing the Actual and Normalised results (NV and % frequency)

The main differences found between the two groups was the Non-Adopters gave a much higher negative NV for Barriers (-429) compared to Adopters (-365); also the Non-Adopters gave negative scores for Device Quality (-143), and Work Activity (-114). Clearly the Non-Adopters gave more negative comments than Adopters. To understand the reasons why, each theme was investigated in order to understand why the Non-Adopters did not engage with their iPad.

\(^1\) Average Adopter time = 46 minutes, compared to Non-adopters=24 minutes
6.4.4.1. Quality (Device and Applications)

A total of 67 Non-Adopter comments were coded across the six sub Quality themes; Utility, Function, Usability, Aesthetics, Usefulness, and Ease of Use, which were split into either device or app related themes. Table 65 shows a comparison between the Non-Adopter and Adopter results, along with the calculated normalised results for both groups, which will be used to compare the findings.

<table>
<thead>
<tr>
<th></th>
<th>Non-Adopters</th>
<th>Adopters</th>
<th>Normalised per P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Device</td>
<td>App</td>
<td>Device</td>
</tr>
<tr>
<td></td>
<td>NV % Freq</td>
<td>NV % Freq</td>
<td>NV % Freq</td>
</tr>
<tr>
<td>Utility/Content</td>
<td>-3 15%</td>
<td>+4 12%</td>
<td>-62 19%</td>
</tr>
<tr>
<td>Functionality</td>
<td>-4 35%</td>
<td>+6 42%</td>
<td>-80 30%</td>
</tr>
<tr>
<td>Usability</td>
<td>-4 12%</td>
<td>-7 21%</td>
<td>-80 12%</td>
</tr>
<tr>
<td>Aesthetics</td>
<td>+1 15%</td>
<td>+1 3%</td>
<td>+30 6%</td>
</tr>
<tr>
<td>Usefulness</td>
<td>-4 12%</td>
<td>+4 18%</td>
<td>+110 23%</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>+4 12%</td>
<td>+1 3%</td>
<td>+85 10%</td>
</tr>
<tr>
<td>Total NPV</td>
<td>-10 (34)</td>
<td>+9 (33)</td>
<td>+3 (830)</td>
</tr>
</tbody>
</table>

Table 65. The 6 Quality Rating themes comparing Actual & Normalised results, for Device and Apps

The Non-Adopters Quality comments were equally split between Device (51%) and App (49%), and received an overall negative NV (-15); whereas the Adopters generated more app related comments (59%) compared to device (41%), and scored a much higher positive NV score (+355). So clearly the Adopters gave more favourable comments than the Non-Adopters. The reason can be attributed to the Non-Adopters overall negative score for Device (-143), which was much worse than the Adopters (+1); furthermore, although the Non-Adopters gave a positive apps score (+128), it was much lower than the Adopters (+354). This suggests that both the device and the apps contributed to Non-Adopter dissatisfaction.

Although both groups received similar ratings for the device themes Utility (NV=-43/-44), Functionality (-57), Usability (-57), and Ease of Use (+57/+60), the main difference shown was for Usefulness, where Non-Adopters gave a negative NV score (-57) compared to the Adopters positive score (+78). Despite both groups reporting the device was easy-to-use, they both shared the same negative device comments for Utility (e.g., battery life), Functionality (lack of multi-tasking) and Usability (software crashes). Clearly the main difference was that the Non-Adopters did not find the device as Useful as the Adopters, as outlined here: “I just wasn’t sure what it’s meant to do - I wasn’t sure what it added... I didn’t find it useful for me personally” [P10], with the main reason being due to the lack of appropriate training; “it wasn’t exactly explained how to use it...if everybody had received a bit of training it would have been much more useful” [P5].
For the app ratings, the Non-Adopters gave lower ratings for *Utility* (+57), *Functionality* (+86) and *Usefulness* (+57) compared to the Adopters (*Utility*=+87, *Function*=+205, *Useful*=+126). The Non-Adopter comments reflected on their lack of motivation in using the apps; “I felt there wasn’t enough things [apps] for it, and it just became an added burden… an added object that you don’t really need” [7], as the apps were found not that useful; “In the beginning I used it a lot… but I found I didn’t need it…I thought the apps would be useful here and there… and the apps weren’t as useful as I thought they’d be” [P8]. The main reasons can be attributed to the lack of understanding of what apps to use; “If I had grasped the basics of knowing which apps to download and how to use them…. when you not sure of the basics, your less motivated to investigate more” [P6]. This ultimately impacted on how useful the device was considered; “I thought that the potential of it is really big, there’s going to be loads of apps and books… but then I didn’t really know what to do with it to be honest” [P7], along with a lack of understating of how the iPad could be of benefit; “work-wise I wasn’t sure what I could do with it… I wasn’t sure what it added to me… I didn’t find it useful for me personally” [P10]. So a poor perceived utility coupled with the lack of training at the appropriate time (within the first few months of ownership) impacted on motivation (which was low), appeared to be the main reasons for non- adoption of the iPad.

It is not surprising that Non-Adopter app ratings for *Usability* were slightly better (-100), than the Adopters (-166), as they were less exposed to using their iPad, so therefore experienced fewer usability problems compared to the Adopters. Comments relating to specific app *Usability* problems, where, “Evernote frequently crashed”, which resulted in “losing a fair amount of notes” [P4], and the mandatory UoM EForms app “which is an absolute disaster” as it is “really difficult to use” [P6] as “half the forms disappeared” [P5]. Interestingly, participants reported their “frustration” [P10] of having to use this mandatory app (which was necessary to gain feedback after completing a hospital placement), calling it a “forced fun”, as they felt they were being “forced to use them…without being explained” [P6].

Although both groups gave similar ease of use scores for device (+57/+60), the Non-Adopters gave a lower ease-of-use apps score (+14) compared to Adopters (+84). Comments report on learnability of unfamiliar apps, “I felt like new apps, such as Evernote and Notability, getting used to the functions and things like that were more difficult” [P6], may reflect reasons for non-adoption. Aesthetic comments for both groups (across device and apps) were similar, with general comments reported; “it still looks attractive… it’s well made and feels pretty robust” [P1].
6.4.4.2. Motivations and Barriers

A total of 65 comments were coded according to either Motivator or Barrier related themes, as identified in the main study. Both themes were split into four sub-themes, and the results were normalised in order to compare the two groups. The Adopters generated more motivator comments (60%), then Non-Adopters (45%), and scored an overall higher NV (+607) compared to Non-Adopters (+100), see Table 66 for all Motivator results.

<table>
<thead>
<tr>
<th>MOTIVATORS</th>
<th>Non-Adopters Device NV</th>
<th>Non-Adopters Apps % Freq</th>
<th>Adopters Device NV</th>
<th>Adopters Apps % Freq</th>
<th>Normalised per Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repository</td>
<td>0</td>
<td>-</td>
<td>+31</td>
<td>4%</td>
<td>0</td>
</tr>
<tr>
<td>Portability</td>
<td>+7</td>
<td>29%</td>
<td>+137</td>
<td>19%</td>
<td>+100</td>
</tr>
<tr>
<td>Immediacy</td>
<td>+6</td>
<td>25%</td>
<td>+281</td>
<td>39%</td>
<td>+86</td>
</tr>
<tr>
<td>Perceived Utility</td>
<td>-11</td>
<td>46%</td>
<td>-9</td>
<td>38%</td>
<td>-157</td>
</tr>
<tr>
<td>Total NPV</td>
<td>+2</td>
<td>(24)</td>
<td>+440</td>
<td>(736)</td>
<td>+29</td>
</tr>
</tbody>
</table>

Table 66. The four Motivator sub-theme comparing Adopter and Non-Adopter results

The Non-Adopters generated a higher frequency of device motivator comments (83%), than Adopters (62%), but scored a much lower NV score for both device (+29) and apps (+71) than Adopters (+312/+295), with device comments showing the weakest. Despite Portability (+100) and Immediacy (+86), showing a strong device-related Non-Adopter contribution, Perceived Utility received a negative NV (-157), generating a high frequency of comments (46%), indicating it was more of a barrier to use rather than a motivator. Unlike the Adopters whose comments related to the benefits of app utility (to support learning), the Non-Adopters comments were more device-centric, and reported the “frustrating” [P8] experience of using the touch screen for typing, as “the keyboard was difficult to get used to” [P6].

Overall the Non-Adopters generated a slightly higher overall NV for Barrier comments (-429) than Adopters (-367), a higher frequency of device Barrier comments (72%) and NV score (-314) compared to the Adopters (62%, -356), see Table 67. However the Non-Adopter scored a much higher negative NV (-115) of App Barrier comments than Adopters (-11), indicating the apps were the stronger barrier to use than the device.

<table>
<thead>
<tr>
<th>BARRIERS</th>
<th>Non-Adopters Device NV</th>
<th>Non-Adopters Apps % Freq</th>
<th>Adopters Device NV</th>
<th>Adopters Apps % Freq</th>
<th>Normalised per Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connectivity</td>
<td>-4</td>
<td>16%</td>
<td>-183</td>
<td>30%</td>
<td>-57</td>
</tr>
<tr>
<td>Context of Use</td>
<td>-10</td>
<td>27%</td>
<td>-194</td>
<td>38%</td>
<td>-114</td>
</tr>
<tr>
<td>Ordinarily</td>
<td>-3</td>
<td>8%</td>
<td>-77</td>
<td>14%</td>
<td>-43</td>
</tr>
<tr>
<td>Reluctance to Change</td>
<td>-5</td>
<td>19%</td>
<td>-44</td>
<td>18%</td>
<td>-71</td>
</tr>
<tr>
<td>Total NPV</td>
<td>-22</td>
<td>(26)</td>
<td>-500</td>
<td>(640)</td>
<td>-314</td>
</tr>
</tbody>
</table>

Table 67. The four Barrier sub-theme comparing Adopter and Non-Adopter results
Reluctance to Change was the strongest Barriers to use for the Non-Adopters, with comments being split between device (-71) and apps (-86), and scoring much higher than the Adopters (-33/-2). The Non-Adopter comments reported a reluctance to change from their individual learning style, “I had my way of learning already set, and this [the iPad] was a bit of a strange addition” [P8], which was closely linked to their individual learning preferences to “keep written notes” [P6], or to “rather hold a physical book...as it feels substantial” [P10]. This was also influenced by the Non-Adopters lack of understanding of how best to use the iPad: “I wasn’t sure what to do with it work wise, I tried putting my notes and books on it, but then realised that I would rather hold the book... because I prefer paper myself” [P10] and, “when you’re doing it on your own it just feels like a constant battle, and you don’t know if you’re doing it right”[P6]. This was further hindered by a lack of advice; “I felt there was no way for me to figure out how to do that myself, no guidance. I was really unsure” [P6], along with a generally lack of acceptance; “I just personally don’t like tablets that much. I think I’ve learnt that over a year of trying to use it and it not really working out that well” [P5]. Therefore, due to the Non-Adopters reluctance to change their learning style coupled with a lack of advice, their perceived utility and usefulness of the iPad decreased, as it was considered no benefit to their learning.

Both groups shared similar comments for Context-of-use, (Non-Adopters: NV-143, Adopters: NV= -138), which was more device centric, with comments relating to the problems of using it within clinical settings; “I personally found it difficult to take onto wards because I couldn’t take notes, and its much easier to just take out your notebook and scribble down a few notes” [P5], and also how it was viewed by patients; “as patients don’t really take to it well, I feel like its so much more of a barrier than a piece of paper” [P6], and how it was perceived; “because you’ve got this kind of technological device and they don’t really know what your doing on it, because people use it as a personal thing... and can look unprofessional” [P28]. The Adopters reported more negative Connectivity comments (NV=-130) compared to Non-Adopters (-57), which could be due to the Adopters using their iPad for more varied activities, so were exposed more to these problems. Comments for Ordinary showed little difference between groups, and were more device-centric (due to a familiarisation effect), with comments similar to those found in the longitudinal study.

The impact of alternative devices also impinged iPad usage, especially when participants updated their mobile phones; “when I got my iPhone it took me away from the iPad...its just much easier to get my phone out. It just fits in my pocket...its just there” [P1], its “more convenient” [P4], along with individual preference; “I still prefer having my laptop, its just easier to make notes and manage things” [P5]. So although not a barrier to use, comparative
devices hindered product adoption; “it’s an extra device that I didn’t particularly need, because I already had other devices which can do what it can” [P10].

6.4.4.3. Positive and Negative Affect

Only 40 Affect related comments were collected from the Non-Adopter interviews, with Negative Affect receiving a higher frequency of comments (65%) compared to the Adopters (54%), along with a higher negative NV score overall (-143/-98), so overall the Non-Adopters gave more negative affective comments. Although the Adopters scored a higher Negative Affect NV (-226), than the Non-Adopters (-200), which was offset by a higher Positive Affect score (+128) compared to Non-Adopters (+57), see Table 68. Clearly the Non-Adopters gave more negative affective comments than positive, but as they invested much less time interacting with the device, their emotional responses may be less marked then the Adopters, who despite giving high positive comments, gave higher negative affect comments, especially due to initial frustrations occurring in the first few months of usage.

<table>
<thead>
<tr>
<th></th>
<th>Non-Adopters</th>
<th>Adopters</th>
<th>Normalised per Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Device NV %</td>
<td>Apps NV %</td>
<td>Device NV %</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>+1 20%</td>
<td>+3 80%</td>
<td>+25 35%</td>
</tr>
<tr>
<td>Negative Affect</td>
<td>-9 65%</td>
<td>-5 35%</td>
<td>-206 65%</td>
</tr>
<tr>
<td>Total NPV</td>
<td>-8 (35)</td>
<td>-2 (5)</td>
<td>-181 (265)</td>
</tr>
</tbody>
</table>

**Table 68. The Affective (Positive/Negative) comparing Non-Adopters and Adopters results**

Negative Affect comments for device were higher for both Non-Adopters (-129), and Adopters (-146), than for the apps (-71/-80), with little difference between the two groups (see Table 68). Negative device-related affective comments were due to the frustration of touch typing; “it’s not something for typing, it can be frustrating”[P8], and lack of internet; “the Wi-Fi thing, not gaining access, that’s defiantly frustrating sometimes”[P9]. While negative app related comments related to the mandatory use of the iPad; “We had to use it in the hospital setting to get signed off at the end of every semester using e-forms...I found that quite frustrating,” [P10, and lack of advice; “some application I tried to download and they wouldn’t download, and there was no explanation for it, which was slightly frustrating” [P3].

Both groups gave similar Positive Affective comment scores for device (+14/+18), but the Adopters gave a much higher positive NV for apps (+110), then the Non-Adopters (+43), where Non-Adopter comments relating to mainly leisure apps; “I really enjoyed the Skype app, and iPlayer, being able to catch up on what I missed” [P10].
6.4.4.4. Interactivity

Only a total of 24 Interactivity comments were collected from the Non-Adopters (compared to the 523 Adopter comments). The Non-Adopter comments were also coded according to the four Interactivity sub-themes (see Table 69), and all comments were then normalised to aid comparison.

<table>
<thead>
<tr>
<th>Non-Adopters</th>
<th>Adopters</th>
<th>Normalise per Participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>Apps</td>
<td>Device</td>
</tr>
<tr>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
</tr>
<tr>
<td>Interactive</td>
<td>-5</td>
<td>83%</td>
</tr>
<tr>
<td>Communicate</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Video Media</td>
<td>+1</td>
<td>17%</td>
</tr>
<tr>
<td>Games</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>TOTALS</td>
<td>-4</td>
<td>(6)</td>
</tr>
</tbody>
</table>

Table 69. The Interactivity results comparing Adopters and Non-Adopters results.

The Adopters gave a much higher NV score (+201) than the Non-Adopters (+71), indicating that interactivity may contribute to satisfaction. When combining both groups the apps scored a much higher NV (+329) compared to a negative device score (-57), which clearly shows that it is the apps that provided most of the interactive experiences. The Non-Adopters gave more negative comments (-57), which mainly related to the touch screen interface, “it can be frustrating, stuff like the touch screen... its not something for typing” [P8]. However, both groups enjoyed watching Video Media and using the iPad to Communicate through speciality apps (such as iPlayer, YouTube, Skype etc.), with the Adopters using their iPads more to record and watch videos as an alternative learning method.

6.4.4.5. Learning and Leisure Activity

A total of 35 Non-Adopter activity related comments were coded into either Work or Leisure Activities, with work receiving a slightly lower frequency (57%) than the Adopters (72%), see Table 70. Overall the Adopters scored a much higher NV (+887) compared to the Non-Adopters (-14), thus reflecting the lack of interaction with the iPad.

<table>
<thead>
<tr>
<th></th>
<th>Non-Adopters</th>
<th>Adopters</th>
<th>Normalised per P</th>
</tr>
</thead>
<tbody>
<tr>
<td>NV</td>
<td>% Freq</td>
<td>NV</td>
<td>% Freq</td>
</tr>
<tr>
<td>Work Activity</td>
<td>-8</td>
<td>57%</td>
<td>+933</td>
</tr>
<tr>
<td>Leisure Activity</td>
<td>+7</td>
<td>43%</td>
<td>+317</td>
</tr>
<tr>
<td>Total NPV</td>
<td>-1</td>
<td>(35)</td>
<td>+1250</td>
</tr>
</tbody>
</table>

Table 70. The Activity (Work and Leisure) themes comparing Adopters and Non-Adopters results

The main difference for the Non-Adopters was reflected in the negative NV (-114) for Work Activity comments, compared to the Adopters positive result (+662), with Leisure Activity
showing a positive NV (+100). This suggests the Non-Adopters were using their iPads mainly for leisure activity; “although I did try and use it to take notes and stuff, that was a one off thing, it was mainly used for leisure and social networking” [P4], and “the only reason I use it is to either play games or when I’m in bed to go on Facebook… or watching YouTube videos, reading the news” [P10]; while it was “rarely used” [P4], for learning.

The negative Non-Adopter NV scores referred to Note-taking comments (-71) and Reading (-86), with comments relating to speed of typing; “I starting trying to make notes with it using Notability… but over time I went back to writing notes, just because it was quicker” [P6], and learning style; “I wasn’t much of a note-writing person” [4], along with individual preference; “I’m not really partial to iBooks, I prefer having a book in my hand. I think it’s just a different feel from reading” [5]. While the only negative NV rating for Leisure Activity given by the Non-Adopters was for games (-43), with comments relating to their novelty; “I played a few games at the start, but then that kind of just died out” [P10].

Non-Adopters reported disengaging with their iPad after around 2-3 months of usage; “we were given it early October, by November I’d stopped bring it in, and by January my little brother took control of it” [P10], and “in the beginning I used it a lot…we were given it at the end of September and just before Christmas I found I didn’t need it…” [P8]. Most participants reported little motivation to continue to use their iPad, “realistically I don’t see myself using this much” [P4], “right now its not even with me, its in London at home, I’ve not used it at all this year” [P8]. Therefore, for the Non-Adopters, the first few months of owning the iPad were crucial, where lack of appropriate training as to how best make use of the device and apps appear to be a contributory factor in their rejection of the iPad.

6.4.5. Non-Adopter Verbal Ratings Results

The verbal ratings compared participants’ Expected Use (pre-study) with Actual Use (current experience) for both Non-Adopters and Adopters, as shown in Figure 76. The graphs show the mean scores from both groups, and clearly shows the Non-Adopters ratings over time dropped considerable more compared to the Adopters.
Comparing the overall means, initially the Adopters (M=5.8) gave slightly higher ratings Pre-Use than the Non-Adopters (M=5.6), with the greatest % difference between the two groups seen in Engaging (14%), Boring (6%) and Attractive (7%), see Table 71. Interestingly, there was no difference between the groups for Usefulness (Pre-Use), and they both gave similar Satisfaction ratings (with only 4% difference). So initially the Non-Adopters perceived the iPad to be less engaging, more boring and was less attractive than the Non-Adopters, but no difference was shown in perceived usefulness and utility. However over time the gap widens from Pre-Use (4% difference) to Current-Use (24%), with the Adopters ratings showing little change over time (M=5.8 to M=5.9), compared to the Non-Adopters, who showed a marked drop (M=5.6 to M=4.5), thus reflecting their disengagement. All Non-Adopter ratings for Current-Use showed a 23% to 33% difference (apart from Frustrating and Ease-of-Use), with the greatest diversity shown in ratings for Attractive (33%), Useful (31%), Engaging (29%) and Enjoyable (30%), clearly illustrating the Non-Adopters loss of interest in using the iPad. Surprisingly, Frustrating showed only a slight difference (9%), as both groups shared similar frustrating experiences, while Ease of Use (7%) showed the lowest difference with both group ratings increasing over time.
A Mann-Whitney test was used to compare the Expected and Actual Use for both groups, and Non-Adopters found significant (p > .05) differences for Attractive (Z=-2.219) and Enjoyable (Z=-2.300), while the Adopters showed significant (p > .001) differences for Ease-of-use (Z=-3.958), Attractive (Z=-3.556), and (p > .01) for Usefulness (Z=-2.271). Both groups found the iPad less Attractive over time, which could be due to a familiarisation effect, however the Non-Adopters enjoyed using the device significantly less, while the Adopters found the iPad more Easier-to-Use and continually more Useful over prolonged use and time.

To compare the two group ratings (Non-Adopter with Adopter), an Independent T-test and the non-parametric Mann-Whitney U test were conducted on the verbal ratings for Expected and Actual Use. Results found little difference between groups for Expected Use, apart from Attractive (p < .05) showing a weak significance, indicating the Adopters found the iPad more attractive prior to use, see Table 72. All items (apart from Ease-of-use), were found significant (p < .05 to p < .001) after Actual Use, with the Non-Adopters giving consistently lower ratings than Adopters.

<table>
<thead>
<tr>
<th></th>
<th>Expected Use</th>
<th></th>
<th></th>
<th>Actual Use</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Adopter</td>
<td>Non-Adopter</td>
<td>t-Value</td>
<td>Sig.</td>
<td>Z-Score</td>
</tr>
<tr>
<td>Ease-of-Use</td>
<td>5.41</td>
<td>5.57</td>
<td>-2.29</td>
<td>NS</td>
<td>-4.63</td>
<td>NS</td>
</tr>
<tr>
<td>Attractive</td>
<td>6.39</td>
<td>5.86</td>
<td>1.307</td>
<td>NS</td>
<td>-2.57</td>
<td>*</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>5.86</td>
<td>5.57</td>
<td>.410</td>
<td>NS</td>
<td>-102</td>
<td>NS</td>
</tr>
<tr>
<td>Enjoy</td>
<td>5.82</td>
<td>6.00</td>
<td>-4.13</td>
<td>NS</td>
<td>-778</td>
<td>NS</td>
</tr>
<tr>
<td>Useful</td>
<td>5.41</td>
<td>5.43</td>
<td>-0.40</td>
<td>NS</td>
<td>5.43</td>
<td>*</td>
</tr>
<tr>
<td>Engaging</td>
<td>5.31</td>
<td>4.71</td>
<td>1.410</td>
<td>NS</td>
<td>-1.363</td>
<td>NS</td>
</tr>
<tr>
<td>Frustrating</td>
<td>5.74</td>
<td>5.57</td>
<td>.281</td>
<td>NS</td>
<td>-5.43</td>
<td>*</td>
</tr>
<tr>
<td>Boring</td>
<td>6.47</td>
<td>5.86</td>
<td>1.558</td>
<td>NS</td>
<td>-1.285</td>
<td>NS</td>
</tr>
</tbody>
</table>

Key: *p<0.05, **p<0.01, ***p<0.001

Table 72. Means and Independent T-Test and Mann-Whitney U-Test Sig Results
(Comparing Expected with Actual Use Results)

### 6.4.6. Brand and User Disposition

Possible reasons for the Non-Adopters loss of interest (over time) could be brand preference or the users personal disposition. The Non-Adopters were also asked to verbally rate the Brand Quality and overall they rated it above average (M=5.6), which was equivalent to the Adopters overall rating (M=5.5), clearly showing that brand had not influenced their decision.

The participants completed the Power User Scale and results for each Non-Adopter participant was compared with the Adopter sample group mean. The 7 Non-Adopter participants were split across the groups as follows: High Power Users: (N=2, M=5.6-5.9), Medium Power Users: (N=3, M=4.2-4.9), and Low Power Users: (N=2, M=2.1-3.3), so technical disposition was not an influential factor for Non-Adopters.
6.4.7. **Summary of Non-Adopter Results**

It is clear from the findings that the Non-Adopters were less satisfied with using the iPad compared to the Adopters, with significant decreases in the majority of ratings after continued use, see Table 73. Also, the Non-Adopters generated an overall negative NV comment score (-430), compared to the Adopters positive score (+1585) across all the qualitative themes. The contributing factors were negative comments towards the Device Quality (NV=15) and Barriers (NV=-429), which were much lower than the Adopters Device Quality (NV+355) and Barrier comments (NV-367).

The main reasons for negative device Quality comments was the Non-Adopters did not find the device as useful as the Adopters, as they reported more negative Usefulness comments (-57/+78), as confirmed by both the questionnaire ratings (40% difference), and the verbal ratings (31% difference) between groups. Although the Non-Adopter reported fewer Usability comments than the Adopters, they showed only a slight (13%) difference in questionnaire ratings. This can be attributed to the Non-Adopters interacting less frequently with the device, so therefore they experienced fewer problems. In addition, the Non-Adopters gave lower Ease-of-Use comments and showed a 19% difference in questionnaire ratings, indicating that the Non-Adopters did not find using the iPad as easy to use as the Adopters.

<table>
<thead>
<tr>
<th>MAIN THEMES</th>
<th>Non-Adopters</th>
<th>Adopters</th>
<th>TOTAL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Device</td>
<td>App</td>
<td></td>
<td>Non-Ad</td>
</tr>
<tr>
<td>Quality (Device &amp; Apps)</td>
<td>-143</td>
<td>+128</td>
<td></td>
<td>-15</td>
</tr>
<tr>
<td>Motivators</td>
<td>+29</td>
<td>+71</td>
<td></td>
<td>+100</td>
</tr>
<tr>
<td>Barriers</td>
<td>-314</td>
<td>-115</td>
<td></td>
<td>-429</td>
</tr>
<tr>
<td>Positive Affect</td>
<td>+14</td>
<td>+43</td>
<td></td>
<td>57</td>
</tr>
<tr>
<td>Negative Affect</td>
<td>-129</td>
<td>-71</td>
<td></td>
<td>-200</td>
</tr>
<tr>
<td>Interactivity</td>
<td>-57</td>
<td>+128</td>
<td></td>
<td>+71</td>
</tr>
<tr>
<td>Work Activities</td>
<td></td>
<td></td>
<td></td>
<td>-114</td>
</tr>
<tr>
<td>Leisure Activities</td>
<td></td>
<td></td>
<td></td>
<td>+100</td>
</tr>
<tr>
<td><strong>Grand Total Normalised Results</strong></td>
<td>-430</td>
<td>+1585</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 73. Summary of the Main Qualitative themes comparing Non-Adopters & Adopters normalised results*

Although the Non-Adopters gave positive app Quality comments overall (+128), it was much lower compared to the Adopters (+354), with low ratings for Functionality, Usability, Usefulness and Utility. Negative comments referred to the learnability of unfamiliar apps, frustration at having to use mandatory apps, and the lack of understanding about which apps to use in order to provide a benefit. The need for appropriate training (in the first few weeks of ownership) to advise how to use the device and apps in order to benefit learning may have improved technology adoption. Favourable Aesthetic comments were reported, with little difference between the groups, although verbal ratings for Attractive showed the greatest difference (33%). This concurs with the questionnaire ratings for Quality, which shows a
24% difference, so the Non-Adopters rated the iPad less attractive and not as high quality than the Adopters.

Clearly the Adopters were better Motivated (+607) than the Non-Adopters (+100), with more comments reported, along with the questionnaire ratings showed a 36% difference, where device (+29) showing a weaker contribution compared to apps (+71). Although Immediacy and Portability were the most prominent motivators of use, for Non-Adopters it was the Perceived Utility of the device that proved to be more of a barrier, rather than a motivator, with comments reporting the frustration of using the device to support note taking and reading.

Unsurprisingly the Non-Adopters reported more Barrier related comments (-429) than Adopters (-367), with the main difference being their Reluctance to Change, which refers to the Non-Adopter aversion to change their set learning style, where the lack of advice and guidance further hindering the process of adoption. This is closely linked to the Perceived Utility of the iPad, which was not strong enough to support various learning activities (note-taking, reading), and also was influenced by the competition of alternative devices.

Both groups gave similar Negative Affective comments (-200/-226), although the Non-Adopters gave less Positive Affective comments (+57) compared to the Adopters (+128), further supporting their lack of engagement. Interestingly, in total (Positive + Negative Affect), both groups gave more negative comments (-143/-98), indicating negative affect was more likely reported than positive as this was more salient in memory. Also, the Non-Adopters would have invested less time using the iPad, so their emotional responses were less likely to be as intense as the Adopters. Both groups gave higher Negative Affect comments for Device, due to their frustration of touch typing, lack of Wi-Fi, and the mandatory use of the iPad (and apps), accentuated by the lack of guidance and training.

The Adopters gave more positive Interactivity comments (+201) than the Non-Adopters (+71) with the apps (+329) showing the difference (device -57), as the apps enabled more positive interactive experiences, such as watching videos, or social communication, while device interactively remained invisible. The Non-Adopters reported negative comments for the device Interactivity, relating mainly to the touch-screen interface (further supporting their reluctance to adapt to new technology). Clearly the Adopters were using the iPad much more for work (+662) than leisure (+225), compared to the Non-Adopters, who favoured it more for leisure (+100), reporting more negative comments for work (NV=-114), due to their negative perception of its usefulness to support work activity. Non-Adopters typically disengaged within 2-3 months of usage, during which leisure activities were more prominent.
6.5. Chapter Summary

The qualitative findings from the main longitudinal study showed that all participants were generally satisfied with the iPad, where the apps provided the advantage. Despite usability problems, initial frustrations were resolved through work-around solutions, or forgiven due to the variety of apps available, and the functionality and content they offered, thus making the iPad a useful device for both learning and leisure use.

All participants in the longitudinal study adopted the iPad, so a small group of Non-Adopters were interviewed to explore the reasons behind non-adoptions. The qualitative comments were compared across the two groups (Adopters and Non-Adopters) to identify any differences. The findings clearly showed that the Non-Adopters were less satisfied with the iPad, giving more negative comments (NV=430) compared to the Adopters, who gave far more positive comments (NV=+1588). This was also reflected in the verbal ratings, as nearly all Adopter verbal ratings significantly increased after continued use, apart from ease-of-use, which increased for both groups. To outline the reasons behind non-adoption, two models (see Figure 77) drawn from the qualitative findings illustrate the main differences between the Adopters and Non-Adopters, and will provide the basis for the summary discussion.

![Figure 77. Summary Model of Adopters (left) and Non-Adopters (right) qualitative results](image)

Each model shows the main groups split between application and device, and the arrows show the strength of NV ratings, which is either positive or negative. The main differences between the two group models, is how Useful the iPad was and the strength of the Motivators and Barriers of use.
Clearly the *Adopters* found the iPad much more *Useful* than the *Non-Adopters*, generating far more positive comments, and showing a significant increase in the questionnaire and verbal (*usefulness*) ratings over time (31% increase). Consequently, the *Adopters* were more *Motivated* to use their iPad, as they reported a higher *Perceived Utility* and gave more *Positive Affective* comments (compared to the *Non-Adopters*), due to the *Usefulness* of the apps to support learning-based activities. Although both groups reported similar *Negative Affective* comments (due to *Connectivity, Usability* and *Functionality*), similar *Barriers* comments were reported (e.g., *Context-of-Use*), but overall the *Non-Adopters* gave slightly more *Barrier* comments than the *Adopters*. This was due to their negative *Perceived Utility* of the apps, as they felt they did not enhance their learning activities (e.g., note-taking and reading), with negative app-related comments referring to learnability of unfamiliar apps and the frustration at having to use mandatory apps. This was further escalated by the *Reluctance to Change* their learning style (another *Non-Adopter* barrier), further hindered by a lack of early advice about how best they could utilise the iPad, which ultimately lead to non-adoption.

Conversely, the *Adopters* gave more positive app-related comments than the *Non-Adopters*, due to their *Usefulness*, where poor *Usability* was tolerated, in favour for their content or specific functionality that supported learning. Although both groups enjoyed the communication apps that enabled social interaction, the *Adopters* were motivated by specific apps that facilitated interactive learning, such as video media, which contributed to their overall product satisfaction. Clearly the *Adopters* interacted far more with their iPad than the *Non-Adopters*, which could account for their higher frequency of app-usability comments, causing them continual frustration. Similarly, the *Adopters* gave more device-usability comments, (although less prominent), despite rating the device significantly more usable than the *Non-Adopters* in the questionnaire ratings. This suggests that the questionnaires captured more general attitudes of the iPad, while the qualitative comments provided deeper insights at the app/device level.

Based on the *Adopter* ratings, far more positive app-related comments were generated than device, indicating that it was the apps contained on the device that facilitated positive UX. The more popular apps were chosen for their dual-purpose (e.g., iBooks) or specific functionality (e.g., Facebook) to support a variety of activities (e.g., Notability) that could be rivalled by no other. Speciality medical apps were favoured for their trusted content and brand name, where poor usability was forgiven, while mandatory apps necessary for clinical learning were commented upon negatively due to their poor usability and effort to learn.

A general drop in *Adopter* activity comments can be attributed to a familiarisation or novelty effect, as overtime the device appeared more ordinary (identified as a weak barrier). The *Adopters* reporting more work-based activities, indicating they viewed it more as a work-
based tool, while the Non-Adopters reported only favourable leisure-based activities. Both groups gave good Aesthetic ratings, however the Non-Adopters rated the iPad less Attractive (33% verbal ratings), and lower Quality (questionnaire ratings) than the Adopters, possible due to their perceived lack of utility. The Non-Adopters reported disengaging fairly quickly after acquisition (2-3 months), where the most salient frustrations (Negative Affect) related to their forced use of mandatory apps (e.g. UoM eForms) for clinical learning, accentuated by the lack of guidance at how best the iPad can be used to benefit learning.

In summary, the Adopters were generally satisfied with using the iPad, due to the Usefulness of the apps, where device Usability was good enough and app problems were tolerated in favor for their function, content and interactivity. High positive app comments far outweighed device comments, indicating that positive UX was mainly app-related. The Motivators of use were a composite of both device and apps, with Portability and Immediacy being more device-centric, while Repository (e.g., iBooks) and Perceived Utility (e.g., Notability) related more to app functionality. The Adopters positive perception of the utility and usefulness of the apps to support both learning and leisure activity motivated further usage, while the Non-Adopters perception of app utility was negative, and became a barrier to use. In addition, the Non-Adopter reluctance to change their learning style was hindered by the lack of early advice and guidance at how best to utilize the iPad to support learning. Poor perception of the utility and usefulness of the iPad (especially to support learning), were the main reasons for non-adoption.

The Adopters showed an initial exploration phase (2-3 months) where various apps and activities were tested, and overtime all activities and affective responses reduced indicating a familiarity or novelty effect. Whereas the Non-Adopters reported disengaging with the iPad in 2-3 months or ownership, where initial novelty quickly waned, and unlike the Adopters whose perceived utility and usefulness increased over time, the Non-Adopters found the iPad ineffective as a learning support, therefore use declined and the product was rejected.

This chapter set out to answer some unanswered questions resulting from the quantitative findings, as outlined at the end of Chapter 5. Clearly PQ ratings increased over time, despite poor usability, where initial frustrations were overcome through work-around solutions, due to the variety of apps available, indicating the apps were more important than the device in generating positive UX. Although over time only a few apps made the difference, due to their utility, usefulness and ability to provide interactive experiences. Differences were found in users’ attitudes towards technology (power users), where more adverse comments were given by negative IT users. However, technical disposition was not the main contributory factor to non-use. It was users’ reluctance to change, along with their negative perception of usefulness and utility of the apps, which resulted in non-adoption.
Chapter 7 Main Discussion

This chapter discusses the main contributions of this thesis and its implication for researchers and practitioners within the HCI field. It investigates the results of the three main studies (reported in Chapter 3 to 6) to present an extended model of user experience. The findings are evaluated in light of the research constraints, along with the lessons learnt from using a mixed method approach within this thesis. The chapter closes with a set of generic universal principles or design heuristics that can be used to guide researchers or practitioners in the design or evaluation of interactive technologies, along with some suggested areas for future research.
7.1. Introduction

This chapter begins with a summary of the key findings that were derived from the three studies presented in Chapter 3 to 6, which are grouped into four main research contributions. The findings are reviewed alongside the aims and objectives, and original research questions outlined at the start of this research. The research constraints are also considered in relation to lessons learnt about the nature of the methods employed within this research and the inferences they generate. An expanded version of the cognitive process model of UX quality judgement proposed by Hartmann, Sutcliffe, & De Angeli, (2008) is presented, which was used as a starting point to this research. The model is discussed in relation to the decision-making theories and how this relates to the inclusion of the first three main contributions, interactivity, user groups and UX over time. The fourth contribution outlines the main insights gained from the methodological approach used within this research. These four key contributions identified within this thesis are then discussed in reference to the relevant literature within the HCI field and beyond. The chapter concludes with set of UX design heuristics generated from the main findings, along with some suggestions for future research.

7.2. Summary of Main Findings

The motivation for this research was guided by the main aims and objectives which first set out to expand the cognitive process model of user judgement for quality proposed by Hartmann et al., (2008), through exploring the additional constructs of affect, immersion and presence. The influence of interaction on a variety of UX constructs was investigated in order to understand what attracts people to use an interactive product, and how interactivity impacts on user engagement. Different interactive design features were examined in relation to various UX measures and how these may influence user judgement of the interactive product quality. User diversity was explored through using a cluster analysis technique and the effect of long-term usage over time was captured using a variety of methods to understand the decision-making process when adopting new technology, and what key constructs may influence overall user satisfaction.

7.2.1. Main Contributions

The main contributions of this thesis are shown in Table 74, which identify seven key findings from each of the three main studies reported in Chapters 3 to 6, which are grouped into four main contributions; interactivity, user groups, UX over time and methodological insights, as now summarised.
This research clearly demonstrated the importance of interactivity in positively influencing UX\textsuperscript{1}. Both quantitative and qualitative results from the first two studies (Chapter 3 \& 4) confirmed that interactivity matters in two ways; first the overall experience of interaction reflected in pre to post-test results; and secondly, the enhanced interactive features used within websites improved user engagement\textsuperscript{2}. However, the importance of interactivity overall masked the preferences for particular styles of interaction, as reflected by the three user groups\textsuperscript{3}, where enthusiasts showed a preference for either the avatar or 3D navigation. It was also shown that user judgement was influenced by their technical disposition (power usage)\textsuperscript{4} indicating that prior experience mattered. Interactivity was also shown to be a contributory factor to long-term satisfaction in the longitudinal study (Chapter 5 \& 6), which clearly identified that the variety of apps contained on the iPad facilitated positive experiences. Popular apps were those that supported dual-use (work/leisure), provided unique functionality and prompted interactivity that enabled communication (CMC), social activities and engaging experiences (multimedia video, games etc.). The dominating factors that influenced long-term UX were usefulness, usability (PQ) and utility (functionality of apps), where hedonic ratings of device remained high\textsuperscript{5}. A negative perception of utility and usefulness, along with a resistance to change, were the main reasons for non-adoption\textsuperscript{6}, further exacerbated by a lack of advice or training. Lastly, the mixed methods approach revealed that user judgement can shift\textsuperscript{7}, depending on the type of measures employed, as quantitative questionnaire responses are more instant ‘affect’ based, while interviews stimulate more reflective judgements.

### Table 7.4

<table>
<thead>
<tr>
<th>Four Main Research Contributions</th>
<th>Study One Chapter 3</th>
<th>Study Two Chapter 4</th>
<th>Study Three Chapter 5/6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interactivity positively influences user experience</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>2. Interactive design features* improve user engagement</td>
<td>✓*</td>
<td>✓*</td>
<td>✓**</td>
</tr>
<tr>
<td>3. User preference for different interactive designs can be explained through user types</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>4. User judgement was influenced by an individuals’ technical disposition</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5. Over time, usefulness, usability (PQ) and utility (interactivity of apps) were dominating factors effecting UX \ satisfaction</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>6. Perceived usefulness and utility are antecedents to product adoption</td>
<td>-</td>
<td>-</td>
<td>✓</td>
</tr>
<tr>
<td>7. User judgement can shift depending on the research method employed</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Key:** ✓* = Interactive features: avatar, active media and 3D navigation (Study 1 \& 2), ✓** = Video and CMC Media (Study 3)

Table 7.4. A summary of the four main research contributions from this thesis
Chapter 7: Main Discussion

The main contributions are now reviewed in relation to the original research questions posed in Chapter One:

- **What are the key constructs that contribute to user engagement and user experience?**
  Affective and hedonic responses to an interactive product were found to positively influence users overall judgement of product quality, although over time usefulness, pragmatics (usability) and utility (due to the interactive apps) dominate.

- **Does enhanced interactivity (design features) lead to more positive user experience?**
  Interactivity clearly matters in two ways; first, the general experience of interaction was demonstrated by an increase in affect from pre-interaction to post-interaction measures; and secondly, inter-site differences revealed that interactive features (avatars/3D navigation) positively influenced users affective and hedonic responses. Clearly the apps contributed the most to positive UX over time, facilitated by those that provided greater interactive experiences.

- **What are the key constructs that contribute to user satisfaction and product adoption over time?**
  Perceived usefulness and utility (interactivity of apps) were the most important determinate of product adoption, where usability was often tolerated. Device accessibility and portability were strong motivators of use, yet motivation levels differed between the adopters and non-adopters. Negative perceptions of usefulness and utility of the apps were the main barriers to adoption, along with users reluctance to change, forced mandatory use and a lack of advice and training at how best to utilise the technology.

### 7.2.2. Research Constraints

This research adopted a mixed methods approach to bridge the gap between generic quantitative ratings with individual qualitative accounts, producing some valuable insights into the reasons behind user judgements. However, using a multi-methods approach presented a number of research constraints that are worth noting.

The first two studies were lab-based experiments, where participants evaluated the websites through following simple tasks requested by the researcher. Although participants were given free time to explore the site, this approach is known to have certain limitations due to the un-naturalistic tasks and environment. To keep the experiments within a realistic time frame, the interactive sessions were relatively short (5-10 minutes), so the effect of
interactivity may have been transient, hence few effects on overall judgement were found. Furthermore, the participants were not given goal-orientated tasks, so the effect of different task modes (goal-mode or action-mode) may have had an impact (van Schaik & Ling, 2009), future research could investigate for such influences.

The first study investigated real-world website designs, but sacrificed a loss of experimental control. Although the sites were from the same domain, confounding aspects, such as content and design style, may have influenced participants’ quality ratings. The Louvre character may have made the content appear more interesting and accessible, which possibly may have overcome any aversions to the avatar itself. While the Google website may have had less appeal (due to its dark palate and atypical website layout), although hedonic measures showed no differences between the sites. Brand and prior experience may have influenced user judgement, although these were tested, by adding them as covariant, and no effect was found. Study Two controlled any confounding variables by using two identical websites and manipulating specific interactive features. However, the bespoke sites quality suffered due to their simplicity, and the avatar design contained some usability errors, which may have biased judgement against the more interactive version. It is a testament of the power of interactivity that the interactive version was preferred in spite of the usability flaws.

Some questionnaire measures (e.g., immersion/presence) produced no inter-site differences, which could be attributed to the difficulty in capturing immersion. Flow-type immersive experiences can be difficult to measure, as they are often fleeting, occurring within-session, but can rapidly dissipate post-interaction (Csikszentmihalyi, 1990). Alternative, less disruptive methods could be used to captures immersive flow-like experiences, such as ESM (experience sampling method), (Csikszentmihalyi & Larson, 1987), facial recognition, or galvanic skin response (GSR), which could be used in combination with measurement scales (Nacke & Lindley, 2008). Furthermore, the sequence order when completing multi-construct questionnaires can impact upon the results. For example, the immersion/presence scale was completed after both the affect and website quality scales, further delaying users reporting any flow-experience, thus limiting the capacity to capture immersive experience. Careful attention is also required in the experimental design, as despite counterbalancing the website order, those who experienced the more interactive site first, may discount the non-interactive site more, resulting in higher negative comments and lower quality ratings.

The longitudinal study provided some valuable insights into UX in an ecological setting using a mix of data collection methods (questionnaires, diary-study, interviews), however, conducting such a study created a number of challenges. Time constraints and limited resources are inherent limitations, because not only are long-term studies very labour
intensive to run; the amount of data generated can be overwhelming. As a result only few longitudinal HCI studies have been conducted, which tend to use low participant numbers and last for shorter durations of time (Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, & Sinnelä, 2011; Vermeeren et al., 2010). Careful planning and implementation are crucial when making such an investment, especially when using a mixed-methods approach.

The relatively small sample sizes used within all three studies (N=40-47) created certain constraints on the analysis methods employed. Although it found some significant results, it would enhance the validity to repeat the study with a larger sample group, which would increase the option for further statistical analysis (e.g., factor analysis). Participants tended to be younger (>30) and were mainly university students, limiting the generalizability of the findings, although no age related effects were found. Prior art or art galley website knowledge was accounted for in the first study, and no influence was found. Previous experience with Google 3D technology and avatars was an uncontrolled influence, although the Google street view technology is fairly widespread and prevalent, and the younger participants tend to have experience with avatars (e.g., 3D games, SecondLife). In the longitudinal study, previous experience with tablet technology (pre-owners N=7) showed no impact on the overall results. However, it is likely that over time and prolonged use some participants moved from novice to expert users, which may have impacted upon the overall ratings, as revealed by the converging for some of the power user results, (where low power users benefited the most). Additional contextual influences may have also impacted on the participants' experiences (e.g., work timetable patterns, social group influences), and these would provide some interesting areas for future longitudinal research.

7.3. Expanded Model

Prior to the main discussion, three (of the four) main contributions are reviewed in relation to the cognitive process UX model that was used as a starting point to this research. These findings extend the original cognitive UX process model by Hartmann et al., (2008) in three ways, first, to account for the interactive experience (interactivity), secondly, to verify the importance of individual differences (user groups), and lastly to further expand how the various constructs within the model may change over prolonged use and time (UX over time), and what factors lead to product acceptance. The expanded model is illustrated in Figure 78, (the green shaded boxes show the additional three main contributions).
Figure 7.8. The expanded model of users’ decision-making process during UX

The first stage is assess (1), which is determined by the UX context of the product (e.g., serious or fun), the task-in-hand and users motivational goals, which have been widely acknowledged to affect UX judgement in several studies (De Angeli, Sutcliffe, & Hartmann, 2006; Diefenbach & Hassenzahl, 2009; Hassenzahl & Tractinsky, 2006; Porat & Tractinsky, 2012). This is followed by select criteria (2), which depend on various product qualities than can determine a variety of influences on the decision making process. Previous criteria (usability, utility, content, brand, aesthetics, customisation) have shown to contribute to user decision-making, as verified by a number of studies (De Angeli, Hartmann, & Sutcliffe, 2009; Hartmann, Sutcliffe, & De Angeli, 2007). This research demonstrated that interactivity influences user judgement, and so interactivity is added to the criteria. The context and criteria are antecedents to users initial perceptions and actual experience when evaluating an interactive product. In addition, the users profile (skill, knowledge and experience) is known to influence user judgement when evaluating their experience (3) with a technological product (Harbich & Hassenzahl, 2011; Hartmann et al., 2008; S. Lee & Koubek, 2010a). This research showed that user preference for different interactive features, and their technical pre-disposition can influence their overall judgement of a product, and so the inclusion of user groups is added to the model. In addition, the criteria people use to make judgements can shift as their experience progresses over time (multiple sessions), where aesthetics and hedonics are important attributes in the initial impressions, but as the interactive experience progresses, usability, utility and usefulness dominate, especially when using task-orientated products. The effect of long-term UX has been shown to influence user judgement (Karapanos, Zimmerman, Forlizzi, & Martens, 2009b; Kujala et al., 2011), this research demonstrated the transitional effect of user judgement and how certain criteria (usability, utility and usefulness) can influence product adoption; so UX over time is added to this model.
These three contributions are now discussed further in relation to the expanded cognitive UX process model. (The cross-references included link to the design heuristics presented in 7.5).

### 7.3.1. Product Qualities – Interactivity

Interactivity is added to the list of *product qualities* or attributes that can influence users evaluations of a product quality. This research showed that interactivity is an important factor in engaging users, which can be linked to specific interactive features influencing both short and long-term use.

The cognitive model of user judgement posits that judgements of interactive technology are based on 'fast-path' decisions, which are made by comparing various attributes and selecting the most favoured, drawing from the dual-process theory, or ELM (Elaboration Likelihood Model) by Petty & Cacioppo, (1986). Ideally 'slow-path' user judgement is made systematically by considering each attribute in turn to gain an overall evaluation (Keeney & Raiffa, 1993), while ‘fast-path’ decisions are often irrational and automatic based on the users intuition (Kruglanski & Gigerenzer, 2011; Sutcliffe, 2009), especially when information maybe absent or limited. This is consistent with the inference perspective, where judgements of a products perceived usability are based on the available aesthetic attributes (Hassenzahl & Monk, 2010; Hassenzahl, 2004). To add to the complexity, a users decision is often contingent upon a variety of factors, such as the task environment (Newell & Simon, 1972), and task context (Payne, Bettman, & Johnson, 1993), so decision-making strategies depend on how salient the task and contextual influences are to the user, as identified by adaptive decision-making (ADM) theory (Payne et al., 1993). In relative low value decisions (i.e. most personal software products), the 'elimination-by-aspects' (EBA) strategy is likely to be favoured, where attributes are selected by their importance or value to the individual. In general product choice this can involve trade-offs between different attributes, such as costs verses aesthetics, as reported by Diefenbach & Hassenzahl, (2009). In trade-off effects or *task framing* (see Heuristic i), users can modify their choice depending on the task, where more serious goal-orientated tasks (e.g., sending an email) favour usability and utility criteria, while more discretionary relaxed, leisure or fun based tasks (e.g., playing games, or social networking) favour aesthetics and interactivity (De Angelii et al., 2006; Hassenzahl, 2008a). In the first two studies (Chapter 3 & 4) the more interactive sites were preferred demonstrating that interactivity is an important product quality criteria, favoured above usability. When given choice scenario situations users trade-off preference between enjoyable (hedonic), engaging (interactive) and usable (pragmatic) designed products (Sutcliffe & De Angelii, 2005), where either the hedonisics, aesthetics, interactivity or pragmatic aspects of a product are favoured (Diefenbach & Hassenzahl, 2009). *Product
framing effects (see Heuristic ii), can have a similar response, where work/goal related products favour usability and utility product attributes, while games/entertainment type products select aesthetics, enjoyment (Hassenzahl, Diefenbach, & Göritz, 2010), and interactivity, as they promote motivational aspects of excitement and curiosity (Jennett et al., 2008). Although the first two studies provided no work/goal related tasks (see constrains 7.2.2), the third ecological iPad study took place in the field combining both work/leisure activity use due to the variety of apps available to support dual-activities. This demonstrates that interactive products can support work-based activities in a holistic way, where both pragmatic goal-orientated task-based activities can benefit from more engaging hedonic based interactions, (Harbich & Hassenzahl, 2011).

7.3.2. User Profile – User Groups

According to ADM theory people’s decision-making is not only adaptive and dependent on the task and context, their background experience (e.g., culture, education, training) can also impact on their judgement (Hartmann et al., 2007). The process decision making framework posit that aesthetic judgment is subject to framing effects of both task and the users’ background (Hartmann et al., 2008). Although usability and aesthetics are important factors, they are sensitive to individual user differences. (De Angeli et al., 2006). This research adopted a user-type profiling approach that used a cluster analysis technique to identify different sub-groups based upon their preferences for interactive styles and their predisposition towards technology. Applying a user-categorisation technique the original process model was expanded to allow for different user groups, indicating that the ‘one-size-fits-all’ approach maybe inappropriate.

7.3.3. Long Term Experience – UX Over Time

One of the main contributions of this research was to extend the criteria by adding interactivity (see Figure 78), which join a number of various generic attributes (product qualities) used to assess the quality of websites and technology products. However, the criteria and product qualities are subject to change over time, from an initial encounter (Hassenzahl, 2004; Lindgaard et al., 2011), to after interaction (Cyr, Head, & Ivanov, 2009; S. Lee & Koube, 2010b; Teo, Oh, Liu, & Wei, 2003), and over longer periods of use (Karapanos et al., 2009; Kujala et al., 2011).
At different phases of use the criteria importance changes, as shown in Figure 79. On the first encounter with the product (pre-use) aesthetics and perceived usability is judged primarily by the products visual appeal and presentation (Lavie & Tractinsky, 2004). People can make very quick and reliable judgements of a product (within 50msec), where first impressions are driven largely by the visual appeal of the product as it requires less cognitive processing, (Lindgaard et al., 2011; 2006), while longer exposure time allows for assessments of more cognitively demanding attributes, such as usability and utility. After the initial interaction hedonics (HQ) and affect ratings increased as a result of interaction, which was more marked for sites containing enhanced interactive features (e.g., avatars, 3D navigation etc.), where usability was adequate, as demonstrated by Study One and Two. Similar pre and post-use studies reflect this general effect of interaction (Ben-Bassat, Meyer, & Tractinsky, 2006; S. Lee & Koubek, 2010b; Lowry et al., 2006; Tuch, Roth, Hornbæk, Opwis, & Bargas-Avila, 2012).

In early-use, involving several sessions of interaction with the product, utility (functionality and content) became more important, while poor usability was often tolerated in favour of rewarding interactive experiences, as reflected in Study Three; and reported in other similar studies (Cyr, Head, & Ivanov, 2009; Teo et al., 2003). However, for long-term use, over multiple sessions of interaction, the products usefulness and utility were dominant factors for judgement of quality and product adoption, as reported in Chapter 5 & 6. Other longitudinal studies conducted that have reported on much shorter time frames using smaller participants groups. For example Karapanos, Hassenzahl, & Martens, (2008) found that the first experiences of using a pointing device for interactive TV, the pragmatic aspects (usability and utility) were the most prominent determinates of goodness, but after a month, user identification (self-identify) became the most dominate aspect of product goodness. In a similar study Karapanos et al., (2009) followed 6 participants over 5-weeks after purchasing of a new mobile phone, finding that hedonics (novelty) quickly faded over time, being replaced by identification (HQI), while usability and usefulness becomes a dominant factor to long-term use. Other studies also report that pragmatic perceptions (usability) improve, while hedonics decrease over time (Kujala et al., 2011; Mendoza, Novick, & Paso, 2005; von
Chapter 7: Main Discussion

Wilamowitz-Moellendorff, Hassenzahl, & Platz, 2006). This transitional view of UX is still governed by the criteria selection process model identified in Figure 7.8, where the context of the task (fun or serious) and user background (skill, knowledge and preference) are influential. However, the criteria importance of the various product qualities shift over time, where hedonics and interactivity are important short-term, but utility and usefulness become more dominant long-term.

7.4. Main Discussion

The key findings outlined in Table 7.4, are discussed according to the four main contributions found within this thesis, interactivity, user groups, UX over time & adoption, and methodological insights, as shown in Figure 8.0. The discussion will review the main findings across each of the three studies in relation to the current literature, (additional cross-references will be shown that link to the design heuristics presented in 7.5).

\[ \text{Figure 8.0. An overview of the five main contributions of this thesis (and how they link to the chapters)} \]

7.4.1. Interactivity

The findings from this research have clearly demonstrated that interactivity matters, as revealed in the first two studies (Chapter 3 & 4). This was shown in two ways; first through repeated exposures where affective and hedonic measures increased from initial perception to post interaction; then secondly, inter-site differences showed a marked increased in ratings for the more interactive sites in the same measures, which were supported by the high volume of positive user comments for specific interactive features. Different types of interactive designs (e.g., avatars, 3D navigation) were also tested, as reflected within the user-group preferences (e.g., enthusiasts). The general effect of interaction has been well established by many studies comparing pre and post interaction that demonstrate the
Chapter 7: Main Discussion

dominance of aesthetics in pre-interaction experience, where usability and utility become more important after interaction, impacting upon user satisfaction (Ben-Bassat et al., 2006; S. Lee & Koubek, 2010a, 2010b; Lindgaard et al., 2011; Lowry et al., 2006). The clear inter-site differences shown in the first two studies (Chapter 3 & 4), indicate that interactivity is an important design aspect that influences users’ affective and hedonic experiences. The strong effect between initial exposure to interaction were also found by Teo, Oh, Liu, & Wei, (2003), where increased interactivity levels had positive effects on users’ sense of satisfaction. Similar results from both studies showed affect ratings were considerably higher after interaction, as well as favouring the more interactive sites; furthermore, the ANOVA interaction indicated that the enhanced interactive features appeared to reinforce positive user experience. Clearly, the sites that contained the interactive features had a strong effect on UX, as the interactive sites for both studies were ranked first over the standard (non-interactive) sites, and generated the highest volume of comments. However, little difference was found between the two interactive sites on global measures (ANOVA & regressions), where interactive style preferences only became manifest in the qualitative results. Overall, interactivity was well liked by the majority, where most users (e.g., enthusiasts and adopters) were more pre-disposed towards novel interaction styles, but a small minority (non-adopters) preferred the more traditional style of interaction (menu-link).

These findings reinforce previous studies (Cyr, Head, Larios, & Pan, 2009; O’Brien, 2010; Teo et al., 2003) that have demonstrated the importance of interactivity on various measures of UX. However, these studies manipulated either the ‘look and feel’ of the websites information visualisation to focus on perceptions of aesthetics (Cyr et al., 2009, P.586); or manipulated different levels of CMC tools, involving either asynchronous (feedback forms) or synchronous communications (online forum), (Teo et al., 2003); or conducted surveys to examine the interactive experience of online shopping (O’Brien, 2010). This research has demonstrated interactivity makes a contribution to UX research in both short and long-term use. The short-term interactive experiences generated from specific interactive features (e.g., avatars, 3D navigation etc.), that share similar experiences to flow and immersion typically found in game design, (Jennett et al., 2008) can create excitement within UX. Technologies that enable interactive experiences such as CMC and social interaction can enhance long-term usefulness and positive UX, facilitating product adoption, as reported in Study Three.

In the first study (Chapter 1), 3D virtual worlds using a fly-through navigation, a virtual character and interactive animated links with audio media, all contributed to increased affective responses and hedonic ratings leading to a more positive user experience, compared to traditional menu-link designs. The second study (Chapter 4) manipulated design features (avatar, active links and media) reinforcing these findings. Although these interactive
features generated opposing preferences, for example, the avatar within the second study (Chapter 4) was considered engaging, but received a high volume of adverse usability comments; similarly the animated character in study one (Louvre website) received conflicting comments, some finding it fun and engaging, while others finding it distracting and irritating. So character based interactions appear to generate divergent opinions that can promote engaging experiences, but may arouse displeasure and annoyance along with poor perceived usability.

Both the interactive sites in the two website studies were the preferred choice, indicating a strong interaction effect, despite adverse usability problems, which were mainly attributed to the interaction difficulties of Google (3D navigation), and the poor avatar design in both studies. In a study of avatars within an online shopping environment, Qiu & Benbasat, (2005) found participants reported a greater sense of telepresence, yet no differences were found in perceptions of flow, which were attributed to the avatars poor realism that lacked vivid facial expressions and limited interactivity. While Luo, McGoldrick, Beatty, & Keeling, (2006) reported that although onscreen characters are likable, if users detect limitations with the avatars (e.g., mismatch in the human-like characters), they become frustrated and irritated. However, in an experiment by Van Vugt, Hoorn, Konijn, & de Bie Dimitriadou, (2006) the aesthetics (beauty) and affordance (helpfulness) of the avatars from the Sims game were manipulated to reveal that user engagement was enhanced by the beauty and perceived affordance of the avatar, which influenced user experience and satisfaction. Similar findings were reported by Khan & Sutcliffe, (2014) who showed the attractiveness of an agent positively influenced its persuasiveness. The inclusion of interactive avatars/agents to provide customer support, online training, and for entertainment purposes has shown to enhance engagement, improve persuasiveness, and provide a more a personal human experience (Fogg, 2002; Qiu & Benbasat, 2005). However, when selecting appropriate avatars it is important they are suitable well designed (see Heuristic vi), and the user is offered the choice as to whether to interact or not, and how to interact with an avatar (Fogg, 2002; McGoldrick, Keeling, & Beatty, 2008).

Overall the interactive features enhanced positive user experience, however good design is crucial. In the first two studies some participants reported feeling frustrated as they felt little control over the interactive experience due to the interaction style. For example the Google fly-through metaphor can present navigational barriers to some users, and the virtual character produced an adverse reaction, especially in the second study, due to poor control, (see Heuristic x). Although user control has been considered in the perspective of two-way communication, the level of user control over a users interactive experience needs to be considered in order to avoid usability problems (Lee, 2005; McMillan & Hwang, 2002).
Ensuring a well-designed interactive experience or by providing alternative navigation options can negate user frustration and aid user control (see Heuristic vii). Both studies showed that poor design of the interactive features generated the most usability comments, although these were forgiven to a surprising extent. Explanations could either be that usability has to be ‘good enough’, and once that level is reached, it no longer has an adverse effect; or the positive affect generated from interactivity has a halo effect on poor usability (see Heuristic iii). Considering the high volume of negative comments on the avatar, particularly in the second study (Chapter 4), which had obvious defects, it seems the second interpretation is more likely. Both website studies ratings were based on retrospective recalled experiences, where affect measures were captured immediately post-interaction, suggesting that the within and post session interaction produced positive affect, but only lasted short term (as reported in the regression results). However, in the longitudinal study (Chapter 5 & 6), affect was captured over long-term use (not within session or post-interaction), thus allowing users affective responses to impact on their attitudes and overall judgement of quality.

Although interactivity had a strong effect on hedonic and affect ratings in the first two studies (Chapter 3 & 4), overall preference ratings were marginal for the interactive websites. This could be that while interactivity had a strong in-session effect, when users assessed the site post-interaction they may have been assessing its utility (function) and content as they reflected on their overall experience. This contrasts with findings from (Cyr, Head, & Ivanov, 2009; Teo et al., 2003) where the more interactive styles received more favourable overall evaluations. Possible explanations could be due to the tasks employed in these studies, which took place in e-commerce settings and used more goal directed tasks, reflecting a more utility-based experience. In contrast, the studies in this thesis, focused upon different interactive features that enabled multimedia CMC (communication-mediated communication), interactivity (avatar and video), which were more focused on exploration.

Due to the longitudinal nature of the third study, interactivity was not directly manipulated. Findings from the interviews showed that positive interactivity comments related mostly to the applications contained on the iPad device, where video-media, games, communication and medical apps were favoured the most. The use of interactive features to enhance user engagement is common practice within game design (Salen & Zimmerman, 2004) and has been reported in several studies with much research focused on immersion and flow (Jennett et al., 2008; Sweetser & Wyeth, 2005). Although gaming apps were favoured for their interactivity, gaming activity decreased over time with negative affective comments related to boredom, which could be attributed to the simplistic nature of some of the games that may not maintain long term engagement (Csikszentmihalyi, 1975; Seah & Cairns, 2008).
When designing for interactive experience certain levels of ‘adaptivity’ or ‘adaptability’ of product could help maintain user engagement over long-term (Arhippainen & Tähti, 2003), and keep interactivity fresh (see Heuristic viii). Adaptivity refers to the capacity for a system/product to adapt automatically without user’s intentional action, (Benyon, 1993; Höök, 1997), while adaptability is the capacity for a product to enable customisation (see Heuristic ix). This was reflected in the iPad study, where the variety of apps enabled adaptability, thus enhancing long-term satisfaction. The ability to customise products through offering a choice of different design alternatives (e.g., MP3 player or mobile phone ‘skins’), that can satisfy a diversity of users’ tastes, can positively influence overall judgements of aesthetics (appeal) and hedonics (identity) beyond the products functionality (Hassenzahl, 2004; Tractinsky & Zmiri, 2006). However, users can be divided between the aesthetic or customisation option in a trade-off choice (Hartmann et al., 2008), indicating differing preferences are found within a user population.

The most popular apps were favoured for their dual-functionality and ability to support learning (e.g., YouTube), along with providing entertainment ‘on the go’ (e.g., iPlayer). The video media apps showed the greatest increase over time, due to their dual-use to provide alternative methods of learning and leisure entertainment. Dynamic interactive media (e.g., animation and video) is known to enhance arousal, attract attention and promote excitement (Reeves & Nass, 1996). The value of interactive video media has been well recognised as an effective support to learning (Bransford, Brown, & Cocking, 1999; Domagk, Schwartz, & Plass, 2010; Schaffer & Hannafin, 1986; Zhang et al., 2006), as it allows students to explore and re-examine content both individually and collectively and improves product utility and user satisfaction. Using interactive technology to communicate with others has been considered to be another important dimension of interactivity (Sohn & Lee, 2005). According to Ha & James, (1998), connectedness can be characterised as “the feeling of being able to link to the outside world and to broaden one’s experience easily” (P. 462). Applications that facilitated connectedness (e.g., Skype, Facetime, etc.) were also favoured because they enabled unique personal interactive communication, which promoted social presence (Short, Williams, & Christie, 1976) and enhance user engagement (Sutcliffe, 2009). While branded medical apps were preferred for their trusted content, where poor usability was often tolerated. When brand image is valued, it can exert a positive influence on other criteria such as usability and aesthetics (Bolchini, Garzotto, & Sorce, 2009), can influence perceptions of trust and reliability (De Angeli et al., 2009) and aid persuasiveness influencing the users’ intention to buy (Bruijn, De Angeli, & Sutcliffe, 2007). (see Heuristic iv).
Chapter 7: Main Discussion

Clearly interactivity is an important factor in contributing to positive UX in both short and long-term use, as it enhances user engagement and facilitates interactive experiences that contribute to product adoption.

7.4.2. User Groups

This research identified three user groups that generate differing levels of user judgement based on two aspects, user preference for specific interactive designs, and the users’ individual technical predisposition. Adopting a user-typological approach enables a better understanding of the user population and adds a new prospective within UX research, which calls for a more flexible design approach to support individual difference (see Heuristic v), rather than the standard ‘one-size-fits-all’ mode.

User Preference: Users’ attitudes to interactivity can vary depending on their background and preferences (Kirk, Chiagouris, & Gopalakrishna, 2012), where aspects such as innovation and curiosity have shown to influence users’ desire to explore new technology (Magni, Susan Taylor, & Venkatesh, 2010). User judgement of the contrasting websites in the first two studies (Chapter 3 & 4) showed interesting differences between the quantitative and qualitative results. In the first study (Chapter 3), the Louvre site was slightly favoured overall in the quantitative measures, whereas the Google site was more salient in memory recall after interaction. Although no significant differences were found between the two sites for the global quantitative measures (beauty, goodness & overall experience) the qualitative findings indicated that the sites were liked or disliked for different reasons. The Google 3D navigation metaphor was appreciated less than the avatar and interactive graphical features contained in the Louvre site, which was ranked the highest overall, indicated differing individual preference. Cluster analysis techniques were developed to analyse individual and sub group differences, leading to the identification of three user sub-groups. Enthusiasts in both studies appeared to like either the avatar in Louvre (and in the second study), or the 3D interactive navigation (Google), which was reflected in their overall choice. Further analysis of the dependant variables and user profile measures was conducted to detect user groups, similar to Brandtzæg’s, (2010) media-user typology approach. Three user groups emerged, enthusiasts who enjoyed the site’s interactive design, adopters who were also positive about the site’s interactive features but were critical of usability issues, and non-adopters who didn’t like the site (Hart, Sutcliffe, & De Angeli, 2013). An interesting difference between the enthusiasts and adopters is the former group were more tolerant of usability problems in their overall preferred site. This indicates that the users’ overall preference towards specific interactive features showed a strong influence on their utilization and enjoyment of an interactive product (Davis & Yi, 2012; Magni, Susan Taylor, & Venkatesh, 2010).
Chapter 7: Main Discussion

Little research within HCI has explored user diversity. Although, there is a growing area of research into recommender systems that extract user preferences either implicitly through their online search patterns or through the system learning user preferences based on individual user feedback (Ricci, Rokach, & Shapira, 2011; Wasfi, 1999). However, these systems suffer from ‘new user problem’, where initially they know very little about a new user, so it can take time to build up a user profile (Rashid et al., 2002). Understanding the diversity of user preferences prior to interaction can provide valuable insights into user engagement. Adopting a user-typological approach through using a cluster analysis technique to distinguish different sub-level group preferences of specific interactive features is a contribution to UX research.

**Technical Predisposition:** Prior experience has been recognised to influence user judgement (Arhippainen & Tähti, 2003; Hartmann et al., 2008), affect user behaviour (Ajzen & Fishbein, 1980; Taylor & Todd, 1995), and shape product adoption (Shih & Venkatesh, 2012). Users perceptions of aesthetics is also influenced by the users’ background (Hartmann et al., 2007), CVPA (centrality of visual product aesthetics), an individuals disposition towards the visual aesthetics of a product (Bloch, Brunel, & Arnold, 2003) and impact on their general design preference (Kim, Lee, & Choi, 2003).

An individuals’ technical disposition can influence their judgement of an interactive product, with the more technical disposed users being more tolerant, or less sensitive to usability problems, as identified in the second and third studies. Three user groups (IT positive, IT negative & IT neutral) were identified in the second study (Chapter 4) using cluster analysis, based on participants predisposition towards technology (Marathe, Sundar, Bijvank, van Vugt, & Veldhuis, 2007), with the IT positive giving the interactive site higher ratings for usability. A similar cluster analysis was conducted in the third longitudinal study (Chapter 5), showing a 3-group split: high (IT positive), low (IT adverse), and mid power users (IT ambivalent). The IT positive users gave higher PQ and hedonic ratings, indicating they were more tolerant of usability problems than the other two user groups, and found the iPad more pleasurable and enjoyable to use. Whereas the IT adverse users gave the lowest ratings for PQ and hedonics, and the highest negative frustrating ratings, indicating they were less accepting of usability problems; while the IT ambivalent users (the largest group) showed an increase in PQ after a mid-study dip, indicating that early usability issues were either resolved or forgiven. These findings can be interpreted in the perspective of Rogers’s, (1995) innovation theory, who categorised participants into 5 groups, according to their speed of uptake and attitude towards technology, where positive evaluators (innovators and early adopters) gave higher satisfaction scores, (followed by early majority, late majority and laggards); with early and late majority (mid group) being more critical, while the laggards
only adopting technology if they have too. It is likely that the non-adopters (Study One), and IT negative sub-group (Study Two) could became long-term non-adopters (as identified in Study Three), due to their negative disposition towards new technology.

The multi-dimensional Power User (PU) scale was chosen in this research as it identifies users under four dimensions (motivation, expertise, efficacy and behaviour), as opposed to the one-dimensional PIIT scale that only measures personal innovativeness (Agarwal & Prasad, 1998). Power users are more efficient, creative and motivated users, and will typically make more productive use of a technological device then innovative users (Zhong, 2013). For examples, Sundar & Sampada, (2010) found that power users prefer customisable tools that allow them to control the active content within a website. The IT positive users (PU) identified in the third study gave higher positive ratings (hedonics, pragmatics, ease-of-use) than the other two groups, and were generally more motivated to use the iPad, indicating that predisposition has an influence on user judgement.

The new perspective of user sub-groups is a contribution to the UX field that bridges population-level quantitative studies with individual-level qualitative studies that focus on subjective interpretations of experiences (Wright & McCarthy, 2010). User technical predisposition and user diversity can impact on user judgment and influence UX as shown in several studies (Marathe et al., 2007); for example, Jumisko-Pyykkö & Häkkinen, (2008) investigated several user characteristics, such as age, where the younger participants reported to be more critical of technology, which was interpreted as an indication of their wider media and technology experience. Whereas Wu, Li, & Fu, (2011) found that some users had higher technical curiosity traits due to their willingness to try out new innovations, as measured by PIIT (Personal Innovativeness in IT), a scale used across many various studies to explore users attitudes towards using technology. Similarly Magni, Susan Taylor, & Venkatesh, (2010) demonstrated the association between PIIT and users increased intention to explore technology.

User differences has also been found in users cognitive style of visual processing, where higher visual preference users are more critical (Jumisko-Pyykkö & Strohmeier, 2013), and individuals with high CVPA (centrality of visual product aesthetics) attach more value, acumen and respond more positively to the aesthetics of a product, leading to a higher purchase intention (Bloch et al., 2003). Although the IT positive users in the third longitudinal study gave more favourable hedonic ratings than the other two groups, suggesting they were influenced more by the visual appeal of the iPad, no differences were found between groups for beauty, suggesting their high hedonic ratings related to their pleasure-experience rather than their general aesthetic appreciation. However, Wakefield &
Chapter 7: Main Discussion

Whitten, (2006) reported no differences in hedonics between cognitive absorption (low/high absorption) in an experimental contrast of hedonic/utilitarian on mobile phones, but they found that cognitive absorption did influence perceived usefulness. The IT adverse users showed an increase in usefulness (TAM) ratings over time, so the iPad became more useful and therefore they benefited the most. This suggests that user groups are transient and change over time, as reflected in the convergence for some of the three user group ratings after 6 months of usage. While user group members may co-exist at a certain time, their membership distribution can change, reflecting a transition across different phases of product use (Karapanos, Zimmerman, et al., 2009b). Providing alternative interactive options (see Heuristic vii) can enhance users sense of control, (Cyr, Head, & Ivanov, 2009), that accounts for different types of user (high/low power users), and accommodate for their transitions over time, as users gain more experience, their sense of control increases (Sundar & Marathe, 2010), (also so Heuristic x).

7.4.3. UX Over Time & Product Adoption

Most of the past UX research has focused on single momentary interactions with a product, with few published studies reporting on long-term (+2 month or more) evaluations of UX. The third study reported in Chapter 5 and 6 addresses the lack of longitudinal UX and uses a mixture of concurrent research methods to investigate the temporal aspects of UX and the reasons behind product adoption/non-adoption.

**UX Over Time:** Understanding the transitional nature of UX is complex as the meaning of UX changes, where initially (perceived) usability and interactivity are the most favoured criteria, yet over time usefulness, utility (functionality) became the dominating factors affecting experience and user satisfaction, while hedonics remained stable (see Figure 79). These findings expand previous studies by investigating a much longer time period compared to the short ethnographic study by Karapanos, Zimmerman, Forlizzi, & Martens, (2009a), who followed the hedonic changes of 6 new iPhone users, to reveal initial novelty shifted to self-identify (meaningful attachment) after 5-weeks of usage, and ease-of-use (the main predictor of goodness) changed to usefulness, becoming the dominant long-term factor. In contrast, Kujala, et al., (2011) reported both pragmatic and hedonic qualities contributed to the attractiveness of mobile phones over 3-12 month period in a qualitative study that used a retrospective memory technique. This differed from other studies, including the longitudinal study (reported in Chapter 5 & 6), as hedonic perceptions showed a mid-study dip (after 3 months), where increased frustration related to an increase in usability problems. This concurs with Mendoza, Novick, & Paso, (2005), who observed novice users creating school webpages over 8 weeks to reveal that frustration decreased over time, as participants found...
ways to resolve the usability problems. Similarly, von Wilamowitz-Moellendorff, Hassenzahl, & Platz, (2006) investigated expert users (N=7) of mobile phone use, and found that ease of use and pragmatic perceptions (usability) improved as users became more familiar with using their devices, while their stimulating experiences (hedonics), decreased over time indicating a novelty waning effect. The more persistent problems (network browser problems, missing features and hard-to-find features) identified by Mendoza et al., (2005), were reflected in the barrier and usability problems in the third study (e.g., connectivity, functionality), caused by the participants unfamiliarity with using a tablet device. However, reasons for overall satisfaction were found to be more complex, due to the combination of device and app experiences. Participants clearly enjoyed using the device, as hedonic perceptions remained high, probably reflecting the brand image (design quality), therefore, hedonics was not a key factor in product adoption. All usability measures (PQ/ease-of-use) improved over time as users (who were mainly novice) became more familiar with using the iPad and associated apps, as found in the longitudinal studies by (Mendoza et al., 2005; von Wilamowitz-Moellendorff et al., 2006). This could indicate a \textit{usability ceiling effect}, where usability was rated good enough, so no longer influenced user judgement (see \textit{Heuristic iii}). The portability and accessibility of the device were consistent motivators of use, despite prevailing usability and barrier issues (context, connectivity). But utility (function) and interactivity of apps were the more dominate motivators, as they enabled dual-use (work/pleasure), enhanced perceived usefulness and influenced overall user satisfaction.

\textbf{Adoption:} All the participants within the longitudinal study (Chapter 5 \& 6) accepted, and continued to use the iPad throughout the 6-month study, so a short non-adoption study reported at the end of Chapter 6 was conducted to investigate the reasons for non-adoption. One of the most widely cited models used to explain technology acceptance is TAM (Davis, 1989), which focuses on changes in Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) over time. The aggregated PU scale was used to measure the iPad’s usefulness, and showed a significant increase over time, in line with PQ and ease-of-use (verbal ratings), indicating that as the iPad became easier to use, its usefulness increased. However, the non-adopter study revealed that negative perceived usefulness (PU) was the main determinate to product non-adoption, which hindered motivation and continued use, while hedonics and ease-of-use showed little difference. These findings are consistent with several studies within the medical context, for example; Yi, Jackson, Park, \& Probst, (2006) found that healthcare professionals appear to be fairly pragmatic, being more concerned with the technology’s usefulness rather than its ease of use, suggesting perceived usefulness was the most significant determinant of a technology acceptance, as reflected in the non-adopter results.
Chapter 7: Main Discussion

This was also consistent with Wu et al., (2011) who found that perceived usefulness had much greater influence on attitudes towards using mobile technology within healthcare environment, as reported by other similar studies (Chau & Hu, 2002; M. Y. Yi et al., 2006).

Although Chau & Hu, (2002) reported no social influence in their findings, the Yi et al., (2006) study on the acceptance of a PDA (personal digital assistant) by medical professionals, found subjective social norms had a positive influence on perceived usefulness and behaviour. This could explain why most of the adopter participants within the longitudinal study accepted the iPad, and only a small sample (N=7) of non-adopter participants could be found. As all adopter participants received the iPad at the same time and remained fairly satisfied with the device throughout the study, they may have shared a collective experience. According to social learning theory, (Bandura & McClelland, 1977) people learn through observing and imitating others' behaviour and attitudes. Since all the participants were part of a network of medical students who shared similar social norms and interpersonal communications, it is likely this affected their acceptance of the iPad. Especially as the medical school was pioneering the benefits of the iPad, so the less technically inclined users may have chosen to persevere with the technology through peer pressure and tacit authority from the Medical School. Another possible explanation could be based on the innovation diffusion theory (Rogers, 1995). As all the participant demographic were younger, more highly educated and therefore more attuned to technology, it could be likely they were in the innovator and early or mid adopter range, and not laggards (non-adopters), as reported by Branchau & Wetherbe, (1990) who investigated the innovation diffusion of over 500 professionals adoption of a spreadsheet software within manufacturing services.

The original TAM framework has been criticised for its simplicity (Bagozzi, 2007), with extended versions such as Venkatesh et al.'s., (2003) acceptance model (UTAUT), along with numerous other research studies adding new factors to enhance its explanatory power; such as cognitive absorption, affinity, compatibility, innovativeness, self-efficacy and anxiety (Agarwal, Sambamurthy, & Stair, 2000; Aldás-Manzano, Ruiz-Mafé, & Sanz-Blas, 2009; Deng, Turner, Gehling, & Prince, 2010; Lu & Su, 2009). Cognitive absorption has been used to measure the holistic experiences with technology (Agarwal & Karahanna, 2000), which is a multi-dimension construct drawing on notions of flow (immersion, enjoyment, control, curiosity). Deng et al., (2010) argue that cognitive absorption is an antecedent to technology satisfaction and an indicator of use, as the more users feel absorbed, the more likely they are to perceive high utility and hedonic experiences, although hedonics only exist in the presence of utility. The non-adopter participants reported spending very little time using the device, with disengagement occurring within 3 months of receiving the iPad, so they had little time to fully appreciate the experiential features of the device. Clearly they experienced low cognitive
absorption, resulting in a low perception of utility and usefulness of the device, and a negative effect on overall satisfaction, contributing to non-adoption.

One of the strongest barriers of use affecting all participants (adopters and non-adopters) was the context of use, where connectivity and use within hospital settings being reported the most adversely, impacting on the usefulness and overall motivation. This agrees with Mallat, Rossi, Tuunainen, & Öörni, (2006), who found that the benefits of using a mobile depended on the situation of use, which impact on its usefulness. Interestingly, participants’ reports of feeling uncomfortable using the iPad in front of patients and medical staff (due to it creating a patient/doctor barrier, and its association to being a leisure-based tool), hindered its usefulness. A common view within healthcare reported by Wu et al., (2011), is that mobile technology is still an emerging technology, where usefulness is dependant on its reliability to be operated at any time and place (Skomorowski, Jordan, Schroeder, & Elliott, 2013). Obviously the connectivity problems impacted upon its usefulness when it was employed within the hospital setting.

However, one of the most prominent barriers of use reported by the non-adopters was their reluctance to change, which could be attributed to their lack of affinity (perceived importance) with using the iPad. Aldás-Manzano et al., (2009) found that users affinity and compatibility to using mobile phones, along with innovativeness (openness to new technology) had a direct and positive influence on their intention to engage with mobile shopping. The non-adopters reluctance to change was heightened by their perception of the lack of support and training, suggesting they may have had a higher sense of anxiety (due to the lack of affinity and compatibility) when they first received the device. Lu & Su, (2009) suggest that anxiety (linked to skillfulness) has a negative effect on behaviour intentions to use technology. Also Branchau & Wetherbe, (1990) found that early adopters of spreadsheet software were significantly more involved in interpersonal communication, and were more active advice-seekers and advice givers. The non-adopters may have been less likely to seek advice, where a lack of training, especially in the first few weeks of acquisition may have further facilitated negative attitudes towards the utility/usefulness of the iPad, resulting in non-adoption. Furthermore, the effect of a user’s first experience with new technology has been shown to be a significant contributor to technology anxiety, which can be reduced by providing appropriate and timely training (Chua, Chen, & Wong, 1999). Therefore, careful implementation and intervention strategies are critical for the more technical adverse users, especially in the early stages of interaction (Cowan & Jack, 2011), (see Heuristic xi).

Another contributory factor for the resistance to adoption reported was the mandatory use for certain apps. Enforcing users to use technology, without giving them any other options
Chapter 7: Main Discussion

can lead to negative attitudes, as reported by (Johnson, Bardhi, & Dunn, 2008; Liu, 2012; Ram & Jung, 1991) who suggest that forced use, especially for those more technologically anxious, can threaten their freedom of choice and decisional control, impacted on user satisfaction and behavioural intentions, generate a negative UX, and ultimately hindering product adoption (Cowan & Jack, 2011; Liu, 2012; Ram & Jung, 1991) (see Heuristic xii). In situations of mandatory use, (as identified in the iPad study), options to minimise the negative effects are to provide alternative options of interaction (see also Heuristic vii), and ensure adequate training that communicates the benefits of use to the potential user.

Interestingly, the non-adopters were evenly distributed across the three Power User groups, indicating technical disposition was not an influencing factor to non-adoption. According to Zhong, (2013) Power User are not heavy or innovative users, but aim to use most of the features on a device more creatively and effectively. However, the main contributory factors for non-adoption were not due to device-usage, but to poor motivation due to their negative perception of usefulness, coupled with a reluctance to change, further exacerbated by a lack of training and forced mandatory use.

Two models were drawn from the qualitative findings (Chapter 6, Section 6.5) to illustrate the main reasons for non-adoption. The main factors influencing adoption are summarised in Figure 81.

![Figure 81. Summary model of main reasons for Adoption or Non-Adoption of the iPad](image)

Consistent with TAM literature, Perceived Usefulness (PU) was the main factor to product adoption, with the adopters having a positive PU (due to the variety of apps that enable learning support), while the non-adopters a negative PU. The aspect of time was also a critical factor, as over prolonged use the adopters PU increased, as initial usability problems were overcome or just accepted (within the first 3 months). They became more engaged (cognitively absorbed) with using the iPad; while the non-adopters reported disengaging with the iPad within 1-3 months of ownership, giving little time to overcome any usability problems, or establish engagement with their iPad. This was further hindered by the
additional barriers to adoption of user predisposition (reluctance to change), lack of timely advice, enforced usage and situation of use; while the main motivators of use were another variety that offered unique functionality, content and an alternative interactive experience, that supported learning, along with an increasing general ease-of-use.

7.4.4. Methodological Insights

One of the main aims of this research was to bridge the gap between objective and subjective approaches to UX research. Tension between the reductionist (quantitative model-based) and the holistic (qualitative experience-based) has been long established within HCI, with recent calls for the development of new data collection methods (Law, 2011). This research developed a mixture of concurrent research methods (quantitative and qualitative) to gain deeper insights into the reasons behind user judgements of product quality. Adopting a mixed-methods approach revealed some interesting methodological insights that can be used to inform future research and provides the fourth main contribution of this thesis.

Questionnaires used to capture quantitative ratings soon after participants completed a task tend to be more affect-based responses based on the participant’s initial impressions of the product (website). Post-test interviews stimulate more cognitive reflection, as participants are given more time to consider their verbal responses, and therefore may provide more detailed and refined judgements. For example in the first study (Chapter 3), participant’s quantitative ratings showed a slight advantage for the Google site in overall preference, but this shifted upon reflection when they were encouraged to articulate their opinions in the post-test interviews, where more positive comments were given to the Louvre site, which emerged as the overall preferred site. This shift of user judgement is consistent with the process model (Hartmann et al., 2008), based upon the adaptive decision making theory (ADM) by Payne et al., (1993), where user judgement can change from perceptual to more cognitive reflective phases (Lindgaard et al., 2011). This had important implication for HCI research, as the type of method used can be critical in determining the nature of the results.

This thesis argues for a mixed methods approach to HCI research where triangulation of results can establish a consistent and richer insight into user experience.

Analysis methods and questionnaire responses can also mask the diversity of individual difference. The ANOVA and regression analysis in the first two studies (Chapter 3 & 4) provided some insights into inter-site differences, where individual preferences were masked by overall judgement ratings. The quantitative results clearly showed that the interactive sites produced more positive UX, as measured by the increase in affect/hedonic responses. So the interactive features (avatar and 3D fly-through) were liked, but little difference between the interactive features was manifest in overall ratings. Although the qualitative findings
identified a polarisation in attitudes and emotion towards the interactive features (e.g., avatar), indicating that user preference impacted upon the overall UX ratings, and user preference was not tied to a specific feature. Cluster analysis was used to clarify individual differences, based on both quantitative and qualitative responses. Differences in the methods used in the longitudinal study (Chapter 5 & 6) were also noted. The quantitative measures were based on holistic ratings of the iPad (composite of app/device), while the qualitative findings enabled greater insight into the different device/apps contributions. Questionnaire ratings (app/device) showed increasing PQ (usability) and usefulness, with hedonic ratings remaining high, indicating users were positively motivated to using the iPad. However the device-related comments were negative, while app-related comments remained positive, despite high usability problems reported, indicating users forgave poor app usability. It therefore appears that the positive quality ratings of PQ and usefulness can be attributed to the apps contained on the device, indicating they were far more influential. There were clear difference in the measurement methods employed, where quantitate ratings clearly masked the device/app difference, which was only revealed by the qualitative findings, as the interviews enabled more reflection, thus more criticism of the device was found. This illustrates the benefit of combining the strengths of both methods, where questionnaires can be used to identify generic differences, while qualitative evaluations provide greater insight into the reasons for individual preferences, attitudes and emotion.

However, using a combination of different methods can produce added complexity that needs careful consideration and procurement. The longitudinal study (Chapter 5 & 6) combined a mix of concurrent methods involving questionnaires, interviews and a weekly diary-study to capture UX over time. Differences were found in results when comparing each method. The diary-study method was chosen to capture fine-grained user experiences, as they unfolded over time, yet this method required careful monitoring (reminder emails) to aid motivation, avoid participant fatigue, and ensure a consistency of data (e.g., no missing data entries). Although findings provided interesting insights into weekly activities and changes in affective responses, there was a drop in the qualitative responses over time, resulting in a measurement bias. The weekly diary-study quantitative questionnaire ratings employed a forced-choice option (to avoid missing data), however the self-reporting open qualitative questions did not, and so the quantity and quality of the qualitative dairy-study data declined considerably over time (especially in the last two months of the study), making it very difficult to access any trends over time. The highest frequency of comments was reported in the first 6 weeks of the diary study when participants were most motivated and eager to report activities. This can be attributed to either diary study fatigue, as over time participants became less diligent, or a reluctance to re-report previous experiences.
Participants’ need to be continually encouraged to continue recording their experiences, however over a long time frame and large participant group this can be very time consuming to manage.

The interviews proved more reliable, as although less frequent (x3), were conducted face-to-face, and linked to participants receiving incentives (Amazon vouchers), which enhanced their motivation. Although the weekly diary prompts collected more activity related comments (76%) than the interviews as they were closer to the event, leisure activities (e.g., Facebook, games etc.) were reported infrequently, as they may not have been salient in participants’ memory, and probably were deemed less relevant, resulting in a methodological bias. Self-reporting diary entries can miss out on participants’ tacit knowledge that can only be gained during interviews. In contrast, the interviews collected more leisure-based activities, and the final in-depth interviews generated the highest frequency of comments; including more negative affective comments, as participants were prompted (by the researcher) to disclose both good and bad experiences. Furthermore, verbal interviews enabled more reflective discourse, compared to (open-question) self-reporting diary-prompts, thus gaining more refined reflections. Therefore comparing the interview and diary results proved difficult, as they can generated contradictory experiences, as identified by Holt, (2013). Furthermore, participants reported more negative affect comments (than positive), as negative emotions generated from bad first impressions, tend to remain more salient in memory, which can lead to negative judgement (Norman, 1994), as identified in other similar studies (De Angeli et al., 2009, 2006). This suggests that the use of diary studies should not be used alone and are best used in unison with other more robust methods.

Previous inference models suggested casual influences (Hassenzahl & Monk, 2010), where perceived hedonics and pragmatic qualities influenced global evaluations of a products goodness and beauty. The models generated from the regression analysis from the first two studies (Chapter 3 & 4) showed a consistent relationship between hedonics, beauty and goodness in both directions: the general to specific (inference model), and the specific to general hypothesis (see Figure 27 & 51). This shows the problem of attributing causal influences from correlation type analysis, as the direction of causality is not always clear (Cohen, Manion, & Morrison, 2007). However the relationship between pragmatics (PQ) and goodness was less clear, showing a stronger association pre-interaction for both studies, where the decreasing effect of PQ found post-interaction may have reflected poor usability, as found in the experimental study by van Schaik, Hassenzahl, & Ling, (2012). Affect had little influence on any of the variables apart from overall experience in the second study (see Figure 51), indicating that although interactivity positively influenced affect (as reported in the ANOVA results), its effect may be temporary showing little influence on judgements of
general product quality. In the third longitudinal study, hedonics was a consistently strong predictor of affect, which strengthened over time, and became the main predictor for all product qualities (beauty, goodness, useful and satisfaction), suggesting hedonics perceptions were mediated by positive affect (see Figure 61). This conflicts with the first two studies results that showed no affect-link, which can be explained by the longer time period in the third study, where affective responses were given more time to manifest, so influencing attitudes and impacting on user judgement. The initial PQ relationship with beauty diminished over time, although reappeared to show a weak relationship with goodness. This maybe due to initial usage being more device-focused, where usability (PQ) was more important, but over time judgements became more app-focused, as more pleasurable (affect/hedonics) experiences dominated. Interpretations of these causal relationships required insights into the context-of-use, which the app/device qualitative data provided. Further studies are required to establish causal effects across varying interactive products using the differing quality perceptions (aesthetics, interactivity, usability etc.) in order to improve the generalisation of these results.

### 7.5. Heuristics

Heuristics are often associated with usability evaluation (Nielsen & Molich, 1990), where evaluative judgements are based on a set of pre-established guidelines or ‘heuristics’ (Nielsen, 1994). Similar guidelines have also provided practitioners with a comprehensive basis to evaluate usability (Agarwal & Venkatesh, 2002; Hvannberg, Law, & Lárusdóttir, 2007; Keeker, 1997). The third aim of this research was to identify generic design principles and/or evaluation heuristics. Several studies have proposed new heuristics based on aesthetic design, for example Sutcliffe, (2001) recommended a set of seven heuristics for assessing and achieving attractiveness in websites. Kim, Lee, & Choi, (2003) identified key design features that, when combined in certain ways, can produce emotionally evocative user responses, while others have produced various design recommendations to improve a websites appeal (Cyr, Head, Larios, et al., 2009; Park, Choi, & Kim, 2004; Sun, 2001). Numerous design principles have been proposed that provide practical guidelines to designing more appealing websites, (Lidwell, Holden, & Butler, 2010; Lynch & Horton, 2008; Shneiderman, 1992; Spool, Scanlon, Snyder, & DeAngelo, 1999), and interactive products (Djajadiningrat, Overbeeke, & Wensveen, 2000).

The results from this thesis show that UX is complex, as identified in the expanded models, where context, task, product quality criteria, individual difference play a role in forming user experience, all of which is subject to change as the experience progresses over time. To
Chapter 7: Main Discussion

facilitate understanding a set of twelve high level principles or heuristics are proposed. These are based on the main findings reported in this thesis (see cross-references), which draws from the literature review, with the main purpose to assist design practitioners and guide evaluation.

i. **Task framing effects:** The first two studies employed more relaxed, leisure based tasks, so interactivity was favoured above usability. User judgement has been shown to be more susceptible to task framing effects, with serious tasks favouring usability and utility, while more fun based tasks, aesthetics and interaction criteria were preferred (Hartman, 2008).

ii. **Product framing effect:** Work related products select usability and utility attributes, while games or entertainment type products prefer aesthetics, pleasure and enjoyment more (Hassenzahl et al., 2010). The third longitudinal study took place in the field (medical learning), so no set tasks were conducted. However, the iPad was promoted as a work-based tool, so utility and usefulness dominated over time, thus illustrating that prior information could promote framing effects (Hartmann et al., 2007). Although, it was the apps contained on the iPad that enabled interactive experiences (CMC and social), providing engaging experiences, which impacted on long-term user satisfaction.

iii. **Usability ceiling effects:** When usability is adequate, this criterion no longer becomes an important attribute in users’ judgement of a product. As identified in the longitudinal study, where usability improved over time due to product familiarity and long-term usage, which concurs with (Mendoza et al., 2005; von Wilamowitz-Moellendorff et al., 2006). Likewise, the ‘what is beautiful is usable’ rule (i.e., halo effect) where beauty (aesthetic) ratings can positively influence pragmatic judgements (usability), impacting on the overall rating of the product (De Angeli et al., 2006; Tractinsky, Katz, & Ikhar, 2000).

iv. **Brand Image:** In the longitudinal study, the most favoured branded medical apps were rated for their trusted content, despite their poor usability ratings. Brand image has been found to positively influence usability ratings of a product (De Angeli et al., 2009), especial on initial encounter. This heuristic is dependant on the strength, quality and visibility of the brand image, which can promote trust and loyalty.

v. **Support individual difference:** Users predisposition (preference and technology disposition) can impact on user judgement (Marathe et al., 2007), as identified in the user-group differences shown within this research. Three user groups showed contrasting preferences for different interactive features, while users technical disposition showed that the more technical inclined users were less sensitive to usability problems. Previous research has shown some users have higher personal innovativeness (Wu et al., 2011) and curiosity (Magni et al., 2010) towards technology, and prior experience can influence user judgement (Hartmann et al., 2008). This heuristic emphasis the advice “**know your**
Chapter 7: Main Discussion

*audience’s preferences and expectations*” (Hartmann et al., 2008), where pre-screening of potential users and participants to identify subgroup preferences and user-diversity can be used to provide more fine-graded personalised design solutions (Karapanos, Martens, & Hassenzahl, 2009).

vi. **Use of Avatars:** It is critical when using avatars that they are designed and implemented well, offer clear and easy functionality that give users control, along with the option to not interact with an avatar (McGoldrick et al., 2008), so to avoid frustration. Test piloting avatars with a target user audience would provide guidance to the appropriate selection of avatar types (e.g., gender, realism and anthropomorphism), and level of functionality (e.g., chat/facial expression etc.), for the required context domain (Khan & Sutcliffe, 2014).

vii. **Alternative Interaction:** Enhanced interactive features that provide alternative methods of navigation (e.g., 3D fly-through and active-object links) was shown to increase user engagement (as identified in the first two studies), but can cause frustration due to usability problems if poorly designed (Sutcliffe & Gault, 2004). These effects need to either employ an engaging interactive metaphor (e.g., fly through in Google), or provide a rich combination of interactive functions (e.g., avatars, 3D graphical world, active objects in Louvre); thereby giving users alternative options to traditional menu-link navigational methods.

viii. **Keep interactivity fresh:** As users gain more experience with technology, early forms of interactivity (e.g., avatar, games etc.) are likely to become ordinary, prompting boredom and rejection (Fogg, 2002). To keep the interactive experience fresh and engaging over long-term use, adaptivity (the capacity for a system to adapt according to users’ behaviour), (Benyon, 1993; Höök, 1997), is suggested. This would enable the interactive features (e.g., avatars, interactive media etc.) to be revised regularly to avoid boredom and frustration, allowing for different levels (novice/expert) of user experience (Qiu & Benbasat, 2005). *(see also customisation (ix) and user control (x)*)

ix. **Customisation:** Adaptability refers to the capacity of a product to enable users’ to customise and adapt a product to their needs (Thevenin & Coutaz, 2002), which can be an important motivator for longer-term use, enhancing user engagement and product ownership (Hartmann et al., 2007; A. Sutcliffe, 2009). The iPad enabled a certain level of customisation/personalisation through self-selection of the variety different apps, which satisfied different users needs and use (work/leisure), enhancing long-term satisfaction.

x. **User Control:** High priority should be given to user control, enabling users the freedom to control their interactive experience with minimum support. User control is an important aspect within interactivity and can lead to satisfaction, save time and enhance trust (Cyr, Head, & Ivanov, 2009; Sohn & Lee, 2005; Teo et al., 2003). However, the extent of the
interface control options should be limited and transparent so not to disrupt the users’ interactive experience (C. Yi, Jiang, & Benbasat, 2011).

xi. **Provide Training:** Offering appropriate and timely training when implementing new technology can enhance user involvement, improve user confidence and reduce anxiety, leading to greater user satisfaction (Torkzadeh & Dwyer, 1994). Initial training sessions should focus on how new technologies (e.g., popular apps) can provide benefit (usefulness, effectiveness and efficiency) to the user, rather than on general familiarisation training, that focus on operating the technology (Chau & Hu, 2002).

xii. **Avoid Forced Use:** User satisfaction can be affected by forced mandatory use leading to negative UX and resistance to adoption, especially for users with high technology anxiety (Liu, 2012; Ram & Jung, 1991). Providing flexibility, user choice and alternative ways to interact with a product can improve users perceptions of control and behavioural intention towards technology acceptance (Botti, McGill, & Iyengar, 2003; Sundar & Marathe, 2010). The use of social media and cooperative learning in the early stages of acquisition, that encourage peer groups to share their experience, knowledge and positive usage experience may also improve attitudes towards new technology and its adoption (Ram & Jung, 1991).

### 7.6. Conclusion and Future Research

The four main contributions of this research were to identify the importance of **interactivity**, clarify three differing **user-groups**, investigate the transient changes of **UX over time**, and provide some valuable **methodological insights** into using a mixed methods approach.

This thesis has expanded the process model of user quality judgement by demonstrating the importance of interactivity in positively influencing UX. Affective and hedonic ratings increased as a result of interaction, demonstrating the powerful effect of interactivity short-term, and showed marked differences for websites that contained interactive features, despite the presence of usability problems. Technologies that offered specific functionality that facilitated interactive experiences (e.g., apps), prompted long-term usefulness and positive UX, especially those that mediated action such as CMC and social interaction.

Results from the inter-variable regression analysis from the first two studies (Chapters 3 & 4) showed causal links between hedonics, beauty and goodness, in both specific to general, and general to specific directions, with the latter supporting the inference models (Hassenzahl, 2004), although the pragmatic-to-goodness link was less clear, only being apparent pre-interaction. Variations occurred between the different interactive sites and tasks, which can
Chapter 7: Main Discussion

mask user-diversity, where judgements may not be consistent across user populations and products, so subject to fixed effect fallacy (Monk, 2004). Future research using different interactive sites/features, across different domains would provide interesting comparisons to these studies and improve the generalizability of the findings. Few effects on overall product judgement were found in the first two studies (Chapter 3 & 4), where affect showed little effect on overall quality judgements, which could be due to the limited interaction time given for each website. Future research could investigate longer interactive sessions where more marked long-term effects maybe found. In contrast to the first two studies, a strong association between affect and hedonics was shown in the iPad longitudinal study (Chapter 5 & 6), which increased over time. Where early-use showed a PQ (usability) influence, when perceptions were more device-focused (usability), but over time judgements became more app-focused, where affect/hedonic experiences became more dominate. The specific to general model supports the hypothesis that initial perceptions of PQ and hedonics influenced affect and attitude over time, which impacts on overall judgement of the products quality.

This research provides a new contribution to the field by adopting a user typology perspective, which bridges the generic population-level quantitative studies, with the individual-level qualitative studies that focus on ethnographic interpretation of experience. User differences were identified using cluster analysis across a combination of quantitative and qualitative ratings, to reveal three sub-groups that was categorised according to their interactive design preference, or their predisposition towards technology. Prior experience impacted on users attitudes towards new technology, influencing their judgements, with the more technical disposed users giving consistently high ratings for usability and hedonics, indicating they were more tolerant of usability problems and enjoyed using the device more. However, over time, all three user-group ratings converged for most items, with the less technical disposed group benefiting the most, indicating that user group membership is transient. The area of individual difference clearly deserves more attention in future research.

Few studies have focused on the transitional effects of UX over time. This thesis found that the main incentives of long-term use were perceived utility (usefulness) and usability of the product, while hedonic ratings remained the same indicating the users were generally satisfied with the device. The complexity of application and device revealed that user satisfaction was determined by the applications contained on the device, despite poor usability. Popular apps were favoured for their flexibility to support dual-use, provide specific functionality or content, and enable interactive experiences. The main determinates to non-adoption were negative perceived utility and the users reluctance to change, further facilitated by a lack or advice or training along with a negative response to force mandatory
use. Further research comparing different contexts-of-use with UX qualities could provide some fruitful new insights in the effects of long-term user experience and satisfaction.

The mixed methods approach adopted by this thesis provided insights into the reasons for users quantitative ratings of product qualities, highlighting the differences between user preferences and their predispositions towards technology. Furthermore, interesting differences in user judgement were shown between the qualitative and quantitative data analysis methods. User preference judgements appeared to shift from questionnaire ratings to post-test interviews, identifying a methodological bias. This can be explained by the differing cognitive responses these methods elicit, where questionnaires are based on the immediate perceptual response to a product, so a more affect-based style of judgement is given, whereas interviews stimulate more cognitive reflection, thus producing a more detailed and longer lasting judgement. This highlights the need for more reliable and valid research methods that can capture both ‘in-the-moment’ experiences (immediate), as well the long-term temporal experiences (UX over time), which need to be quick, practical and easy to apply. By developing new alternative methods that triangulate a variety of high-level statistical methods along with rich ethnographic techniques to capture contextual patterns as they unfold over time, would provide new insights into UX research. This thesis advocates using a mixed methods approach, however, further understanding at how best to combine existing and new alternative methods would facilitate future UX research. Twelve high level heuristics were proposed as an outcome of this research with the aim to provide generic principles to aid design and evaluation of new or existing interactive products. However, they are tentative and so contingent on further evaluation and open to future research.
Bibliography


Bibliography


Cooper, A. (1999). The inmates are running the asylum: [Why high-tech products drive us crazy and how to restore the sanity]. Vol. 261. Indianapolis: Sams


Bibliography


Bibliography


248
Bibliography


Bibliography


Holt, T. B. (2013). *Capturing the User Experience: Experiences with diaries and interviews*.


ISO (2001). 9126 DIS. In *Software Product Evaluations-Quality characteristics and guidelines for their use*.


Lee, S., & Koubek, R. J. (2010b). Understanding user preferences based on usability and aesthetics before and after actual use. *Interacting with Computers, 22*(6), 530–543.


Lidwell, W., Holden, K., & Butler, J. (2010). *Universal principles of design: 125 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design.* Rockport publishers.


Bibliography


Magni, M., Susan Taylor, M., & Venkatesh, V. (2010). "To play or not to play": A cross-temporal investigation using hedonic and instrumental perspectives to explain user intentions to explore a technology. *International Journal of Human-Computer Studies, 68*(9), 572–588.


Bibliography


Norman, D. A. (1998). *The invisible computer: why good products can fail, the personal computer is so complex and information appliances are the solution*. Cambridge, MA: MIT Press.


Bibliography


Bibliography


Short, J., Williams, E., & Christie, B. (1976). Theoretical approaches to differences between media. The social psychology of telecommunications (pp. 61–76). London: Wiley and Sons Ltd.


Bibliography


Bibliography


Experience – Towards a unified view, Workshop at the 4th Nordic Conference on Human-Computer Interaction (pp. 74–78).


Appendices

APPENDIX 1: Interactive Tasks for Study 1 & 2

Study One

Task One: Orientation: Simple navigational task where participants were asked to navigate each website making sure they used each of the interactive features contained in that site.

Task Two: Examine: Additional task that directed the participants to examine an art object using specific interactive features (zoom, audio guide, 3D animation of an object).

Task One Louvre (4 pages) (Interact with: avatar x2 active objects x2)
Page 1: Homepage >Click on Cartoon Avatar that pops up on home page
Page 2: Interactive Room with animated cartoon and objects > click on any animated object
Page 3: Artefact is displayed with avatar giving a description of this > click on animated trap door (on the left of the page)
Page 4: Takes the user back to the interactive room (page 2)

Task Two Louvre (4 pages) (Interact with avatar x1, animations x1-2, zoom x1)
Page 1: Homepage >Click on ‘A Closer Look’ from menu at top of page
Page 2: Scroll down to bottom of page to click on >Page 2
Page 3: Click on > ‘A closer look at the Consecration of the Napoleon’ (image at bottom of page)
Page 4: Click on > View Feature (located under the text of the image) and use the tools under the picture to explore the picture (animation of picture with audio voice over, zoom feature)

Task One Google (5 pages) (interact with 3D Space x2, active map x1)
Page 1: Homepage >Select ‘The Palace of Versailles’ from left hand scroll down menu
Page 2: Select ‘Explore the Museum’ option at bottom of page
Page 3: Brings up the 3D view of the museum with the arrow keys to explore
Page 4: Click on ‘Navigate Floor Plan’ top right to open floor plan
Page 5: Click on ‘Hall of mirrors’ from the floor plan> brings up Hall of mirrors room.

Task Two Google (5 pages) (interact with 3D Space x1, zoom x1)
Page 1: Homepage >Select ‘The Metropolitan Museum of Art’ from left hand scroll down menu
Page 2: Select ‘Explore the Museum’ option at bottom of page
Page 3: Navigate to the picture ‘The Harvesters by Pieter Brugel’ click > on the small cross +
Page 4: (A 2D image of the picture appears) Use the zoom tool to zoom in and out of the picture
Page 5: Click on Description at top of page to reveal information about the picture

Task One National (5 pages) (Interact with basic map x1)
Page 1: Homepage >Click on Paintings (from top menu)
Page 2: Click > on Paintings room by room (located in middle of page)
Page 3: Brings up Floor Plan > Click on any room of your choice
Page 4: A page of paintings and descriptions is appeared
Page 5: Click> on ‘Floor plan’ at top left menu to return to map

Task Two National (5 pages) (Interact with basic zoom x1)
Page 1: Homepage >Click on Paintings (from top menu)
Page 2: Click > on Paintings room by room (located in middle of page)
Page 3: Click on > ‘Room 45: Van Gogh and Cezanne’
Page 4: Scroll to find thumbnail of ‘Surprised by Henri Rousseau’ click> on thumbnail
Page 5: A large image of picture is shown – use the zoom tool (right) to examine the picture

Study Two

Interactive Task: participants were asked to visit 3 pages of the website making sure the experienced all the interactive features within the site (homepage, interactive map and gallery).

Interactive Task (4 pages) same for both sites
Page 1: Homepage >Manchester
Page 2: Select > Culture and Location - explore this page
Page 3 Select > Campus Map (from top menu) explore this page
Page 4: Select > Gallery (from top menu) explore this page
**APPENDIX 2: The Affect Scale Questionnaire**

**The Affect Scale Used in Study One & Two (Chapters 3 & 4)**

M1. Can you please rate OVERALL your current MOOD on the scale below?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Calm</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
</tbody>
</table>

M2. Please rate the website against your experiences on the following ASPECTS:

<table>
<thead>
<tr>
<th>ASPECTS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Fun</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Attractive</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Curious</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Interesting</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Pleasurable</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Absorbing</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Exciting</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Engaging</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
</tbody>
</table>

**The expanded Affect Scale Used in Study Three (Chapters 5 & 6)**

A1. Please rate your EXPERIENCE of using the iPad (so far) on the following ASPECTS:

<table>
<thead>
<tr>
<th>ASPECTS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Fun</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Enjoyable</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Attractive</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Curious</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Interesting</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Pleasurable</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Absorbing</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Exciting</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Not Engaging</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Very Frustrating</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>Very Boring</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
</tbody>
</table>

*(Includes Negative Effect items Frustrating and Boring, plus Enjoyable)*
APPENDIX 3: The Website/Design Quality Scale

The Design Quality Scale Used in all Studies
Taken from AttrakDiff2 Scale (Hassenzahl & Monk 2010; Hassenzahl 2004).

P1. How well did the website hold your ATTENTION? (Please rate from 1 – 7 below)

<table>
<thead>
<tr>
<th>*Hardly At All</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>All the Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

P2. Please rate the following AESTHETIC qualities based upon your perceived judgment:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cheap</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Premium</td>
</tr>
<tr>
<td>Confusing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Clearly Structured</td>
</tr>
<tr>
<td>Tacky</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stylish</td>
</tr>
<tr>
<td>Complicated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Simple</td>
</tr>
<tr>
<td>Dull</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Captivating</td>
</tr>
<tr>
<td>Amateurish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Professional</td>
</tr>
<tr>
<td>Unpredictable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Predictable</td>
</tr>
<tr>
<td>Conservative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Innovative</td>
</tr>
<tr>
<td><strong>Challenging</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Easy</td>
</tr>
<tr>
<td><strong>Impractical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Practical</td>
</tr>
<tr>
<td><strong>Isolating</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Integrating</td>
</tr>
<tr>
<td><strong>Commonplace</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
</tr>
<tr>
<td><em>Unimaginative</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Creative</td>
</tr>
<tr>
<td><em>Cluttered</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Spacious</td>
</tr>
<tr>
<td><em>Poor Colour</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Good Colour</td>
</tr>
<tr>
<td><em>Unbalanced Design</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Balanced Design</td>
</tr>
<tr>
<td><em>Poor Images</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Good Images</td>
</tr>
<tr>
<td><em>Design Discord</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Design Harmony</td>
</tr>
</tbody>
</table>

A3. Can you rate your OVERALL view of the website for the following two aspects?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Good</td>
</tr>
<tr>
<td>Ugly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Beautiful</td>
</tr>
</tbody>
</table>

Shows the amendment made between the first two studies and study three:
*Items dropped from the third Longitudinal Study (Chapter 5 & 6)
**Items added to the scale for the third Longitudinal Study
APPENDICES

APPENDIX 4: Usability, Content, Quality and Overall Experience Scales

The Usability Scale used in all three studies
Taken from existing usability scales (Lavie & Tractinsky 2004; Tractinsky & Zmiri 2006).

F8. Please rate the USABILITY (EASE OF USE) within the following aspects:

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Slightly Disagree</th>
<th>Neutral</th>
<th>Slightly Agree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convenient to Use</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Clear Functionality</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Easy to Learn</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Easy to Navigate *</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

The Contents & Service Quality Scale used in Study Two
Taken from (De Angeli et al. 2009; Bernier 1996; Lavie & Tractinsky 2004).

F9. Please rate the CONTENT found within the following aspects:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good Quality Content</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Content is Relevant</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Too Much Content</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

F10. Please rate the SERVICE QUALITY within the following aspects:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliable Information</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Can Trust the Site</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Contains No Mistakes</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

The Overall Experience Scale used in all the first two studies
Taken from O’Brien, (2010)

F11. Please rate your OVERALL view of the website:

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Slightly Disagree</th>
<th>Neutral</th>
<th>Slightly Agree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would visit this website again</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I would recommend this website to friends</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>My experience of using the site was stimulating</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
APPENDIX 5: The Immersion (Flow) and Presence Scale

The Immersion/Presence Scale used in Study 1 and 2 only
Taken from various sources: (Witmer & Singer, Michael 1998; Trevino & Webster 1992; Jennett et al. 2008; Csikszentmihalyi 1975).

F0. Please rate the website against your experiences on the following ASPECTS:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOT in Control</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>In Control</td>
</tr>
<tr>
<td>Challenged</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Easy</td>
</tr>
<tr>
<td>Time Flew</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Time Went Slow</td>
</tr>
<tr>
<td>Absorbed</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Detached</td>
</tr>
<tr>
<td>Natural</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Artificial</td>
</tr>
<tr>
<td>Compelling</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Boring</td>
</tr>
<tr>
<td>Stimulating</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Frustrating</td>
</tr>
<tr>
<td>Fast Pace</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Slow Pace</td>
</tr>
<tr>
<td>Focused</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Un-Focused</td>
</tr>
<tr>
<td>Oblivious to outside events</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>Aware of outside events</td>
</tr>
</tbody>
</table>

*Amended from Study One to a bi-polar scale.

F4. To what extent were you aware of yourself in your surroundings?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little</th>
<th>Moderately</th>
<th>Quite a bit</th>
<th>Very Much So</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

F5. To what extent did you feel consciously aware of being in the real world when using the site?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little</th>
<th>Moderately</th>
<th>Quite a bit</th>
<th>Very Much So</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

F6. To what extent did you notice events taking place around you?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little</th>
<th>Moderately</th>
<th>Quite a bit</th>
<th>Very Much So</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

F7. To what extent did you feel the website was something you were experiencing rather than something you were doing?

<table>
<thead>
<tr>
<th></th>
<th>Not at all</th>
<th>A little</th>
<th>Moderately</th>
<th>Quite a bit</th>
<th>Very Much So</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
APPENDIX 6: The Perceived Usefulness Scale (PU) & Overall Experience

The PU Scale used in Study Three (Longitudinal Study)
Pu scale was taken from Technology Acceptance Scale (TAM) by Davis (2008).

E1. Please rate your EXPERIENCE (so far) for the iPad’s USEFULNESS for each statement:

<table>
<thead>
<tr>
<th>USING the iPad has...</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Slightly Disagree</th>
<th>Neutral</th>
<th>Slightly Agree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Improved my PERFORMANCE</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>2) Increased my PRODUCTIVITY</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>3) Enhanced my EFFECTIVENESS</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>4) Improved the QUALITY of my work</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>5) Saved me TIME</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>6) Make it EASIER to complete my work</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>7) Been USEFUL for my course learning</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>8) Involved little EFFORT to LEARN how to use</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Overall Experience Scale used in Study Three (Longitudinal Study)
Experience/Satisfaction scale based upon (Cyr et al. 2010).

E7. Please rate your OVERALL EXPERIENCE (so far) of using the iPad?

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Slightly Disagree</th>
<th>Neutral</th>
<th>Slightly Agree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel MOTIVATED to continues to use my iPad</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I would RECOMMEND the iPad to my friends</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>My experience of using the iPad is ENJOYABLE</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>OVERALL, I am very SATISFIED with my iPad</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
APPENDIX 7: Diary Study Mini-Questionnaire

Q1. Can you please select what ACTIVITIES you have used your iPad for to SUPPORT your LEARNING? Please indicate HOW OFTEN you have carried out each activity THIS WEEK.

<table>
<thead>
<tr>
<th>Select One Option</th>
<th>Several Times a Day</th>
<th>Once a Day</th>
<th>Several Times a Week</th>
<th>Once this Week</th>
<th>Not this Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sending / Checking your emails</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Browsing the Internet (For Information Search)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Making Notes (Word Processing)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Accessing the on-line VLE (Blackboard)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Reading (Books, PDFs, Journal Papers, etc)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Collaborating within Groups</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Q2. Can you please select what ACTIVITIES you have used your iPad for LEISURE and ENJOYMENT? Please indicate HOW OFTEN you have carried out each activity THIS WEEK.

<table>
<thead>
<tr>
<th>Select One Option</th>
<th>Several Times a Day</th>
<th>Once a Day</th>
<th>Several Times a Week</th>
<th>Once this Week</th>
<th>Not this Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browsing the Internet (News, Shopping, etc)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Social Networking (Facebook, Twitter, Skype, etc)</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Watching Videos</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Playing Games</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Taking / Sharing Photos</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Listening to Music</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Are there any other ACTIVITIES you have used your iPad for not listed above?

- For Learning: 
- For Pleasure: 

Q3. Please LIST any NEW APPS you have downloaded onto your iPad THIS WEEK:

Q4. Please LIST any APPS you have DELETED or STOPPED using THIS WEEK:

Q5. Please RATE how you currently FEEL about your IPAD THIS WEEK?

<table>
<thead>
<tr>
<th>Enjoyable</th>
<th>Useful</th>
<th>Engaging</th>
<th>Boring</th>
<th>Frustrating</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Please DESCRIBE up to 4 ACTIVITIES (GOOD/BAD) while using your iPad? How did you FEEL about this experience? You can enter the text description here, or you can use EVERNOTE (or similar APP) to record an AUDIO or TEXT description and EMAIL IT to the researcher THIS WEEK.

Thank You.

- Good Experiences?
- Good Experiences?
- Bad Experiences?
- Bad Experiences?
APPENDIX 8: Interview Questions

Semi-structured interview questions were used as a guide within all the three studies.

**Study One & Two**

- What are the First TWO things you remember most about the websites?
- What things did you LIKE about any of the websites?
- Why?
- Was there anything you DID NOT LIKE about any of the websites?
- Why?
- Did you experience any PROBLEMS with any of the websites?
- Which websites do you PREFER and why?
- Did you notice any difference between the sites for:
  - Aesthetic Design
  - Content
  - Ease of Use
  - Engagement Experience
  - Brand
- Can you RANK them in order of PREFERENCE?
- What is your BACKGROUND (work/study)? Art Appreciation?

**Study Three**

The same questions were used for Early/Mid Study Interviews, but additional questions were included in the Final Interview, that allowed flexibility to explore any issues that arose.

All interviews began with Verbal Ratings: *(Ease-of-Use, Attractive, Enjoyment, Usefulness, Engaging, Frustrating, Boring, Satisfaction)* on a 1-7 scale.

- What ACTIVITIES have you used the iPad for?
- How does the iPAD differ from using a PC?
- What new or existing APPS have you been using?
- What are your top 2-5 FAVOURATE APPS and why?
- Do you think the iPad has changed the way you do things? (Learning/Leisure)
- What do you ENJOY the most about using your iPad? (GOOD Experiences)
- Have you encountered any PROBLEMS while using the iPad? (BAD Experiences)
- How do you find using the TOUCHSCREEN interface? (Intuitiveness)
- Have you PERSONALISED your iPad?
- Can you RATE your OVERALL EXPERIENCE since receiving your iPad (1-7)

Additional Final Interview Questions:

- Retrospective Graphs – Past experiences (affect/activities)
- What MOTIVATES you to continue to use your iPad?
- What BARRIERS have you experienced that have hindered you using the iPad?
- How has using the iPad ADAPTED into your Work/Leisure life?
- List 2/3 things you most like about using the iPad?
- List 2/3 things you do not like about using the iPad?
APPENDIX 9: Diary Study Time-Line Graphs

Aggregated ratings for Affect (top), Learning (middle) & Leisure Activities (bottom).

Affect Likert Ratings: (1=Low, 7=High)

Activity Likert Scale Ratings: (1=Not this week, 2=Once this week, 3=Several times this week, 4=Once a day, 5=Several times a day)
APPENDIX 10: Activity Usage

Activity Graphs: shows changes in Learning and Leisure activities over time, across 3 devices (computer/laptop, mobile phones, and iPad). The Significant differences are also shown.

**Learning Activities**: email, Internet, notes, virtual learning environment (VLE), reading and group use.

**Leisure Activities**: Internet, social networking, videos, games, photos and music.

Key:
- ⋆ = Computer/Laptop
- ◆ = iPad
- □ = Mobile Phone

Significant Device Differences:
- PC & Phone: ○ = p<.05, ◆= p<.01, □ = p<.001
- iPad & PC: ◆ = p<.05, ◆= p<.01, □ = p<.001
- iPad & Phone: □ = p<.05, ◆= p<.01, □ = p<.001
- Time Decrease p<.01: (Early-Final /Mid –Final)
Appendices

Summary of activity usage over the three devices showing the significant differences, rank order of means and time pattern changes.

<table>
<thead>
<tr>
<th>Scale Items</th>
<th>RANK ORDER</th>
<th>TIME PATTERNS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>iPad</td>
<td>PC</td>
</tr>
<tr>
<td><strong>Work Activity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Email**</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>W Internet***</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Notes***</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>VLE***</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Reading***</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Groups***</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Leisure Activity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L Internet**</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Social Networking</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Video***</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Games***</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Photos***</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Music</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medical Apps***</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Key: *** = p < .001, ** = p < .01, = p < .05, ↓ = Sig Decrease, ↑ = Non Sig Patterns, ≈ = No Change

**Device Usage Summary Table:** Showing Sig. Difference, Rank Order & Time Patterns