Designing a Value Centred Inspection Method for Children

by

Obelema T S Akobo WODIKE

A thesis submitted in partial fulfilment for the requirements of the degree of Doctor of Philosophy at the University of Central Lancashire

May, 2016

Student Declaration Form

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution.

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work.

Signature of Candidate: ______________________

Type of Award: Doctor of Philosophy

School: Physical Sciences and Computing

ABSTRACT

There are currently several evaluation methods suited to children within the Child-Computer Interaction (CCI) community. However, these methods are user based leaving a gap in inspection method suited to children, that is, where children will act as the inspection method evaluators. This research focused on how to design an effective and easy to use inspection method where children will perform the evaluation based on their values.

To ensure that the above is met, a user centred approach and a mixed methodology was explored and finally resulted in the creation of the **I**nspection **M**ethod for **Ch**ildren (IMCH) with an accompanying guideline. This six stages method could be used by CCI researchers as a guide to develop similar methods for children, by industries to perform inspection method evaluation with children on technologies designed for children and could be used by designers to gather design criteria for children's technology. The process undertaken within this research to develop the new method is also novel and could be adapted by new and old researchers when adapting method to suit children.

Future work will focus on carrying out evaluation with wider age range of children in the method to ensure suitability of the method for more children. Comparative studies of the method with other usability method to determine the effectiveness of the method and as a refinement process to produce a validated and refined IMCH method.

Table of Contents

1 CHAPTER ONE: INTRODUCING THE THESIS 1

1.1 Introduction 1

1.2 Research Motivation 1

1.3 Research in Context 4

1.4 Aim 5

1.4.1 Research Questions 5

1.4.2 Objectives 6

1.5 Contribution/Novelty 6

1.5.1 Major 6

1.5.2 Minor 6

1.6 Methodology 6

1.7 Thesis Structure 9

1.8 Conclusion 10

1.8.1 Contribution of This Chapter to Thesis 10

2 CHAPTER TWO: UNDERSTANDING CHILDREN IN RESEARCH 11

2.1 Introduction 11

2.2 Children 11

2.3 Children’s Developmental Stages 12

2.3.1 Piaget’s Cognitive Development theory 13

2.3.1.1 The Preoperational Stage (age 2 to 7years) 16

2.3.1.2 The Concrete Operational Stage (age 7 to 12years) 16

2.3.2 Vygotsky’s Social Interaction Approach 16

2.3.2.1 More Knowledgeable Other (MKO) 18

2.3.2.2 Zone of Proximal Development (ZPD) 18

2.3.3 Bandura’s Social Learning Theory 19

2.3.4 Hanna *et al's* work ... 21

2.3.4.1 Elementary school aged children (6 to 10year olds) ... 21

2.3.4.2 Middle school aged Children (11 to 14year olds) ... 21

2.3.5 Markopoulos and Bekker's work ... 21

2.3.5.1 The emerging-autonomy stage (ages 3–7) ... 22

2.3.5.2 The rule/role stage (ages 8–12) ... 23

2.4 Child-Computer Interaction ... 23

2.4.1 The Origin of CCI ... 23

2.4.2 Defining CCI ... 24

2.4.3 Popularity of CCI and Origin of IDC ... 24

2.4.4 CCI Involvements, People and Practices ... 25

2.4.5 The Subjects of CCI ... 26

2.4.5.1 Input Devices with Children ... 26

2.4.5.2 Design with Children ... 27

2.4.5.3 Evaluation with Children ... 28

2.4.6 Challenges of CCI ... 29

2.4.7 The Future for CCI ... 30

2.5 Conclusion ... 31

2.5.1 Contribution to Thesis ... 31

3 CHAPTER THREE: UNDERSTANDING EVALUATION METHODS ... 32

3.1 Introduction ... 32

3.2 Evaluation Methods in HCI ... 32

3.2.1 Usability Evaluation Methods (UEMs) ... 33

3.2.1.1 User Based (Testing) or Empirical Evaluation Method ... 34

3.2.1.2 Inquiry Based Evaluation Method ... 34

3.2.1.3 Analytical or Expert or Inspection Based Evaluation Method ... 34

3.3 Evaluation Methods with Children ... 35

3.3.1 Fun Toolkit.....36
3.3.1.1 Smileyometer.....36
3.3.1.2 Fun Sorter.....36
3.3.1.3 Again-again Table.....37
3.3.2 Contextual Laddering Method.....37
3.3.3 Think Aloud with Children.....38
3.3.4 This or That (Pairwise Comparison Scale) Method.....38
3.3.5 The Problem Identification Picture Cards (PIPC).....39
3.3.6 MemoLine.....39
3.4 The State of Usability Inspection Method (UIM) with Children.....40
3.4.1 Heuristic Evaluation (HE) with Children.....41
3.5 Measuring the Success or Effectiveness of UEMs.....43
3.5.1 Defining Method Success for this Thesis.....46
3.6 Conclusion.....47
3.6.1 Contribution of This Chapter to This Thesis.....48

4 CHAPTER FOUR: RESEARCH METHODOLOGY AND DESIGN.....49
4.1 Introduction.....49
4.2 User Centred (UC) Approach.....50
4.2.1 UC within this Research.....50
4.3 Research Approaches.....50
4.3.1 Mixed Method Approach.....51
4.3.1.1 Mixed Method Research Design.....52
4.3.1.2 Typologies of Mixed Method Designs.....54
4.4 Other Research Concerns.....55
4.4.1 Sampling.....55
4.4.1.1 Considerations for Sequential Sampling.....56
4.4.1.2 Types of Mixed Method Sampling.....56

4.4.1.3 Sampling for this Research 57
4.4.2 Data Type 59
4.4.3 Reliability 59
4.4.4 Validity 60
4.4.5 Ethics 61
4.5 Methods for the Research Stages 63
4.5.1 Literature Review 63
4.5.2 Stage one – Identifying a gap in UEM with children 63
4.5.3 Stage two – Assessing Children's Performance in a UIM Evaluation 64
4.5.3.1 Observation Method 64
4.5.4 Stage 3 – Measures to improve UIM to work with children 65
4.5.4.1 Specific Methods applied in this stage 65
4.5.4.2 Design Session with Children 66
4.5.4.3 Designing the IMCH 67
4.5.4.4 Testing the newly Designed Method 68
4.5.5 Stage 4 – Write up Guideline and Finish 68
4.6 Conclusion 68
4.6.1 Structure of Chapter Flow towards Creating the New (IMCH) Method 68
4.6.1.1 Chapter 5 68
4.6.1.2 Chapter 6 68
4.6.1.3 Chapter 7 69
4.6.1.4 Chapter 8 69
4.6.1.5 Chapter 9 69
4.6.1.6 Chapter 10 69
4.6.2 Chapter Contribution to Thesis 70

5 CHAPTER 5: PILOTING HEURISTIC EVALUATION METHOD 72
5.1 Introduction 72

5.2 Pilot Study 1 (Heuristics with adult experts) 73
5.2.1 Method 74
5.2.1.1 Participants 74
5.2.1.2 Apparatus 74
5.2.1.3 Procedure 76
5.2.1.4 Data Analysis 77
5.2.2 Results 77
5.2.2.1 Evaluators’ Data 77
5.2.2.2 Observers’ Data 79
5.2.3 Discussion 81
5.2.4 Conclusion 81
5.3 Heuristic Evaluation with children 82
5.3.1 Method 82
5.3.1.1 Participants 82
5.3.1.2 Apparatus 83
5.3.1.3 Procedure 83
5.3.1.4 Data Analysis 84
5.3.2 Results 86
5.3.2.1 Result from Evaluators’ Data Analysis 86
5.3.2.2 Result from Observers’ Data Analysis 87
5.3.3 Discussion 88
5.4 Conclusion 90
5.4.1 Contribution of Chapter to thesis 91

6 CHAPTER SIX: GATHERING TEACHERS’ INPUT 93
6.1 Introduction 93
6.2 Method 93
6.2.1 First Focus Group (FG) Study with School Teachers 94

6.2.1.1 Participants 94

6.2.1.2 Apparatus 94

6.2.1.3 Procedure 95

6.2.1.4 Data Analysis 96

6.2.2 Second Focus Group Study with School Teachers 97

6.2.2.1 Participants 97

6.2.2.2 Apparatus 97

6.2.2.3 Procedure 98

6.2.2.4 Second Focus Group Data Analysis 98

6.3 Result for FG1 and 2 98

6.3.1 Input within a wider evaluation context 98

6.3.1.1 Teaching children to carry out the evaluation process 98

6.3.1.2 Preventing bias from their peers 99

6.3.1.3 Preventing children from being engrossed in the game during the evaluation process 99

6.3.1.4 Teaching children to empathize with younger children whilst carrying out the evaluation process 100

6.3.1.5 Making the whole process engaging and fun for the children 100

6.3.2 Input in the context of the heuristic evaluation method 101

6.3.2.1 Presenting and explaining the heuristic set so that children would understand 101

6.3.2.2 Severity scales that will be appealing to children 101

6.3.2.3 Explaining severity scales to children 102

6.3.2.4 The format of the evaluators’ data collection form 102

6.3.2.5 Problem merging process that will be engaging and fun 103

6.3.2.6 The order of the evaluation process 103

6.4 Discussion 103

6.5 Conclusion 104

6.5.1 Contribution of Chapter to Thesis....105

7 CHAPTER SEVEN: EXPLORING TEACHERS' IDEAS TO DESIGN THE NEW METHOD 107

7.1 Introduction....107

7.2 PART A: Creation of IMCH version 1....108

7.2.1 Storytelling....108

7.2.2 Values....110

7.2.3 Scripting the Storyline....111

7.3 PART B Testing Parts of the Method with Children....113

7.3.1 Study 1 (Younger Children)....113

7.3.1.1 Method....113

7.3.1.2 Result....121

7.3.1.3 Discussion/Conclusion....127

7.3.2 Study 2 (Older Children)....128

7.3.2.1 Method....128

7.3.2.2 Result....137

7.3.2.3 Discussion/Conclusion (PRODUCE METHOD V2)....141

7.4 PART C Input from Independent Teacher on IMCH V2....144

7.4.1 Process....145

7.4.2 Conclusion (Produced Method V3)....147

7.4.2.1 INSPECTION METHOD FOR CHILDREN (IMCH) V3....147

7.4.3 Conclusion of the Chapter....150

7.4.4 Contribution of Chapter to Thesis....150

8 CHAPTER EIGHT: TESTING THE IMCH V3....151

8.1 Introduction....151

8.2 Study with IMCH V3....151

8.2.1 Method....153

8.2.1.1 Participants 153

8.2.1.2 Apparatus 154

8.2.1.3 Procedure 155

8.2.1.4 Data Analysis for the IMCH V3 157

8.2.2 Result 177

8.2.2.1 Child Evaluators' Game Criteria 177

8.2.2.2 Child Evaluators' Usability Problem Data 179

8.2.2.3 Result: Child Evaluators' Usability Problem link to Game Criteria 180

8.2.2.4 Result from Analysing Observers' Data 181

8.2.3 DISCUSSION 184

8.2.3.1 Discussing Child Evaluators' Problem Data 184

8.2.3.2 Discussing Observers' Problem Data 187

8.2.3.3 Judging Ease of Use and Effectiveness of the method 188

8.3 Conclusion 189

8.3.1 Contribution of Chapter to Thesis 192

9 CHAPTER NINE: CONFIRMING FINDINGS AND RETESTING IMCH 195

9.1 Introduction 195

9.2 IMCH V4 Described 195

9.2.1 What the method should do: 195

9.2.2 How the method works: 196

9.3 Method 197

9.3.1 Participant 197

9.3.2 Apparatus 198

9.3.2.1 The Severity Scale 198

9.3.3 Procedure 200

9.3.4 Data Analysis 204

9.3.4.1 Game Criteria 204

9.3.4.2 Child Evaluators' Usability Problems Data 210
9.3.4.3 Confirming Problems from Previous Study 213
9.3.4.4 Identifying Severity Ratings for Problems Found 213
9.3.4.5 Observer's Data 215
9.4 Result 216
9.4.1 Game Criteria (Result) 216
9.4.2 Child Evaluator's Data (Result) 217
9.4.3 Result for Severity Rating 218
9.4.4 Observation Result 218
9.5 Discussion 219
9.5.1 Children's Problems Found 220
9.5.1.1 Critical Problem finding 220
9.5.1.2 Severity Ratings 220
9.5.2 Observations 221
9.5.2.1 Comparing Observed Problems to HE and Previous 221
9.5.3 Comparing IMCH V4 Study to IMCH V3 Study 222
9.5.4 Judging Ease of Use and Effectiveness 223
9.6 Conclusion 223
9.6.1 Contribution of Chapter to Thesis 224
10 CHAPTER TEN: THE METHOD – Inspection Method for Children (IMCH) VERSION 5 225
10.1 Introduction 225
10.2 What the Method Should Do 225
10.3 How the Method Works 226
10.4 Guidelines for Carrying out the IMCH Evaluation with Children 226
10.5 Conclusion 229
10.5.1 Contribution of Chapter to Thesis 229

11 CHAPTER ELEVEN: CONCLUDING THE THESIS 233
11.1 Introduction 233
11.2 Summary of Research...233
11.3 Answering the Research Questions 234
11.4 Contributions of this Research 235
11.4.1 Major...235
11.4.2 Minor...236
11.5 Limitation of this Research...236
11.6 Future Work...237
11.7 Researcher's Comment...238

12 BIBLIOGRAPHY 239

13 APPENDICES 255
13.1 Appendix 1 Chapter 5 Documents...ii
13.2 Appendix 2 Chapter 6 Documents..xii
13.3 Appendix 3 Chapter 7 Documents..xv
13.4 Appendix 4 Chapter 8 Documents - Children's Data for the first CVBIM EXPERIMENT lvii
13.5 Appendix 5 – Chapter 9 Document: Last CVBIM study with children lxxxiv

14 PUBLISHED WORK xciv

List of Tables

Table 2.1 Schematic representation of Piaget's Cognitive Developmental Stages 15
Table 5.1 Number of Individual problems found 78
Table 5.2 List of problems and evaluators who reported such problems 78
Table 5.3 Merged usability problem of adult expert evaluators 79
Table 5.4 Observers' data collected for individual expert evaluation......................................80
Table 5.5 Observer's data for the problem merging phase 80
Table 5.6 Number of Problems Reported by the Children 86
Table 5.7 Summary of Evaluators Merged Problems Reported 87
Table 5.8 Number of Observed Problems: General themes, Sub themes and Actual Problems in Phases..87
Table 5.9 Example of Repeated Sub Theme in a General Theme...88
Table 7.1 Description of younger children’s severity scale drawing....................................119
Table 7.2 Severity representation of problems rated using the traffic light system 120
Table 7.3 Game Criteria and Groups 121
Table 7.4 Themes and number of criterion for younger children’s game criteria 122
Table 7.5 Classification of Game Criteria into Themes and Matching Existing Heuristics/Guidelines 123
Table 7.6 Summary of younger children’s severity drawings 124
Table 7.7 Children’s text for their severity drawing...124
Table 7.8 Summary of usability problems rated using the traffic light colours 126
Table 7.9 Children’s criteria for a good game compared to existing heuristics 129
Table 7.10 Criteria Categories and Group 130
Table 7.11 Identifying exact group for each criterion 131
Table 7.12 Matching criterion to existing heuristics/guidelines..132
Table 7.13 Children’s drawing described 136
Table 7.14 An example of text used to describe each level picture......................................137
Table 7.15 Number of levels used and number of pictures identified for each level 137
Table 7.16 Result of merging game criteria..138
Table 7.17 Result of categorising children’s severity drawing...139
Table 7.18 Thematic Description of children’s drawing and frequency 140
Table 7.19 Picture text type and frequencies 141
Table 7.20 IMCH method Versions and Variation..149

Table 8.1 Detailed Criteria for Each Group....159
Table 8.2 (Extract) Similarities of Game Criteria between Groups....160
Table 8.3 Example of categorised problems merged as one....162
Table 8.4 Example of single problem reports from group 4....163
Table 8.5 Code and Details of Merged Problems for Each Group (1 to 3)....164
Table 8.6 Code and Details of Merged Problems for Each Group (4 to 6)....164
Table 8.8 Merged Problems and their Severity Ratings....170
Table 8.9 (Extract) Identifying game criteria themes and matching them to existing heuristics/guidelines....171
Table 8.10(Extract) Group 1's Observed Data for problem sub category....173
Table 8.11 Problems Observed Around Game Play....175
Table 8.12 Problems Identified for Game Criteria Area....176
Table 8.13 Problems around Merging phase and Severity Rating....177
Table 8.14 Summary of Children's Game criteria data....178
Table 8.15 Themes for Children's Game Criteria....178
Table 8.16 Summary of Children's usability problems reported....179
Table 8.17 Themes Identified for Children's Problems found....180
Table 8.18 Areas in which problems were observed for Child Evaluators....181
Table 8.19 Problems Observed around Game Play....182
Table 8.20 Comments observed for child evaluators....183
Table 8.21 Comments observed for child evaluators....184
Table 8.22 Observed comments for the facilitator....184
Table 8.23 Comparism of the Four Versions of the IMCH....193
Table 9.1 Coded Game Criteria (EXTRACT)....205
Table 9.2 Linking IMCH V4 Found Problems to Game Criteria....206
Table 9.3 Game Criteria Merged between Groups....208
Table 9.4 Game Criteria and Themes (EXTRACT))....209
Table 9.5 Merged Problems and Associated Groups....211
Table 9.6 Coded Merged Problems and Groups....212
Table 9.7 Children's Problem Category....212
Table 9.8 Current Problems Confirming Previous Study Problems....213
Table 9.9 Final Severity Ratings for Merged Problems....213
Table 9.10 Children's Explanation of the Severity Steps....214

Table 9.11 Observed Problems Encountered by Evaluators but was not Recorded..............215
Table 9.12 The Required Observed Problems..215
Table 9.13 Summary of Children's Game Criteria 216
Table 9.14Game Criteria Themes Identified 217
Table 9.15 Summary of Child Evaluators' Data 218
Table 9.16 Number of Problems Reported in Different Problem Areas................................218
Table 10.1 Comparison of the Five Versions of the IMCH...230

List of Figures

Figure 1.1 Routes Undertaken for this Research 8
Figure 2.1 Zone of Proximal Development (McLeod, 2007) 19
Figure 3.1 Smileyometer: A Visual Analogue Scale 36
Figure 4.1Stages of this Research 49
Figure 4.2 Stage 2 - Methods used and Type of Data Collected 64
Figure 4.3 Stage 3 - Methods Used and Chapters Implemented 65
Figure 4.4 The Roles of Children in the Design of Technologies for Them (Druin, 2002) 66
Figure 4.5 Methods Used at Different Stages of the Research and Data Collected 71
Figure 5.1 Images of iPad Fingu game (a) Actual game play (b) A successful play 75
Figure 7.1 The IMCH Method Version 1 112
Figure 7.2 Jamming Music (JamMo): A music making game 114
Figure 7.3 LeesCircus: A beginning reading game -drag letters to trashcan view 114
Figure 7.4 Milo and the Magical stone screenshot 115
Figure 7.5 Anti-heuristic space invader game 115
Figure 7.6 Study setting and seating position of participants 116
Figure 7.7 Design Session with Older Children A: Children drawing severity scale B: Judging the Severity of Problems 128
Figure 7.8 Traffic light facial severity scale 143
Figure 7.9 Comparing Versions 1 and 2 of the IMCH method 144
Figure 8.1 IMCH V3 Method Process 152
Figure 8.2 Maths Blaster Game a. Game play view b. Completing a level view 154
Figure 8.3 Ice maths ninja game – a. Sub game selection page b. Actual game play for subzero mines sub game 155
Figure 8.4 First IMCH Study with older children A. Game Criteria Being Written Down. B. Merged Problems on the Board with final Severity 156
Figure 8.5 Children Merging their Problems 157
Figure 8.6 Merged Problem Linked to Game Criteria 169
Figure 8.7 Version 4 of the IMCH 192
Figure 9.1 IMCH V4 Method Process 196
Figure 9.2 Ice maths ninja game – a. Sub game selection page b. Actual game play for sub-zero mines sub game 198
Figure 9.3 Coloured Face Severity Rating Scale 200

Figure 9.4 The Evaluation Session (a) Child Evaluators Performing Evaluation (b) Coloured Face Severity Scale in View 201
Figure 9.5 Children's Merged Problems with Final Severity..203
Figure 10.1 Coloured Face Severity Rating Scales 226
Figure 10.2 The IMCH V5 Process and Order 226

ACKNOWLEDGEMENT

Work in this thesis has gone through processes to reach its conclusion and this won't have been possible without the firm support of several people whom I will not fail to thank in this section.

I will love to thank all the children and teachers who participated at any point of this research, without whose contribution, this research won't have been completed. A big thank you to members of the ChiCI research group who supported any part of this work.

A special thanks to my supervisory team for their expertise and advice towards my work. A huge thank you and an applause to my supervisor Dr Gavin Sim, whose critical review of my work always helped my work towards a progressive turn. His relentless counsel and help has always saved my head from drowning in confusion. He has always been available to reply my emails and receive my calls whenever needed and allowed meetings even at short notice. A big thank you to my second supervisor Prof Janet Read for her insightful input which has always given my work a bounce back and a lift at different stages of the work.

A big thank you to my wonderful parents for their concrete support all through my education and even throughout this research (I love you both always), my mother in law, in-laws (Mr and Mrs B Wodike), my loving and very supportive siblings (you sparkle), my brothers and sisters in-law, other family members, Dr Ernest Nnadi, my darling friend Mrs Onyinyechukwu C Nnadi and other friends for their prayers, support and encouragement.

A big and a precious thank you to my darling son Ethan for understanding when mummy can't be available to do the hide and seek and peekaboo with you. When mummy has to run off to sort out research work and stay out late just to meet deadlines. A warm welcome to you my adorable Elnathan, you came just before this ship finally went to sail. A heartfelt thank you and big gratitude to my darling, sweet, precious husband who has been my rock and support all through this journey. Thank you love for all your tolerance, endurance and support just so that I make it successfully. I love and appreciate you deeply.

Thanks to everyone, may God bless you all.

DEDICATION

The work in this thesis is dedicated to God Almighty, the giver of life, wisdom and strength.

Thank You Lord for the grace, mercy and strength throughout this journey, May your name be praised forever. Amen

DEFINITION OF TERMS

Terms	Definitions for This Research
Children	Persons aged between 7 and 11 (younger: 6 to 8, older: 9 to 11)
Double Experts	People who are knowledgeable in the usability method and the technology being evaluated
Triple Experts	Having current knowledge about being a child, the usability method being used and technology being evaluated
Novice	Someone who has no knowledge or understanding about a subject matter
Ease of Use	Method is understandable and can be easily used for the purpose it was designed
Effectiveness	Method can be used to find real usability problems
Misses	Problems that exist but were not reported (omitted)
False Alarms	Problems that were reported but do not exist
Hits	Problems that exist and were reported
IMCH (new method)	**I**nspection **M**ethod for **CH**ildren

1 CHAPTER ONE: INTRODUCING THE THESIS

1.1 Introduction

This thesis contributes to the adaptation of usability evaluation method with children as evaluators. Apart from producing a value based inspection method that will be used by children, this research also produced some evaluation tools that could be used by Human-Computer Interaction (HCI) and Child-Computer Interaction (CCI) practitioners when carrying out evaluation with children in this context, which could also be used in a wider context. Also the research produced new strategies that could be adopted when designing evaluation tools or making adaptations to existing evaluation methods to become suitable for children. The research adopts an exploratory approach using a user-centred technique to determine the effectiveness of the heuristic evaluation method for children and further adopted an empirical approach to inform the redesign of the method and validation of the adapted method.

This chapter is an overview of the research carried out in this thesis. Section 1.1 explains the motivation of this research; section 1.2 puts the research into context; section 1.3 states the aim of the research with 1.3.1 and 1.3.2 highlighting the research questions and objectives respectively; section 1.4 gives a clear contribution of this work with 1.4.1 and 1.4.2 as its major and minor contributions respectively; section 1.5 describes the methodology of this thesis; 1.6 is a breakdown of the whole thesis pointing what each chapter will discuss; and 1.7 concludes this chapter.

1.2 Research Motivation

Having worked as a secondary school teacher, I had the opportunity of working with children aged 9 to 13years old. In a class activity session, children were given opportunity to analyse the presentation and articles of their mates, I observed children were honest with their opinions and made suggestions that worked well in practise (class). On getting the offer to do this research I was excited about the research area and the target research group with the possibility of getting honest input from them. I also thought it was an opportunity for children to voice their opinions in the design of technologies tended to them. As my position at this time was that children have been left out in all realms when educational technologies are developed for them.

However, the review of related literature showed me HCI is now concentrated on user centred approach in the design and evaluation of technologies; which means there is no design that fits all rather design is driven by the knowledge of the target users (Markopoulos & Bekker, 2003b). So children are already being involved. However, this drove my curiosity to investigate how involved children have been, then I realized the current stance and flexibility of HCI has led to the origination of the Child-Computer Interaction (CCI) community where research is suited to children. In this community, methods originally designed for adults are adapted to be used with children (this was interesting to me that I was highly motivated to carry on the research).

Through this adaptation process, several evaluation methods with children have been developed and proven effective. For example, survey methods (Read, 2007; Zaman, 2009), interview method (Zaman, 2007), verbalisation technique (Barendregt et al., 2007; Donker & Markopoulos, 2002; Donker & Reitsma, 2004), retrospective method (Vissers *et al*., 2013). However, most of these methods are user based, leaving a gap in usability inspection method (UIM), with the Structured Expert Evaluation Method (SEEM) as the up to date UIM designed to evaluate technologies for children. Although SEEM still uses adults as evaluators contrary to the proposition of CCI community that believes children should have a say in the design and evaluation of their technology. In addition, the DAAR model created by Woolrych & Cockton, (2002) to improve the assessment of UIMs proposes the use of the right evaluators as one of the routes to ensuring trustworthiness of UIM evaluations (Woolrych & Cockton, 2002). Therefore, using children rather than adults as the evaluators for technologies designed for children in a UIM evaluation seem an appropriate approach.

Some attempts have been made to perform a heuristic evaluation with children, however, these studies have reported issues that needs fixing for the method to be effective with children. For example, work by MacFarlane & Pasiali, (2005) carried out heuristic evaluation with children aged 13-14 years on a web based tutorial and illustrates that children can do a heuristic evaluation on technologies designed for themselves. (Wodike *et al*., 2014) also carried out a HE study with teenagers and in their work, teenagers (aged 11 – 13years) were empowered (trained to act as facilitators over their peers as an adapted process). Though the teenagers found few problems but the study was reported as unsuccessful due to challenges teenagers encountered, and recommendations were provided on better ways to involve teenagers in a heuristic evaluation study. In addition, Salian *et al*., (2013) analysed the effectiveness of the heuristic evaluation method in its state with children aged 9-11years and illustrates that children of this age can perform a heuristic evaluation on technologies designed for younger children.

All these works suggest that adaptation be made to the HE method to make it suitable for children. The first and second works, highlight that the heuristics should be rephrased, the second also recommends the need to make the process more fun and engaging for all the children. The third suggests that adaptation should be made to the method process, reporting that children had difficulty understanding heuristics and applying it and the severity ratings to problems found.

The HE method has a requirement of using double experts for an effective evaluation. That is, expert in the domain being evaluated and expert in usability (Nielsen, 1992). In some instances where children's game is being evaluated, there will also be the need to understand the class of users (children). In which case double expertise is insufficient but rather triple expertise: need to understand children, the domain (game) under evaluation and usability (Wodike *et al*., 2014). Therefore, it is necessary to use children instead of adults as the evaluators considering children are not short adults but a particular set of humans with their own likes, dislikes, curiosity, needs, use of language, and opinions (Druin, 1996a).

They understand the world differently than adults, they are separate human beings living in separate worlds from adults and even do different activities with computer from what adults do (Bruckman, Bandlow, & Forte, 2003; Janet C. Read, Markopoulos, & Druin, 2011; Janet C. Read, 2005). As seen in Druin's (1999) work, they even view and represent data differently from adults (who have been working with children). In Zaman's view, adults are unable to judge whether a game for children will be fun, challenging and user friendly, since they have lost the feeling of being children. This involves the cognitive, physical, emotional developments, and the (media) context in which they grew up (Zaman, 2005).

In the CCI literature several works are documented that show children are experts in handling the world around them and this expertise is significant in designing meaningful artefacts for them (Brodersen & Iversen, 2007). In the design area, children have been involved as testers, users, informants, design partners and stakeholders. In evaluation, they have been used in studies that use survey methods (e.g. fun toolkit, laddering and this or that) for getting user opinion and preference about technologies. They have also been involved in verbalisation method (e.g. think aloud and problem identification picture cards (PIPC)) for finding usability problems and have also been involved in the retrospective evaluation study (e.g. MemoLine) for gathering long term user experience about technologies.

Given that children could be experts in technologies designed for them, and experts at being children, the research initially aimed *to investigate whether older children can effectively carry out a heuristic evaluation on technologies designed for younger children if given appropriate tool*. However, outcomes from the pilot study carried out with children (as reported in chapter 5 of this thesis) and the heuristic evaluations with children and teenagers (cited earlier), has driven the change in the research direction. This research after careful review of literature and exploration of stakeholders' views, hypothesised that accessing children's everyday life (through narration/story telling) and allowing them perform inspection method evaluation based on their values (where values is defined as something that a person or group of people consider to be important in life (Iversen *et al*., 2010);) could eliminate HE problems (issues with understanding the heuristic set and method process) encountered and help them perform an effective UIM evaluation. The research used the heuristic evaluation method as a platform to design a value centred inspection method where children will act as the evaluators. This new trend could also be applied to other UIM evaluation methods on other applications for children in the CCI community and for real use. This new direction has produced the current research aim reported in section 1.3 below.

1.3 Research in Context

This research has its wider context in HCI although it is majorly situated within the concept of Child-Computer Interaction (CCI) and user centred approach. The research also cut across other disciplines: educational psychology, sociology and educational technology.

Mackay & Fayard (1997) reports that HCI is multidisciplinary, deriving its concept from both natural science and design discipline. Mazzone and Read *et al* proposes that it has its roots in Ergonomics, Socio Technical and Human Factor, with its attention on ways to enhance performance of machines manoeuvred by humans and provide guidance and research into work based system (Mazzone, 2012; Read *et al*., 2011). Its initial interest focus on highly powered machines such as airplanes, military and war machines (Mazzone, 2012). Research in HCI moved further into providing guidance in work places, deriving ways to reduce mistakes in systems when people use them (Mazzone, 2012; Read & Bekker, 2011). The revolution in the use of computers from work based fixed machines to personally owned systems, has brought about the change in focus of HCI research which now looks at how humans interact with computers. Methods are being developed to improve the experiences and usability of computer usage. The flexibility of HCI brought about the formation of the CCI community.

The CCI community is a relatively new field in the HCI space. It started with interest in the use of technologies within education and schools. CCI involves the design and evaluation of technologies where the humans are children (Read, 2005). It brings together different aspects of HCI: Educational Technology, Educational Psychology, Interaction Design, Fun Learning, and Sciences. It is focused on developing innovative work via investigating the different context of children and technology use (Read & Bekker, 2011).

Before the 20th centuries, work in CCI was carried out in HCI research groups by HCI researchers; and work done is reported in HCI journals, journals in computing and education, conference proceedings and books. However, recent years of CCI is undertaking processes to develop into its own discipline with its own methods and solutions (Read & Bekker, 2011). There are now CCI research groups with PhD researchers, research assistants and senior researchers who facilitate research work with children. Also it holds its annual conference termed as the Interaction Design for Children (IDC) and has its own journals and books. The advocacy of CCI is the need to produce rigorous and robust methods in interaction design and evaluation (Read & Fine, 2005). This research undergoes ethically approved processes within CCI to investigate and produce rigorous usability evaluation method for children that will benefit both old and new CCI and HCI researchers and practitioners in real use.

1.4 Aim

This research intends to make contributions that will benefit the CCI community by producing a suitable inspection method for children. Therefore the holistic aim of this study is *to investigate whether children can perform an effective inspection method evaluation (IMCH) on technologies designed for them based on their values*. In view of this, the following are the research questions for this research.

1.4.1 Research Questions

In order to achieve the aim of this research, the following has been set as the main research question:

RQ1. Can children perform a heuristic evaluation?

RQ2. How can children's performance in the heuristic evaluation be assessed?

RQ3. In the event of poor performance, what measures could be taken to produce a suitable UIM for children?

RQ4. Can children's values be incorporated into the new UIM?

1.4.2 Objectives

The following objectives have been mapped out to provide answers to each research question:

RO1. Assess children's performance in a UIM (HE) evaluation to determine the suitability of the method for them.

RO2. To gather information from multiple sources (stakeholders) that will inform the design of a child centred UIM.

RO3. To consider how values can be incorporated into the new child centred UIM.

1.5 Contribution/Novelty

This research has major and minor contributions detailed as follows:

1.5.1 Major

The major contribution (MAC) from this research is the method and an accompanying guideline on how to use the method

1.5.2 Minor

The minor contribution (MINC) is an insight of process (es) that could be used for the adaptation of evaluation method suited to children.

1.6 Methodology

This research followed a mixed method and a user centred design approach to achieve the outlined objectives and better answer the research questions. Qualitative and quantitative data were collected in some studies while in some others it was a single type of data. However, data collected usually informed the questions and or design of subsequent studies. For example, In the HE study with adults and children, evaluators' data collected consisted of qualitative (usability problems predicted) and quantitative (heuristic and severity number attached to problems predicted) data. Children's data were used to determine their understanding of the heuristic set and severity rating (this was determined from children's ability to appropriately allocate heuristic number to usability problems predicted). Observers' (qualitative) data collected from the same study, was used to ascertain children's ability to perform the HE method and the method suitability for children in its original state, and issues children encountered (for example: issues with understanding the heuristic set, severity scale and or facilitator's instructions) during the study. These informed the questions drafted for subsequent focus group studies. Experiments were carried out with children where qualitative data (game criteria and children's drawings) and quantitative data (frequency of child respondents during the game criteria session) were collected. This and the focus group data further informed the unstructured interview session with the independent teacher.

Several methods were used to analyse data collected according to the needs of the research. Observers' data collected were analysed using a closed and open card sort technique and thematic analyses. Usability problems gathered from the HE pilot study with adult and children were analysed using a data merging technique and thematic analyses. Data collected from teachers were analysed using qualitative content analysis in NVIVO. However severity data collected from the children (reported in chapter 7) were analysed and interpreted following a standard statistical approach while children's drawing were manually coded following a coding scheme inspired from literature.

The research used children as its major participants in the HE study, design studies and in the new **I**nspection **M**ethod for **CH**ildren (IMCH) study. Children were recruited from two schools in the UK with different ethnicity. Teachers who participated in the focus group studies were also recruited from two schools and the independent teacher for the unstructured interview session was from a third school.

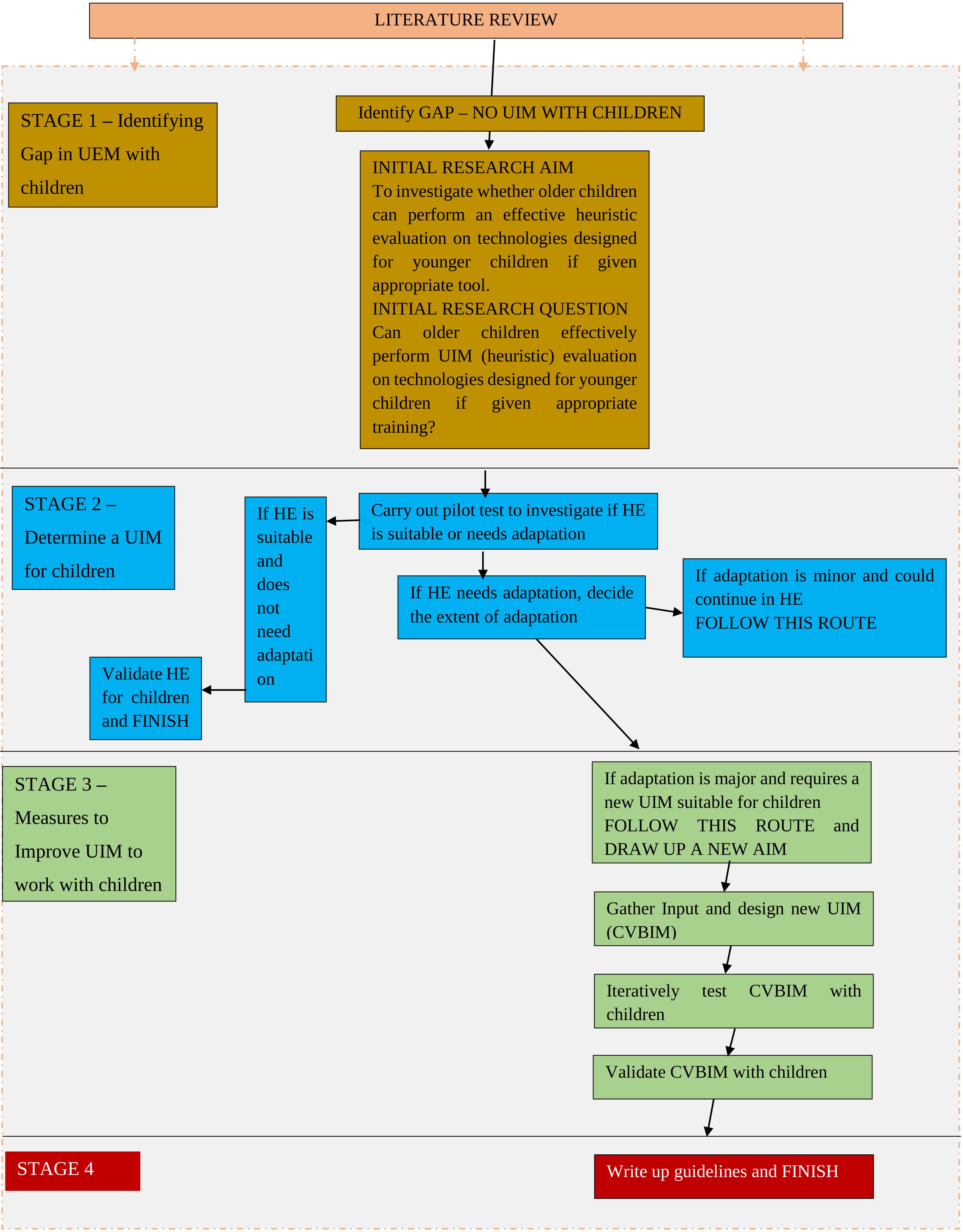

Figure 1.1 Routes Undertaken for this Research

1.7 Thesis Structure

Following this introductory chapter is chapter 2 which focuses on the user class of this research (children). The chapter is divided into 2 sections: the first section reports an overview of research carried out with children and the second section which ends the chapter is a review of literature in the cognitive development of children as reported by theorist and the report of their cognitive development given in HCI.

Chapter 3, which presents an overview of the usability versus user experience, usability evaluation methods (UEM) and the usability inspection method (UIM) as a subset in UEM. Research carried out in usability and user experience with children. Identified the gap that exist in UIM with children, also identified several UIMs in HCI, finally the chapter ends with heuristic evaluation method as the widely used UIM and the focus for this research.

Chapter 4 presents an overview of the research methodologies and techniques deployed for this research. These approaches are reported in connection to the research stages presented in figure 1.1

Chapter 5 reports the initial pilot study carried out for this research. The first study within this chapter provides an understanding of carrying out HE study with adult experts and provided an avenue for comparison with study done by the children. The second study with the children presented an overview of the HE method in its original state (without adaptation), when carried out with children.

Chapter 6 and 7 presents an overview of the adaptation process that involved gathering data from multiple stake holders in the matters that concern children's learning and education. Chapter 6 reports an iteration of focus group study with school teachers, where data collected informed the study with children and discussion session with an independent teacher, both reported in chapter 7.

Chapter 7 is divided into three sections: Part A, B and C.

Part A, details an overview of narration as a method for gathering requirements from children based on their values to inform the redesign of the intended UIM tools and process. This generated the version 1 of the Inspection Method for CHildren (IMCH).

PART B, reports the studies carried out with children using narration to test part of the new method which produced version 2 of the IMCH.

PART C, discussed the new method scrutiny process with an independent teacher to produce IMCH version 3. The section ended the chapter giving a description of IMCH V3 method process reporting IMCH study carried out with children in the fourth version.

Chapter 8 reports the first studies with children (older and younger) using IMCH in the 3rd formation state.

Chapter 9 reports the second IMCH (V4) study with children and highlighted the state of the method in this version

Chapter 10 explained and described the method (IMCH) at the completion of the thesis, stating what it should do, how it works and its accompanying guideline.

Chapter 11 presents a discussion of the processes undertaken for this research and presents the findings to inform future work for new and old researchers interested in developing UIM with children. It is also the concluding chapter that summarises this work and presents answers to the research questions. Finally, it highlights the activities that produced the research contributions, stated the limitation of the work and future route of the research.

1.8 Conclusion

This first chapter has presented an introduction of this thesis, highlighting the research motivation, the context on which the research is situated, the aim and objectives of the research also stating the research questions. This research is typically motivated by the researcher's interest to produce novel work providing solution to existing problem in CCI, in this case in a robust usability inspection method for children.

The research aimed at producing a thorough inspection method for children, by which processes undertaken and trend could be applied in other similar forms. Chapter two will review literature on research with children, and their cognitive development in relation to this work.

1.8.1 Contribution of This Chapter to Thesis

The chapter introduces this thesis and clearly states the aim and objectives of this thesis.

2 CHAPTER TWO: UNDERSTANDING CHILDREN IN RESEARCH

2.1 Introduction

The target users of this research as stated in chapter one are children. Therefore, this chapter aims to explore the usefulness of involving children in research and identify the expectations of the research with this target group in the stated research area. The chapter has the following objective:

- To define the term children and state clearly what age ranges are classified as children
- To review theorist and researchers' perspective on children's cognitive in relation to this research
- To review literature on the involvement of children within CCI

The chapter is therefore divided into 3 sections: the first section reports an overview of involving children in research, the second section is a review of literature on the cognitive development of children as reported by theorist and CCI researchers, the third section which ends the chapter reports the CCI community and how it involves children in research when designing and evaluating technologies for children.

2.2 Children

According to the NSPCC factsheet, there is no legal definition for a child in the UK (i.e. England, Wales, Northern Ireland and Scotland), as each region have their own set guidance on the responsibilities to keep children safe. Though they agree that a child is anyone who has not reached their 18th birthday (NSPCC, 2013). This suits the definition provided by the UN Convention on the right of a child, which defines a child *as any human being below the age of 18* (UNICEF, 1990). Children have been involved in different fields of research for example, in clinical research (Knox & Burkhart, 2007), social research (Mason & Hood, 2011), fire safety research (Harpur *et al.*, 2013), educational research (Azevedo & Ferreira, 2013), nutritional research (Kafka *et al.*, 2011) and also in computing research (Frauenberger *et al.*, 2012). Oftentimes researchers skip the definition of a child but rather focuses on the target group of children for their research. In the CCI community the target age range of children used is between the ages of 5 and 12years old (Read & Markopoulos, 2013). However, *this research target of children are children aged between 7 and 11year olds*.

Children are physically and mentally immature and classified as vulnerable group (Knox & Burkhart, 2007) who are susceptible to any information, therefore their needs ought to be specially safeguarded and cared for, including appropriate legal protection, before as well as after birth (UNICEF, 1990). In research some modalities are put in place and considered to ensure the right of the child is safe guarded. For instance research that involve children are ethically reviewed to ensure appropriate consent is given by either the child or guardian of the child and in some cases by both parties (Read *et al*., 2013).

In this research, all ethical issues that concerns the child was appropriately considered, full details of ethical measures taken for this research is provided in chapter 4.

2.3 Children's Developmental Stages

Technology use by adults and children require attention, perception, memory, information processing, decision making and even more (Gelderblom & Kotzé, 2008). In addition, evaluating technologies involve the use of the technology, processing information, making decision based on perception and then make a report (either verbal or written). In the instance where children will carry out the evaluation, it is crucial to understand that children are not a homogenous group for which a single theory and practise should be recommended (Markopoulos & Bekker, 2003a) but rather vary in cognition at different stages of childhood. This variance is a pointer to understand the cognitive development of the particular group of child users when involving them in design or evaluation of technologies. Also, Gelderblom & Kotzé (2008) demonstrates the valuable benefits of having a wealth of theoretical knowledge on children's cognitive development when formulating frameworks for the design of children's technology. It is believed that this is also profitable in an intention to design evaluation technique or method for them.

Several scholars have looked at children's development and invented theories, for example: Skinner (1974), and Watson (1913) behaviourist approaches, Maslow's humanist approach (Maslow, 1943), Bandura's social learning approach (Bandura, 1971), Vygotsky's social interaction approach (Vygotsky, 1978) and Piaget's cognitive approach (Piaget, 1971).

Reviewing all of these theories go beyond the scope of this thesis and since, this research boarders around learning, interaction and cognition that concerns children; Piaget, Vygotsky and Bandura's approaches will be reviewed. Also, Piaget and Vygotsky's developmental theories are one of the most cited work in the CCI community. In addition, work carried out by CCI researchers to analyse the developmental stages of children in connection with the use of computer and technology will also be reviewed. This include works by Hanna, Risden, & Alexander (1997) and (Markopoulos & Bekker, 2003a)

2.3.1 Piaget's Cognitive Development theory

Jean Piaget's (1971) cognitive development theory is one of the most cited in child cognitive development. Maier (1978) argues that Piaget's cognitive development work is rooted in biological and psychological view point. It is criticised as being solely focused on individual development and did not recognise the importance of the social factors in development (DeVries, 2000). However, his epistemological work focuses more on cognitive development of the child (Piaget, 1971), that is, how specifically cognition works (Wang & Rubart, 2006) for children rather than on child development generally. Cognitive development by Piaget is a flexible increase of one's perception of the world around them and their continuous adaption to their environment (Dodonov & Dodonova, 2011), that is, cognitive structures are created and adapted during an interaction between a subject and its environment (Wang & Rubart, 2006). Though the rate of development and what chronological point it occurs in a child's development varies but experience rather than maturation is the essence of this cognitive development (Maier, 1978). According to Piaget, cognitive human behaviour is traceable to a combination of the following factors(Piaget, 1971;Maier, 1978):

- Maturation of Bodily Process: difference in the nervous system
- Experience: bodily interaction with the physical world
- Social transmission where humans take care of and educate individual and affect the nature of the individual's experience.
- Equilibration which is described as the force that moves development along (McLeod, 2009) and is also seen as self-regulation where an individual's first attempt to understand a new experience is by using previous knowledge and when such comprehension does not fully explain the new experience then they change their previous conception to situate the new happening in a more balanced agreement with their personal conception of events.

He believes human knowledge does not come from sensation neither is it from perception but rather proceeds from the entirety of their actions of which perception and sensation are only an indication (Liben, 1983). Similarly, in Maier's work he stated that human learning or development is neither purely social nor purely maturational; but rather development evolves from individuals' experience of themselves and the patterns of living. His conception on human development could be seen in light of the two aspects (Adaptation and Equilibration) (Maier, 1978):

Adaptation is seen as the intellectual features in any behaviour change which occurs in the interaction with the changes in the environment; it is the creation of an action and the action itself. This is further broken down into or occurs through assimilation and accommodation (Maier, 1978; McLeod, 2009) where:

- Assimilation is the mental effort of experiencing an event in terms of past internal experience (Maier, 1978); that is, the tendency of a child to interpret an experience via existing knowledge structures which is assumed to be the child’s initial tendency (Feldman, 2004).
- Accommodation is the realisation that current knowledge is insufficient for adequate understanding and it must be changed (Feldman, 2004). In other words, it involves the impact of the environment on the child where the child adjust to current event by changing his/her initial conception to fit more correctly into the demands of the actual event (Maier, 1978).

When a child is faced with an event, assimilation occurs where previous knowledge is used to tackle that event but when the event involves more advanced thinking, then equilibration sets in where the child accepts that the previous knowledge is not sufficient enough and therefore needs to change his previous concept to accommodate the new concept. At this stage the child learns and take mastering of the new concept. These two processes (assimilation and accommodation) always act together (Maier, 1978).

Within his theory, Piaget advocates that children are different from adults in their view of events and opinion (McLeod, 2009; Piaget, 1971). He further demonstrates that the cognitive ability of adults is far more developed than that of children such that though a child and an adult give the same answer to a particular question, their answers might mean different things entirely (Maier, 1978).

He further proposes that children know best what they are all about and are capable and ready to learn more (Maier, 1978) but their knowledge of the world around them depends on the development of coordinated activities (Liben, 1983). He also argues that the cognitive ability of children are not the same as they pass through cognitive developmental stages and become more aware or knowledgeable. However, this does not suggest that older children are more intelligent but rather older children would have experience more actions that provides more recognition of objects around them (Maier, 1978).

He explains this further by providing stages of cognitive development in children from birth to age sixteen which is described as the vehicle for analysis and not a core process at the heart of his development theory (Dawson-Tunik *et al.*, 2004). These cognitive development stages are popularly presented in four stages (see table 2.1) (Boom, 2004; Dawson-Tunik *et al.*, 2004; Feldman, 2004; Kesselring & Müller, 2011):

Table 2.1 Schematic representation of Piaget's Cognitive Developmental Stages

Piaget's Cognitive Development Stages	***Approximate Age Description***
Sensorimotor Stage	Birth to age 2years
Pre-operational Stage	aged 2 to 7years
Concrete Operational Stage	aged 7 to 12years
Formal Operational Stage	aged 12 to 16years

However, Maier's (1978) work identified five stages and called it phases. In his work, the preoperational stage was split into two stages and age brackets. He referred to these two stages as the preconception stage which bothers on children aged 2 to 4years old and the intuitive thought stage which comprise of children aged 4 to 7years old. Although analysis of the forma and latter stages show developmental details recorded for the preoperational stages (of other researchers) is similar to that recorded for the preconception stage and the intuitive thought stage of Maier's (1978) work. Also in (Feldman, 2004) effort to describe the pre-operational stage, it was split into two halves.

This relates Maier's (1978) argument that the developmental stages are only points for understanding the sequence of development that is, to denote the course of development but do not present development itself. As no individual is ever aware of being in one stage or another but merely interact as though they knew their stages. In addition, Piaget does not claim that a particular stage is reached at certain age, though the stage description is an indication of the age an average child would reach at a particular stage (McLeod, 2009).

For the purpose of the target group of this research, the earlier stage categorisation will be used focusing on the preoperational (aged between 2 and 7years old) and concrete operations (children aged between 7 and 12years old) stages of his cognitive development.

2.3.1.1 The Preoperational Stage (age 2 to 7years)

Piaget describes this stage as an egocentric stage where children in this stage have difficulty in in taking the view point of other people (Feldman, 2004) that is, they know the world only as they see it and do not know any alternatives (Maier, 1978). Contrary to Piaget's view of the preoperational stage being egocentric Vygotsky in (Feldman, 2004) thinks though the egocentric view is real but the broader generalisation is inappropriate as he puts the egocentric speech at the midpoint of the pre-operational stage which is about age 3 or 4 (Feldman, 2004).

In this stage, ages from 2 to 4years old children have the ability to engage in different activities such as: imitation in the absence of the model, pretend play, drawing, psychological functions based on mental image and language (Kesselring & Müller, 2011). Assimilation is the paramount process of thinking. Towards the end of this stage Children begin to use words as part of their thinking process. They achieve a new level of thinking and can project themselves into other roles and begin to think in terms of other people. They judge experience from the outside appearance and in terms of the ongoing events (Maier, 1978).

2.3.1.2 The Concrete Operational Stage (age 7 to 12years)

Piaget (1971) refers to operations in this stage as logical operational or principles used when solving problems. He argues that children in this stage do not only use symbol representationally but can also use the symbols logically. That is, they have the mental capacity to order and relate experience within an organised whole (Maier, 1978). Although the child must still perform these operations within the context of concrete situations (Piaget, 1971). In this stage it is stated that children are able to fully employ thought structures rather than relying primarily upon perceptual or body-motor cues as they did when they are younger (Maier, 1978).

2.3.2 Vygotsky's Social Interaction Approach

Lev Vygotsky a Russian cognitive psychologist (Vygotsky, 1978) and a constructivist (DeVries, 2000) has also made great impact in the cognitive development of children. His work on learning and cognitive development which was originally in Russian was carried out around 1920s and 1930s, same time as the work by Piaget (McLeod, 2007). However it remained unknown until the 1960s to 1970s when it was translated into English (Vygotsky, 1978).

Most of his work which is socio-cultural in approach is in agreement in many ways to Piaget's theory of cognitive development (DeVries, 2000). For example just like Piaget, Vygotsky believes cognitive development or learning is not necessarily by maturation as it is only a precondition but not the result of learning (Vygotsky, 1978, 1997). He also agrees that action is important as a beginning of diverse forms of intelligence (Lourenço, 2012). However, he suggest Piaget's developmental work as being individualistic rather than socially oriented (DeVries, 2000; Lourenço, 2012) as his work argues that social factors are important in learning and development (DeVries, 2000; Vygotsky, 1978). Although this criticism is judged in Literature as an unfair comparison of both theorist (Piaget and Vygotsky) work (DeVries, 2000; Matusov & Hayes, 2000) and Vygotsky's work sees the individual as the goal of development (DeVries, 2000). Therefore, both (Piaget and Vygotsky's) work is judged as similar (Lourenço, 2012).

Vygotsky's theory suggest that children development of learning is more external than internal as learning occurs through participation in various forms of social interaction with peers, parents and people in the society (Vygotsky, 1978, 1997) using tools (e.g. pencil, hammer) and signs (e.g. language, pretend play and mathematical formulae) (Lourenço, 2012). Since tools, values, signs and believes vary from culture to culture and these affect cognitive development, Vygotsky believes cognitive development varies from culture to culture (McLeod, 2007).

He further argues that the environment in which a child grows up can influence how s/he thinks and what s/he thinks about (McLeod, 2007). For example, prior to receiving formal education, a child assimilates the names of objects and items in his/her environment which allows learning to take place.

According to Vygotsky, the interaction or collaborative dialogue that occurs between a child and a more knowledgeable person or skilful tutor who is able to provide verbal instruction or model a behaviour can enhance learning for the child or promote cognitive development. The child makes effort to understand the instructions or behaviour provided by the tutor (often a parent, guardian or teacher) then internalizes the information, using it to guide their own performance. For example, a child who is given her first task, who performs poorly while trying to perform the task. Later guided by a parent or more knowledgeable peer who provides instruction and clues on how to perform the task and applauds the child when the right steps are taken and suddenly allows the child to independently accomplish the task (McLeod, 2007).

This Vygotsky's theory of cognitive development rests on two principles: More Knowledgeable Other (MKO) and Zone of Proximal Development (ZPD).

2.3.2.1 More Knowledgeable Other (MKO)

The more knowledgeable other (MKO) refers to someone who has more understanding or a higher ability level than the child, in relation to specific task, process, or concept.

Often times with a child, his/her peers, older siblings, parent (s) or guardian or an adult may be the individuals with more knowledge or experience. For example, a peer who knows how to play a game better, an older sibling who can have more experience in solving puzzles, a parent who can guide the child on the right step to take concerning his/her homework, a teacher who teaches the child, only to mention a few.

An MKO does not necessarily mean an older person as a younger person could be more knowledgeable at a task than one who is older. For example, when it comes to the latest kids' app on a technology, children could be better knowledgeable on how the app controls work than their older parents. It is important to note that the role of an MKO could switch as the task changes as the key to MKOs is that they must have (or be programmed with) more knowledge than the learner about the current task being learned.

The MKO could also be an electronic tutor or mobile device, it need not be a person at all. As the trend these days is the use of such (electronic devices) for passing on information aside person to person technique; and most classrooms now employ the use of technology devices for teaching and learning (McLeod, 2007).

This MKO is a crucial part of Vygotsky's theory because it is seen as what proceeds the learning when a child becomes stuck having carried out the task independently on a previous mastery or knowledge. This drives the child's development from the known to the unknown. However, this concept is fundamentally grounded in Vygotsky's second principle which is well referenced: zone of proximal development.

2.3.2.2 Zone of Proximal Development (ZPD)

Vygotsky defines this concept as the *distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers*.(Vygotsky, 1978 p. 86)

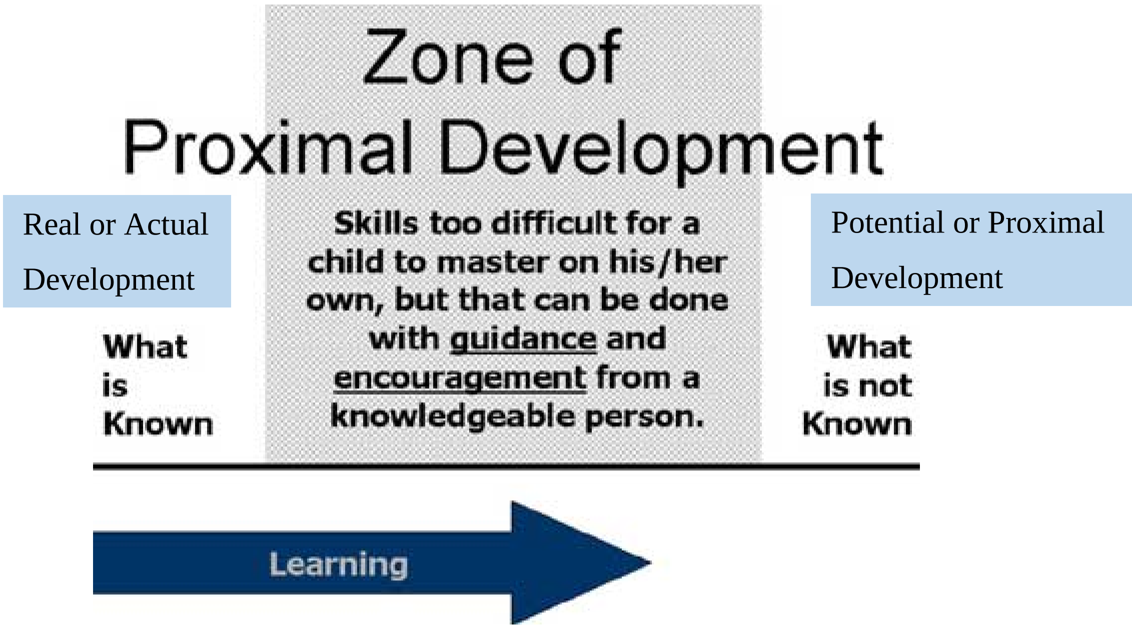

Figure 2.1 Zone of Proximal Development **(McLeod, 2007)**

This ZPD is an important concept that explains the difference between what a child can achieve independently and what a child can achieve with guidance and encouragement from a skilled partner (MKO) (McLeod, 2007). ZPD stresses what the child is capable of doing by him/herself (real or actual development), against what s/he can do with assistance from others' (the MKOs) guidance (potential or proximal development) (Lourenço, 2012). It provides the teacher or educator with a clearer perspective of what the child is capable of doing independently (Vygotsky, 1978).

It is believed that if a child is capable of doing task A independently then it means the function to do that task has matured but ZPD defines the functions which have not yet matured but are in the process of maturation (see fig 2.1). It is also suggested that ZPD is the point at which sensitive guidance and instructions should be provided (McLeod, 2007) as what the zone of proximal development of a child is today will become the child's actual development tomorrow. That is, what a child is capable of doing today with assistance, the child can do by him/herself tomorrow (Vygotsky, 1978).

2.3.3 Bandura's Social Learning Theory

Albert Bandura, a Canadian psychologist is a behaviourist whose work situates more on behavioural learning. Unlike Piaget and Vygotsky's he believes behaviour is learnt from the environment through observation (McLeod, 2011). Children tend to observe people around them which are referred to as models. These influential models can either be siblings or parents from home, children TV characters from watching TVs, peers and teachers from school and friends from playgroups (McLeod, 2011). From observing these models they decide on what new behaviour to act out.

However, humans are neither driven by inner forces nor tossed about helplessly by environmental forces (Bandura, 1971) because if human behaviour is controlled solely by external outcomes, then people would behave like weathervanes, constantly changing directions to adapt to whatever sudden social influence that is imposed on them. In reality, people have self-reflective and self-reactive capabilities that enable them to exercise some control over their thoughts, feelings, motivations, and actions (Bandura, 1991). In addition, behaviour is formed from continuous reciprocal interaction between behaviour and his controlling conditions e.g. environmental influences (Bandura, 1971).

He further suggests that man's capacity to learn through observation enables him to acquire large integrated units of behaviour by examples without building up gradual patterns running into trials and error. Similarly emotional responses can be developed observationally through witnessing the affective reactions of models who are undergoing painful or pleasant experiences(Bandura, 1971).

Learning by observation occur due to the following reasons (McLeod, 2011):

Children observe, attend to and imitate models who they perceive to be like them, especially people of the same sex.

Children are likely to imitate and repeat a behaviour based on the response received from people around. This response can either be reinforcement or punishment. Reinforcement can be internal (happiness the child feels for being appreciated for imitating a behaviour) or external (reward received from a person around); it can also be positive or negative. On one hand, a positive outcome can come in the form of reward (verbal or material) and is likely to have great impact on behaviour. On the other hand, a negative reinforcement are unpleasant reward that do not match the child's need, which could have little impact.

Children will also consider what happens to other people when deciding whether or not to imitate their actions. This is known as vicarious reinforcement (McLeod, 2011).

This Vicarious means could be used to eliminate fearful or defensive behaviour by allowing the child observe someone in a fearful activity without any adverse consequences (Bandura, 1971).

2.3.4 Hanna *et al's* work

Hanna et al (1997) in their work proffered their knowledge as developmental psychologist and their practises in carrying out usability testing with children to report on their observation of children's behaviour in connection to performing evaluation. Their work focused on children aged between 2 and 14 years old in three categories (Preschool aged children (2 to 5years), elementary school aged children (6 to 10years) and the middle school aged children (11 to 14years)). However, for the age ranged specified for this research only the second and the third group will be reviewed.

2.3.4.1 *Elementary school aged children (6 to 10year olds)*

Elementary school children are relatively easy to include in software usability testing. Their experience in school makes them ready to sit at a task and follow directions from an adult, and they are generally not self-conscious about being observed as they play on the computer. They will answer questions and try new things with ease. In this age range, children will develop more sophistication about how they can describe the things they see and do. Six- and seven-year-old children will be more hands-on— ready to work on the computer but a little shy or inarticulate when talking about the computer. Ten-year-old children may have extensive computer experience and be ready to critique your software (Hanna *et al.*, 1997). This is therefore useful for this research as the evaluation of technologies involves the use and critique of the technology.

2.3.4.2 Middle school aged Children (11 to 14year olds)

Middle aged children are very easy to include in usability testing. Most will be comfortable with computers and with unfamiliar adults. Children this age can be asked to perform, and actually enjoy, specific tasks after a period of free exploration. Some older children in this age range may be able to "think aloud" during the session, while others may be self-conscious about having people watch them and listen to what they say. These children may bring a very high level of computer expertise, or distinct expectations for what they will be doing, to a usability session (Hanna et al., 1997).

2.3.5 Markopoulos and Bekker's work

Markopoulos and Bekker in their article made a characterisation of children in relation to the design of interactive technologies for them, considering their humour, changing interest, character and settings. However, since the products to be evaluated are interactive technologies for children, it is useful that this work is reviewed to get an insight of technologies that will be appropriate for the chosen age group. They made an argument that children's interaction with technologies differ with age and further made a distinction of four age group or stages of development in children. Their discussion focuses on the developing skills, needs and knowledge of children and because there are large differences between children, the division in stages are an approximate and the various development theories also assume different boundaries. The stages will be discussed in light of this research age group (Markopoulos & Bekker, 2003a). Therefore, only the two middle ages will be discussed.

2.3.5.1 *The emerging-autonomy stage (ages 3–7)*

Children in this stage, enjoy fantasy and magic. They are fairly self-centred and does a lot of parallel play. They have a need for stimulation, love and safety, though they are developing a greater need for independence. It is very important that products for this age group is kept simple and the products should be based on concepts that are not too abstract and are tuned such that they are not yet up to the fully developed reasoning skills of this age group. Ideas based on the past or future are still difficult to understand, so concepts around themes playing in the present and close to home will be most appealing.

Children develop their knowledge about letters, words and books between the ages of 2 to 6 years. They gradually develop an understanding of words that sound the same at their beginnings or ends, break words apart, or combine words into new words or phrases, they also begin to put characters into words.

Between 3 and 5, they start developing conversational strategies such as adjusting speech in relation to social expectations. Between 3 and 6, children start to use more complex grammatical sentences in which two sentences are combined using connective words such as 'and' or 'because'. Also between ages 5 and 9 they develop the ability to gradually change the topics of discussion.

Children in this age group also develop their initial writing skills, starting with scribbling single characters around the age of four. Then they develop the ability to write words and create sentences and leave spaces between words (Markopoulos & Bekker, 2003a).

2.3.5.2 The rule/role stage (ages 8–12)

Interests of children in this age group shift gradually from fantasy to reality. They play in pairs and groups and become more interested in competition. Children start developing a sense of logic and reasoning and simple abstractions. They are sensitive and have a need for peer acceptance and success. Finally, there is a shift from a main influence of parents and school to a bigger influence from friends.

Products for children between ages 8 and 12 years old can be more complex and challenging; also variation and competition play an important role. They become more aware of the age-appropriateness of products. Since concepts such as the past and the future can be grasped, themes such as science fiction become more obtainable. Around the age of 8, children shift from learning to read, to reading to learn. From 8 to 12 children start to understand more abstract and longer terms and more complex sentences.

They develop the ability to critically analyse what they read. They also develop the ability to signal subtle differences using pronunciation. Over the years children (ages 6 to 10 years) develop an increasingly large vocabulary and understanding of multiple meanings of words. Subsequently, from ages 7 to 9 years they become more proficient at spelling words correctly, writing complete sentences, and in using capitals and punctuations. Children between the ages of 9 and 10years are still not very good at planning their story and start telling the story straight away. In addition, the use of language can be more complex and abstract.

2.4 Child-Computer Interaction

2.4.1 The Origin of CCI

Child-Computer Interaction (CCI) is a community where research is suited to children. It is a branch of HCI in which researchers and developers focus on the interaction between children and computers (Read, 2005). Most work in CCI have its underlying methods and theory in HCI though it is modified when in use with children (Read & Markopoulos, 2013). Although it has only just emerged into a discipline and is growing significantly (Read *et al*., 2011). Work in this area dates back to late seventies, eighties and nineties with the first major works as Logo programming language, Lego Mindstorms (Papert, 1980), and Constructionist Child development theories (Papert, 1988), Revelle & Strommen, (1990) and Resnick (1991). These are accepted and recognised as the originators of CCI (Druin, 2002; Read & Bekker, 2011) and at that time published in HCI journals, and also journals in computing and education. The field goes beyond being linked to HCI, computing and education, it is also linked to psychology and sociology.

2.4.2 Defining CCI

CCI researchers argue that children are not short adults rather differ greatly from adults in their activities, needs, perception, wants and abilities (Druin, 1996a, 1999b; Read, 2005). This has sustained the research in this area and allowed several attempts towards defining CCI. For example, it is reported that the first definition is by the contributions made at the IDC 2002 inaugural workshop (conference) (Read, 2005), this is believed problematic if the conference changes (Read, 2005). A later definition by Read resolves that *it is an area of HCI where the humans are children* (Read, 2005). In a paper capturing the nature of CCI, there were two definitions of CCI by Read & Bekker (2011): the first definition was explored in HCI terms defining it as *a discipline concerned with the design, evaluation and implementation of interactive computing systems for children's use and with the study of major phenomena surrounding them (p. 3)*. This definition however is seen as unable to stand on its own (Read & Bekker, 2011). The second definition was more on CCI agenda, where it is defined as *a study of the Activities, Behaviours, Concerns and Abilities of Children as they interact with computer technologies, often with the intervention of others (mainly adults) in situations that they partially (but generally do not fully) control and regulate (p. 7).*

However, the recent definition of CCI states that *it is an area of scientific investigation that concerns the phenomena surrounding the interaction between children and computational and communication technologies. It combines inputs and perspectives from multiple scientific disciplines informing and supporting an area of research and industrial practice that concerns the design of interactive systems for children* (Read & Markopoulos, 2013 p. 2).

2.4.3 Popularity of CCI and Origin of IDC

As discussed earlier, the activities of CCI dates back to the late 70s, but it was until the end of the twentieth century with works by Druin & Solomon (1995, 1996) and Druin (1996a) that made the field popular (Read & Bekker, 2011). Around that time, Druin (1996b) also established the Chi-Kids community as a part of the ACM SigChi group which ran yearly between the years of 1996 and 1999 at the major HCI conference "CHI". In addition to Druin's work, Hanna et al., (1997) made an influential publication in 1997 on what to consider when carrying out usability evaluation with children on their own technologies.

In 2002, Markopoulos & Bekker (2003) from the Eindhoven University of Technology held a workshop on the Interaction Design and Children (IDC), which had over 100 people in attendance (Read & Bekker, 2011; Read *et al*., 2011). In the following year (2003), Bruckman *et al*., (2003) made a publication in a chapter of the HCI handbook titled “HCI for Kids” and in this same year, the inaugural conference on Interaction Design and Children was held at the University of Central Lancashire, Preston, UK (Read & Bekker, 2011; Read *et al*., 2011). This conference then moved across to the United States in 2004, to be hosted by the HCI lab of the University of Maryland. The 2004 IDC conference focused on the importance and current challenges of allowing children to be integrated at the early stage of the technology design process.

Since then, the IDC have been hosted annually across the globe (Read & Bekker, 2011; Read *et al.*, 2011). The activities of CCI is not only published at the IDC but also in other notable HCI venues: British HCI, Interact, NordiCHI and CHI. There is also a record that CCI publications at CHI is on the increase each year (Read & Bekker, 2011; Read *et al*., 2011). It is essential to stress that IDC and CCI are two different terms as IDC focuses on design while CCI on the other hand focuses on theory (Read & Bekker, 2011).

2.4.4 CCI Involvements, People and Practices

In 2005, Read proposed the need for a special area of study for CCI instead of it being only a subset of HCI, identifying the difference between children and adults on three key issues: Activities, Behaviour and Concerns. Stating that children do different activities with computers than adults do, behave very differently around computers than adults do and have different concerns about the use of computers (Read, 2005). In 2008, a special interest group (SIG) on CCI (Read & Bekker, 2011; Read *et al*., 2011) was proposed and accepted at the 2008 CHI conference which held in Florence Italy (Read *et al*., 2014) and in 2009, the CCI community created an international IFIP SIG under the TC13 group (http://www.idc-sig.org), where research interest is cross disciplinary inviting members and publication from different disciplines.

Currently, the CCI community has her own international journals published by renowned journal publishers such as Elsevier, Springer and IGI Publishing, examples of such journal publications are:

- Elsevier International Journal of Child-Computer Interaction, with ISSN 2212-8689

- Springer
  a. Child-Computer Interaction: Methodological Research, special issue on cognition technology and work journal (Markopoulos et al., 2008). Some articles of this journal can also be found here: http://www.journals.elsevier.com/international-journal-of-child-computer-interaction/recent-articles/
  b. Tangibles for Children, special issue, personal and ubiquitous computing (Zaman *et al*., 2009) a

Sometimes CCI findings are also published in HCI journals for example, International Journal of Mobile Human-Computer Interaction, published by IGI Publishing.

Several books have also been published by the community, with several courses, panels and special interest groups on Child Computer Interaction being held at CHI, Interact and IDC. There is also the record of at least one university that offers Child Computer Interaction at MSc level (Read & Bekker, 2011). Currently, the community is estimated to have around 500 researchers with about 150 PhD researchers and research assistants (Read *et al*., 2011).

The community is being chaired by people who have made significant contributions to move the community forward. People who have written books and papers, regular attendees of CHI conferences, people who have chaired committees and subcommittees at CCI related conferences, served as committee members and people who combine expertise from various areas of CHI, for example, surface computing, user experience, tangible interaction, ubiquitous computing, participatory design and education (Read *et al*., 2011).

2.4.5 The Subjects of CCI

CCI embraces different subject areas with current research themes on interactive techniques, design practises, and evaluation methods (Read & Markopoulos, 2013). Researchers in CCI cater for children's want and need when designing and evaluating interactive technologies for them by making adaptations to methods originally developed for adults in order to suit children.

2.4.5.1 *Input Devices with Children*

It started with early work in input devices which investigated children's performance as users of interactive technologies originally designed for adults, this revealed the need to adapt the method to suit children (Read & Markopoulos, 2013). For example, work by (Revelle & Strommen, 1990) examined the effect of practise on children's computer control when using input devices and proved the mouse to be superior, this is supported by Jones (1991) who evaluated the performance of children aged 6, 8 and 10years old, when using input devices (mouse, trackball or joystick) and proves that mouse is a superior pointing device compared to others. Although, the cognitive demand of using a computer mouse might create a more challenging experience for younger children who are yet to develop the necessary processing skills (Revelle & Strommen, 1990). Additionally, Strommen (1993) argues that children must attend to both the cursor's movement on the screen and the rules necessary to operate the specific device. On another aspect, in comparing two interaction styles (point and click to drag and drop), Inkpen (2001) proved that the point and click interaction style is faster and allows fewer errors than the drag and drop style. This therefore confirms (Strommen, 1993) statement, who argues that with cognitive modelling the degree to which cursor movement devices resembles natural movements (such as pointing), makes use easier for young children.

2.4.5.2 *Design with Children*

In design, early work were those by Druin & Solomon (1995); Druin (1996); Oosterholt, Kusano, & de Vries, (1996); Scaife, Rogers, Aldrich, & Davies, (1997) that situates more on designing with children. Oosterholt et al (1996) describing the process of co-design with children points the need to enter their world and to recognise the sophistication of the current generation. Druin et al, further made adaptations to three existing methods (contextual inquiry, technology immersion and participatory design) originally designed for adults and carried out an exploratory study to understand the wants and needs of children, to get an understanding on what children do with large technologies and to get direct information on what they would like to see in future technologies designed for them (Druin et al., 1998). Researchers in the design area advocates for more active roles for children when designing with and for children. These roles have been classified by Scaife et al (1997) as users, informants and design partner and has been described a little differently by Druin (1999b) as users, testers, informants and designer partners. The influential work by Druin (1999a, 1999b) on cooperative inquiry and participatory design has sustained and moved this debate forward such that children are now being involved in the different roles. For example, as users (Baauw et al., 2005), testers (Donker & Markopoulos, 2002; Donker & Reitsma, 2004; Markopoulos & Bekker, 2002) (Gilutz & Nielsen, 2002), informants (Dindler *et al*., 2005; Guha et al., 2004; Mazzone, 2008; Scaife &

Rogers, 1998), design partners (Dindler *et al*., 2010; Druin, 1999a; Guha *et al*., 2013; Read *et al*., 2013) and as stakeholders (Iversen *et al.*, 2010).

Other design efforts are tailored devices that fit the capabilities and activities of children with Hourcade *et al's* (2008) Point-Assist mouse that adjusts its speed to suit each individual child's ability in accordance with the difficulty level of the pointing task which is ascertained from sub-movements. Tangible interfaces and interactive table-tops are also an area of interest for the CCI community (Read & Markopoulos, 2013). For example, Zuckerman *et al*'s work investigates using children's natural skills for designing tangibles as an appropriate approach (Zuckerman *et al.*, 2005), Brederode *et al* (2005) further argued that tangibles support cooperation. On the other hand, Marco *et al.*, (2009) argue that multi-touch table-top displays is also appropriate for young children's play.

2.4.5.3 *Evaluation with Children*

In evaluation with children, several efforts have been put towards designing suitable evaluation methods for and with children (Read & Markopoulos, 2013). An example of early work in this area is the influential work by (Hanna et al., 1997) that proffers guidelines on how to effectively involve children in a usability testing from an industrial point of view. Other works include: the verbalisation technique (talk/think aloud method) (Als, Jensen, & Skov, 2005; Baauw & Markopoulous, 2004; Donker & Markopoulos, 2002; Donker & Reitsma, 2004), survey methods and interviewing technique (Read, MacFarlane, & Casey, 2002; Zaman, 2007, 2009) to evaluate children's user experience, likeability and preference of technologies, and the use of picture cards to illustrate usability problems found during a user testing (Barendregt et al., 2007). There is also the retrospective (MemoLine) method with children for evaluating the long term user experience of children's technology use (Vissers *et al*., 2013). All of these methods have similar component in that they are user based leaving a gap in inspection method, with SEEM (Baauw *et al*., 2005) as the up to date inspection method originally designed to evaluate children's technology but still uses adults as its evaluators. Some researchers have attempted to carry out the heuristic evaluation with teenagers (MacFarlane & Pasiali, 2005; Wodike *et al*., 2014) and with children (Salian *et al*., 2013) but these reported the need for adapting the HE method to be more suitable for children (as reported in chapter 1, section 1.1). Therefore, it is factual to state that there is no validated inspection method currently suited to children.

The CCI community desires effective and rigorous methods for the design and evaluation of children's technology (Read & Markopoulos, 2013; Read & Fine, 2005) and according to Jensen & Skov (2005) there is need for new methods and greater detail from researchers when describing methods employed in their work. Also as upheld by CCI researchers that children are separate from adults and should have a say in designs and evaluations carried out on technologies suited to them. This suggests the need for suitable inspection method where children will act as the evaluators.

2.4.6 Challenges of CCI

Although the community is constantly addressing issues by involving in diverse forms of research to better meet the need for children in the area of technology usage; but according to (Read & Markopoulos, 2013) children's place and usage of technology in the future is in an unimagined state therefore, the community suffer a key challenge of providing a body of research that will 'better inform' the designs of future technologies and the 'shape' of future spaces. Empirical work, design driven research and the development of robust methods are all needed.

Other challenges which have also been documented by (Read & Markopoulos, 2013) that needs be considered when carrying out future research in this community are as follows:

- **Closer link of theory to design –** The community is enthusiastic about building cool things and creating dynamic fun applications but perhaps has forgotten some of its roots. Despite the wealth of published work on child development, educational theories and perspectives on interaction design, the link between such theoretical works and interaction design practices is weak. Models and guidelines that could guide the design of interactive artefacts for children are few and far between. This reflects the relevant scarcity of empirical research on children to develop models that can guide design.
- **Children and their participation –** The social science research has recorded a considerable debate on the way children participate in their work. Children's participation, as social actors, as designers, as testers and as users is understudied and under explored in CCI. Following the issues of accessibility, CCI research can explore new opportunities offered by interactive technology to support the development of varying groups of children, to support their participation in the design process and their emancipation as stakeholders in technology design and evaluation.

- **Supporting 'Family' Communication, Play and Learning –** Play and learning have been a traditional focus of the CCI community. The emergence of mobile and pervasive technologies, tangible and embodied interaction has created an immense space of opportunities that is yet to be explored especially in the context of bringing families, in the broadest sense, into playful learning and communication spaces. Relevant design explorations need to be matched by child and family appropriate technology development, and methods to support the design and evaluation of related applications.
- **Stories in the Cloud –** Storytelling has been a feature of CCI research since its early days. Initially concerned with writing and retelling stories in text spaces, the new stories are combinations of digital extracts. Children are major users of social media and a large part of research in this community explores appropriate communication technologies, for communicating with collocated and remote friends, for communicating and socialising with family. Privacy, security and trust, risk and risk taking are all features of this environment but so are concerns about memories and keepsakes as children and those around them only capture and keep digital scents as opposed to 'boxes of physical stuff'.

2.4.7 The Future for CCI

The rate at which computers and technologies are being used today is quite different from what it was decades ago; as children of those times differ greatly from children of the present. In other words, children are changing and so is their capabilities (Read & Markopoulos, 2013). Their daily use of computers is increasing both at school and at home (Subrahmanyam et al., 2000). Children are increasingly becoming owners of personal computers, mobile devices, and personal web accounts which they use for entertainment, completion of school work, interaction and communication (Subrahmanyam et al., 2000). Therefore, it is crucial that as the community is adapting to these new technologies, it makes provisions to also adapt to this changing user group (Horton, 2012).

Rigorous and thorough methods are required within CCI to determine the suitability and usability of the future technologies for children. Methods adopted from adult based methods need to be appropriately modified to suit the wants and needs of children. In addition child centred method should be constantly tested and improved to ascertain its suitability for the children at each given time.

2.5 Conclusion

This chapter has reviewed literature on the target group (children), analysed theoretical approaches of their social, behavioural and cognitive development. The chapter has also reviewed literature on the CCI community looking at its history, subjects, popularity, involvements, challenges and future which concludes the chapter.

In this chapter it was revealed that children are different in many ways compared to adults. For example, in the way they represent data (Druin, 1999a), in the way they answer questions (Maier, 1978), and that they do different activities, have different behaviour and concerns with computer compared to adults (Read, 2005). Therefore, they need suitable methods when evaluating their own technologies. Literature also show that in designing for children, there is need to consider their cognitive and physical capabilities and also explore their ideas on how best to create suitable designs and methods for them.

Knowledge was gathered on the vibrant work carried out by the CCI community on developing rigorous methods for the design and evaluation of children's technology. Also, the CCI community is very interested in the future technologies for children and in future children, which shows the community intends to grow with the change in technology and children. Therefore, it solicits for rigorous methods not just in design but also in evaluation for and with children.

The next chapter (three), will review the literature on evaluation methods in HCI and CCI in general and more in-depth on how children have been involved in evaluations using different evaluation methods.

2.5.1 Contribution to Thesis

This chapter has helped to reveal the target group (children aged 6 and 11years old) of this research. Their capabilities, their developmental process and how they are involved in research especially in CCI research. An understanding of this gives a direction on how children could be successful involved in this research given the research aim.

3 CHAPTER THREE: UNDERSTANDING EVALUATION METHODS

3.1 Introduction

The chapter presents a review of related literature in evaluation methods and reports its importance to this research. The chapter is divided into five sections: Section 3.2 introduced the chapter by explaining evaluation method and its use within HCI. Section 3.3 Reports on general aspects of usability evaluation methods and identified usability evaluation methods for children

3.2 Evaluation Methods in HCI

The commonly used approach in HCI nowadays when designing quality interactive products is the user centred design (UCD) approach (Barendregt et al., 2007; Nielsen, 1993a) which allows the product user to be at the centre of the entire design process (Barendregt *et al*., 2007). The design process of interactive products is usually iterative and involves design, evaluation and redesign, placing evaluation at an integral part of the process (Hartson *et al*., 2001). This concept of evaluation dates far back to the start of systems analysis and human factors (Hartson *et al*., 2001) and can either be summative or formative.

A summative evaluation involves the test of a finished product against certain success criteria or the comparison of a product with an alternative product based on certain criteria (Barendregt *et al*., 2007; Hartson *et al*., 2001). Formative evaluation is the testing of different aspects of a product at its prototype stage to identify issues that will be corrected before a final product is accepted for release (Barendregt *et al.*, 2007; Hartson *et al.*, 2001). When carrying out a summative or formative evaluation of interactive products, there are several outcomes that can be reached; however there are two popular outcomes recorded in literature. It is either a product is evaluated for usability or for user experience; in other words, there are two major aspects of evaluations in HCI: usability and user experience.

Usability or user experience evaluation is an important part of interactive product design process, which in some cases consists of iterative cycles of designing, prototyping and evaluating (Ivory & Hearst, 2001). Usability have been since the 1970s (Scholtz, 2004) although the term 'usability' came into general use in the 1980s where 'user friendliness' and 'ease of use' have been the previously used terms in professional and technical writing (Lewis, 2006). According to Nielsen and Mack, usability is a fairly broad concept that refers to how easy it is for users to learn a product, how efficiently they can use the product once they have learnt it and how pleasant it is to use (Nielsen & Mack, 1994). Apart from this explanation of usability, other definitions exist (e.g. Nielsen, 1993a; Skov & Stage, 2005) but the acceptable definition is by ISO who defined the term 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* (ISO 9241-11, 1998). There are different attributes to compare or assess the level of usability of an entire product, for example within ISO's definition the identified attributes are: efficiency, effectiveness and satisfaction (Barendregt, 2006). However, Nielsen, (1993a) and Scholtz (2004) identified five attributes: Learning, Memorability, Error, Efficiency and Satisfaction. Although Scholtz believes one usability attribute might be more critical in an evaluation than the other depending on the particular application being evaluated. For example, if an application will not be used frequently, then memorability is key as it will be important that the user (s) is able to remember the actions needed for a desired task (Scholtz, 2004). Usability has been the trend of user interface evaluation and some researchers argue it should be the main goal of interactive products (Sousa *et al*., 2015). However, HCI researchers sees the need to evaluate more aspects of the user interface and interactive products which brought about the 'user experience' evaluation.

User experience (UX) is a recent form of evaluation (Xu *et al*., 2009) with different attempts of definition (e.g. Alben, 1996; Bernhaupt, 2010; Hassenzahl & Tractinsky, 2006). Although, the acceptable definition by ISO reads thus: *a person's perception and the responses resulting from the use or anticipated use of a product, system, or service* (ISO 9241-210, 2010). This type of evaluation explores a user's functional, psychology and emotional needs of using a product (Väänänen-Vainio-Mattila & Wäljas, 2009). In other words, it involves capturing a person's physical, sensual, social, aesthetic and emotional experiences (Xu *et al*., 2009) or capturing a person's feeling about a product use (Vermeeren *et al*., 2010). This thesis will however, focus more on the usability aspect of evaluation.

3.2.1 Usability Evaluation Methods (UEMs)

In evaluating usability, some methods are required which are referred to as Usability Evaluation Methods (UEMs). UEMs are typically used formatively during the prototype design phase (Hartson *et al*., 2001) and popularly have three classifications (Barendregt *et al.*, 2007; Scholtz, 2004):

- User testing or Empirical or User Based Method
- Inquiry Based Method
- Predictive or Analytical or Expert or Inspection Based Method

Although Nielsen (1994) identified four classifications: Automated (e.g. the use of programs), Empirical (the use of real users), Formal (use of models and formulas) and Informal (the use of experts based on rule of thumbs). However this thesis will focus on the three classifications previously identified: user based, inquiry based and expert or inspection based evaluation methods as these are the commonly used terminologies in most HCI/CCI publications (e.g. Barendregt et al., 2007; Barendregt, 2006; Hartson et al., 2001; Scholtz, 2004).

3.2.1.1 *User Based (Testing) or Empirical Evaluation Method*

The User Testing or Empirical evaluation method is an evaluation that involves the real or end users finding usability problems (Capra, 2006; Scholtz, 2004) where findings are measured on certain usability constructs (de Kock et al., 2009) to determine the usability findings. Usually this type of evaluation is task based carried out within a controlled setting (Jacobsen et al., 1998). Examples of user testing methods include: user study (usability test) or observational technique (Barendregt *et al.*, 2007), think aloud, co-discovery, active intervention and retrospective method (van Kesteren *et al*, 2003).

3.2.1.2 *Inquiry Based Evaluation Method*

The inquiry based evaluation method involves the use of models to gather users' written or verbal opinion on their likes, dislikes, needs and understanding of a product. This method tends to identify broad usability opinions or problems about the entire product (Barendregt *et al.*, 2007). Examples of inquiry evaluations are: Focus groups, User Satisfaction Questionnaire (Barendregt *et al.*, 2007), Interviews (Scholtz, 2004).

3.2.1.3 *Analytical or Expert or Inspection Based Evaluation Method*

The analytical or expert based method popularly referred to as usability inspection method (UIM) involves the use of usability experts to predict usability problems with set guidelines for the purpose of modifications of the product before a final version is released without the need to involve real users (Cockton, Lavery, & Woolrych, 2002; 2003; Hollingsed & Novick, 2007; Mack & Montaniz, 1994; Nielsen, 1994). There are different usability inspection methods as listed below. However only the Heuristic Evaluation (HE) Method will be discussed, because it is the only inspection method that has been attempted with children.

- Cognitive Walkthrough (CW)
- Pluralistic Walkthrough
- Heuristic Walkthrough
- Guideline Reviews
- Structured Expert Evaluation Method (SEEM).
- Formal Usability Inspection
- Feature Inspection
- Consistency Inspection
- Standards Inspection
- Heuristic Evaluation (HE)

Amongst these three UEM classifications, the debate on what method to use have always been between the empirical and the expert based method (Nielsen, 1994). On one hand, the empirical method is argued in literature as the most effective and reliable method for gathering real usability problems and is used as a yardstick for other UEMs (Jacobsen et al., 1998). However, it is viewed as the most resource consuming method as it is not cost effective and does not have the option of having real users as real users must be recruited. The inspection method on the other hand, is argued to be the most informal (Nielsen & Mack, 1994) and cost effective method where evaluations could be performed without the presence of the actual users (Nielsen, 1994). Although there is yet to be a validated usability inspection method with children.

3.3 Evaluation Methods with Children

Previously UEMs were more adult centred where method developed were designed to suit adults alone even for evaluations of technologies that is suited to children (van Kesteren et al., 2003) that is, adults are also used to evaluate technologies designed for children. However, the emergence of the CCI community has evolved the development of UEMs such that there are now many methods that is effective for evaluating technologies with children (Xu et al., 2006). Such methods include:

1. Fun Toolkit (Read *et al.*, 2002),
2. Contextual Laddering Method (Zaman, 2007),
3. Think Aloud (Donker & Markopoulos, 2002; Hanna *et al*., 1997),
4. This or That method (Zaman, 2009b),
5. Problem Identification Picture Cards (PIPC) (Barendregt *et al*., 2007),

6. MemoLine (Vissers *et al*., 2013).

3.3.1 Fun Toolkit

The Fun Toolkit (Read *et al*., 2002) is a suite of tools designed to gather the opinion of children about technology; specifically it measures three dimensions of fun (Read *et al*., 2002). In the first version of the Toolkit, four tools were used: The Smileyometer, Funometer, Again-again table and the Fun Sorter. However, the Smileyometer and the Funometer were found to be very similar therefore the Funometer does not appear in the most recent version.

3.3.1.1 *Smileyometer*

The Smileyometer is the most widely used tool within the Fun Toolkit (Read, 2007) and is a Visual Analogue Scale (VAS) based on a 5 point Likert scale with 1 being awful and 5 being brilliant, see figure 3.1. It is reported to have been co-designed with children, whom assisted in the design of the faces. Its key attributes are (Read, 2007):

- Easy to complete
- Quick to complete
- Requires limited reading ability
- Requires no writing

In use it is advised to be administered before a child uses technology in order to measure the child's expectation of the technology and also afterwards to measure the child's experience (Read, 2007).

Figure 3.1 Smileyometer: A Visual Analogue Scale

3.3.1.2 *Fun Sorter*

The Fun Sorter allows children to rank items against one or multiple constructs. It is designed to collect the child's opinion about a technology or an activity.

To complete the Fun Sorter, children need to interpret the construct being measured (e.g. fun, ease of use) and then write a description of the technology in the appropriate blank cell. In the case where a child is poor at reading or writing or sometimes to save time and prevent children from too much writing (Read, 2007), pre-prepared picture cards are provided for children to place on the appropriate cell. After ranking the technologies, a ranked score can be applied to each technology/construct under consideration. It is recommended that a single Fun Sorter is used for each technology or construct.

3.3.1.3 *Again-again Table*

The Again-Again table is a table that requires the child to tick either 'yes', 'maybe' or 'no' for each technology or activity, following a question that is written above the table which is 'would you like to do it again?' It is reported that the idea of this tool originates from psychology work that supports the fact that we are most likely to return to an activity that we have liked (Read, 2007). It is related to the endurability of an activity, as well as the engagement felt whilst doing it.

This tool is most useful where three or more technologies or activities are being compared. Items compared are usually represented in the rows as the options to tick takes the column position. It is suggested that items should be presented on a single sheet after the children have experienced all technologies. However, it is recommended that too many items shouldn't be compared as too many rows to consider might put the children off.

3.3.2 Contextual Laddering Method

Contextual laddering method popular referred to as the laddering method (or laddering) (Zaman, 2007), was adapted from the means-end theory proposed by Gutman (1982). Means-end theory is mainly used with adults within marketing and consumer research to help understand and describe how consumers perceive products. This is achieved by revealing the core underlying values that motivate consumers to desire certain product. The laddering method with children, seeks to evaluate whether children like a particular technology, whether the technology is usable and also aims to understand why the technologies are liked. In contrast to the means-end laddering, laddering with children is viewed in a particular context usually the context in which the technology is being used; therefore, it is referred to as the contextual laddering method. laddering can evaluate many aspects of the software including the likeability, usability (Zaman, 2007), as well as the user experience (Abeele *et al*., 2011).

The method involves user testing in the first stage (using observational techniques) followed by an interview session (likeability test). The questions within the interview are called the laddering questions, as the facilitator will ask the child a question as to why they liked the technology, when the child responds, another why question is asked to elicit attached conditions to the response. This process continues until all possible conditions are uncovered.

There is no appropriate number of questions reported, it is only recorded that children were frustrated when they were asked ten questions. As a result, fewer questions are recommended

(Zaman, 2007). It is also dependant on the ability of the facilitator to notice signs of frustration, ask suitable questions and make the child at ease.

3.3.3 Think Aloud with Children

The think aloud method which is used for usability evaluations is well documented in literature. It has been used solely with children (Donker & Reitsma, 2004; Zaman, 2005), complementarily (Vanden Abeele *et al*., 2011) and comparatively (Als *et al*., 2005; Baauw & Markopoulous, 2004; Donker & Markopoulos, 2002) with other methods. It was adapted from the standard think aloud method designed for adults (Duncker, 1945; Ericsson & Simon, 1993), which is based on the older introspection method (Barnard & Sandberg, 1994; Donker & Markopoulos, 2002). Introspection is based on the idea that one can observe events that take place in consciousness, more or less as one can observe events in the outside world (Barnard & Sandberg, 1994). Some of the adaptations in TA with children includes: children literally verbalizing their total experience with the application during the evaluation. This technique is referred to as talk aloud (Donker & Markopoulos, 2002). Another adaptation is using constructive interaction technique whereby children are allowed to think aloud naturally while the collaborate to solve tasks (Als *et al*., 2005).

TA with children works by children using a technology whilst at the same time verbalizing their experience. During this process, usability problems encountered are usually captured. The method is carried out alongside observational techniques to note problems children encountered. This observation could either be active (person observing) or via video recording. Some researchers believe this method of evaluation is quite challenging for children due to its cognitive demands (Donker & Reitsma, 2004) especially for younger children (Hanna *et al*., 1997) as they could forget to think aloud unless being prompted (Barendregt *et al*., 2007). Though prompting could make them state non-existing problems in other to please the facilitator (Donker & Reitsma, 2004). Also the technicality of the method require that the facilitator is experience (Barnard & Sandberg, 1994), therefore inexperience or new researcher in this field might not be qualified enough to do the process.

3.3.4 This or That (Pairwise Comparison Scale) Method

The ‘This or That’ method (Zaman, 2009a) was developed based on the pairwise comparison method. This method has been used in different contexts to gather input from participants on the value of one attribute over another (Chan *et al*., 2012; Kakiashvili *et al*., 2012; Tu & Zhou, 2000). It has its empirical history from Thurstone’s law of comparative judgment (Woods *et al*., 2010). The ‘This or That’ method is used to gather feedback from children (pre-schoolers) on their user experience of one digital technology over another. This method consists of asking the child direct questions to stimulate their choice between two conditions. This is actually done by the facilitator who actively points to two alternative technologies asking specific questions ending with this or that and allowing the child to respond by pointing to the preferred technology. There are five questions in total relating to fun and the scale has been proven to be internally consistent. However, there is some lack of clarity from the literature as to how these five questions were derived and how they were judged to be suitable for children of preschool age.

3.3.5 The Problem Identification Picture Cards (PIPC)

PIPC (Barendregt *et al*., 2007) is a formative evaluation method designed to help children aged 5 to 7 years report usability and fun problem while playing a computer game. Problems are reported verbally and nonverbally using picture cards that represent different problems children might have encountered. These picture cards serve as memory aids, for usability they were selected based upon Norman and Draper's (1986) perception, cognition and action model, and for fun, Malone and Lepper's (1987) taxonomy for what makes a computer game fun was used. The suitability of these pictures was determined based on emotions of children that could be portrayed in a picture and expressions that the children used in an earlier evaluation, along with Stienstra and Hoonhout's fun questionnaire (Barendregt et al., 2007).

Children get an explanation of what each picture represents and are informed for which situation each can be used. The box of picture cards is placed on the table next to the computer on which the game is played. Above each compartment of the box, the concept represented by the picture is printed. Children are expected to play the game, verbalize their experience and or use the picture cards to represent their experience. During the process, children could ask for explanation of the meaning of any card.

3.3.6 MemoLine

The MemoLine (Vissers *et al*., 2013) is a method that was adapted from the UX curve method (Kujala *et al*., 2011), that guides children to retrospectively report on their longitudinal user experience of an application. In the UX curve method, a timeline in the form of a graph is used in which users indicate positive and negative experience over a period of time. The UX curve required adapting for children, as the authors believed that children would lack the knowledge to draw a mathematical curve with negative values. Therefore, as an adaptation process to make the method child friendly, the curve was replaced with a timeline and coloured pens were

provided to draw periods of different experiences. For example a green colour indicated a positive experience, a red colour indicated a negative experience and a grey colour indicated non-usage.
The authors reported that children could face difficulties when trying to recollect events that occurred in a particular time frame. Therefore they asked children to identify different activities they had participated in during the period of the evaluation and they plotted these on the timeline to provide a visual recognition cue. Four constructs of UX evaluation were identified: these being Usability, Challenge, Quantity of play, and General Impression. For each construct a timeline is provided, that is accompanied by a question that explains the construct and a legend to explain what each colour signifies.

These methods are formative though some of them could be used as a summative method (e.g. The Fun toolkit, Laddering and This or That methods). Most of these methods are also most appropriate to evaluate the user experience rather than usability which this work is focused. Example of the usability methods are: Think Aloud and the PIPC method, and these two methods are user based leaving a gap in inspection method with children (which is the focus of this research) as SEEM is the up to date inspection method for evaluating technologies designed for children. However, it uses adults as its evaluators.

3.4 The State of Usability Inspection Method (UIM) with Children

UIMs as stated earlier are expert evaluation methods that allow a small group of experts inspect an application based on a set of guidelines or criteria. It is widely researched within the HCI community and its benefits have been documented (Cockton & Woolrych, 2001; Nielsen, 1994):

- UIMs can be used early in the development life cycle, including for use with low-fidelity prototype (Cockton & Woolrych, 2001; Nielsen, 1992). In this instance errors found could be corrected before huge resources are invested in the development of the application.
- UIMs can be used without exhausting or biasing a group of test participants.
- UIMs can be used to identify potential problems that can then act as a focus for user testing. Since more than one evaluation method is recommended for a thorough usability evaluation of product (s) (de Kock et al., 2009; Woolrych & Cockton, 2002).

A number of UIM exist as reported in section 3.2.1.3 above, however, only one of these methods have attempted to use children as the UIM evaluators on children's technology: the heuristic evaluation method.

3.4.1 Heuristic Evaluation (HE) with Children

The heuristic evaluation method was originally developed by Nielsen and Molich (1990) who developed nine heuristics and later Nielsen (1992) refined the heuristic set with the addition of another heuristic. The procedure requires a small group of experts to inspect an application, aided by a heuristic set (guidelines) with intention of refining the application to produce a more usable one (Nielsen, 1992). The method has two stages, the experts individually evaluate the system predicting problems while using the application and allocating a severity then the experts collectively aggregate their problem set into a single list of problems and agreeing on a final severity. Experts in this instance are people experienced in carrying out usability evaluations and also experienced in the application being evaluated, referred to as double expertise (Nielsen, 1992).

Heuristic evaluation is the most cost and time effective UIM (Jeffries et al., 1991) and the most easy to learn (Nielsen, 1992) UIM. Apart from these benefits, it has documented a weakness of the reliability of effectiveness measure (de Kock et al., 2009; Woolrych & Cockton, 2002). However, the HE is the most widely used UIM (Chattratichart & Lindgaard, 2008; de Kock et al., 2009) and has been used successfully to evaluate applications for adults and children in different contexts. Though for both user groups, just like SEEM, adults have always acted as the evaluators (Alsumait & Al-Osaimi, 2009; Korhonen, 2011) thus may be a useful addition within the CCI community.

According to (Nielsen, 1992) the HE evaluator should be a double expert (expert in the evaluation method and expert in the domain being evaluated). However, it is debated that when the evaluation is on children's technology, then double expertise might not be enough but rather triple expertise (understanding of the game, the user target-children and an understanding of usability) (Wodike *et al*., 2014). It is therefore, useful to let children play the role of the evaluators on technologies designed for them. Though it is questionable if children could become usability experts but they could act as the evaluators given that novices have successfully acted as HE evaluators (Nielsen, 1992; Pinelle *et al* , 2009). Also it has been shown in chapter two that children have the cognitive capability to be successful usability evaluators, especially older children (children from age 7) (Hanna *et al*., 1997).

MacFarlane & Pasiali, (2005) carried out a heuristic evaluation with 15 teenagers (aged 13 and 14years old) on an online language tutorial. In this study they were interested in evaluating usability and fun so they created a heuristic set for the study based on Nielsen's (1994) usability heuristics and Malone's (1982) heuristics for fun. The children spent a few minutes exploring the software and 20 – 25 minutes on the evaluation tasks. Although, the authors suggested findings from the study were encouraging as the children were almost all interested participants, they did not clearly state whether the study was a success and what the criteria for success was. However, they recommended that the heuristics might need rephrasing for future use with children.

Wodike et al., (2014) also carried out the heuristic evaluation with 20 teenagers (aged 12 and 13 years old) on a ninja game (NINJAGO) that the authors judged as appealing to the age group, using Pinelle *et al's* (2008) game heuristics and Nielsen’s severity rating scale. In this study, 5 teenagers were empowered (trained) in the method to further act as facilitators over their teen peers. Their findings show the children were able to find usability problems but they encountered so many issues that led to the conclusion that the evaluation was unsuccessful. The authors recommended that in the future the method should be made more fun and engaging for children and that the tools (heuristic set and severity scale) used within the method should be made more suitable for children.

Salian *et al*., (2013) used the method with 14 children (aged 10 and 11years old) whom were to evaluate a music making game (JamMo) for between 15 – 20 minutes using Korhonen & Koivisto's (2006) playability heuristics for mobile games to describe problems found and Nielsen’s severity rating scale to rate problems found. This work showed that there was limited success as children encountered various problems including:

1. Children’s struggle to understand the language used within heuristic set.
2. Children inability to understand the severity rating construct.
3. Children’s inability to understand the merging process.

These issues would therefore suggest that the heuristic evaluation in its present state is not suitable for use with children. Salian et al., (2013) recommend adaptations be made to the method to make it suitable for children. The recommendations given were:

- The need to reduce the number of heuristics and rephrase the heuristics terms
- There is also the need for alternative severity rating that children can understand.

- It was also inferred that number of papers with documentation required to carry out the evaluation be reduced and recommended that forms used for recording problems be simplified.
- Finally their work recommended that more interactivity should be allowed for the merging phase; to encourage discussion amongst the children during this phase.

In view of the studies reported above, this work intends to explore the HE method with children to assess children's performance in a UIM evaluation (this is to answer RQ1 and RQ2). However, how will suitability and effectiveness of the method be determined from the children's performance?

Just like every other usability evaluation method, it is documented in literature that the HE method suffer the challenges of Evaluator Effect and Validity. In addition, similar to other UIM, determining the effectiveness of the method is an issue.

Evaluator Effect is described as the differences in evaluators' detection of problem and severity ratings (Hertzum & Jacobsen, 2003). This subject has been explored by different researchers, and literature show that the effect vary with evaluator experience (Ling & Salvendy, 2009), problem severity, task-type, system-type, and other variables important to usability practitioners and researchers (Jacobsen *et al*., 1998).

3.5 Measuring the Success or Effectiveness of UEMs

Several research works have been carried out to assess the effectiveness of usability evaluation methods. Examples of such work include: Sears' (1997) work which measures usability evaluation methods on three criteria: Reliability, Thoroughness and Validity. These criteria are explained thus:

Validity

Sears defines "validity" (method for the purpose of this research) as evaluators' capability on focusing on relevant issues; that is, evaluators' ability to identify only the issues that impacted them during the evaluation. This is measured as the proportion of real usability problems found by a UEM compared to all the issues identified as problems (see formula below).

- Validity = $\frac{Number\ of\ Real\ or\ Serious\ Problems\ Found}{Number\ of\ Issues\ Identified\ as\ Problems}$

Calculating Validity:

Validity will be calculated using the usability figures from (Sim, 2009, p 128), where

- The number of Real Problems found by the evaluators = 6

(Real problems that was found by the evaluators are confirmed as real problems either by a user testing or through falsification testing. In an ideal scenario, this is part of the real/actual problem that exist). In this (Sim's) study, the number of real problems were decided not through falsification testing but via the outcome of the evaluations. That is, these said real problems were found by evaluators of a user study and by the evaluators of a heuristic evaluation.

- Number of issues identified as problems (both real and unreal problems) = 24 this is the total number of problems predicted via the heuristic evaluation study

- Number of Real Problems that Exist = 22

(Those that is confirmed as real problems prior to the evaluation because these problems were intentionally put into an already tested application for the purpose of the study or was identified via a user study) In this (Sim's) study, the study participants of the user study were able to find a total number of 22 problems which is determined as the total number of real problems that exist in the application that was evaluated.

Validity = # of real problems found / # of issues identified as problems = 6/ 24 = 0.25

Thoroughness

Thoroughness by Sears is defined as evaluators' capability to evaluate the entire aspects of the interface. It is measured by the ratio of real problems found to the number of problems that exists (This is represented with the formula below)

- Thoroughness $= \frac{Number\ of\ Real\ or\ Serious\ Problems\ Found}{Number\ of\ Real\ Problems\ that\ Exist}$

Calculating Thoroughness:

Thoroughness will be calculated using the same figures used for validity above

- The number of Real Problems found by the evaluators = 6
- Number of issues identified as problems (both real and unreal problems) = 24 this is the total number of problems predicted via the heuristic evaluation study

- Number of Real Problems that Exist = 22

Thoroughness = # of real problems found / # of serious problems that exist = 6 / 22 = 0.27

Also according to Sears, calculating the denominator of thoroughness may be difficult as this may differ with different studies (Sears, 1997).

Reliability

Reliability is when a method is able to produce similar result in the similar condition. This is determined when different evaluators or group of evaluators are able to find similar number of problems when using the same evaluation method. This is calculated with the ratio of standard deviation of numbers of problems found to the average number of problems found (below is the formula for measuring the reliability of an evaluation method).

- $R_{Temp} = 1 - \frac{Standard\ Deviation\ (Number\ of\ Real\ Problems\ Found)}{Average\ (Number\ of\ Real\ Problems\ Found)}$

  Reliability = Maximum (0, R_{Temp})

Reliability = Maximum (0, R_{Temp}) where R_{Temp} is calculated as 1 – (standard deviation of the # of real problems found / average # of real problems found) – However, since the full details of the numbers are not obtained here in Sim's study to decide what is the standard deviation or average number of the real problems found, it is difficult to calculate the reliability. Although Sears rightly pointed that many technique could be used to calculate the reliability of evaluation techniques and also stated that another approach to measuring reliability might be to look at the specific problems identified by each evaluator rather than just the number of problems found.

Hartson et al. (2001) carried on Sear's formula in measuring the evaluation methods, though they further extended Sears' formula by including:

Effectiveness = Thoroughness × Validity, Down Stream Utility and Cost Effectiveness, stating it is for practical matters and for practitioners with tightly constraint budget.

They defined Thoroughness following Sears' definition as a measure showing the proportion of real problems found using a particular evaluation method to the number of real problems existing in the target interaction design.

Validity was defined as a measure of how well a method does what it is intended to do.

These are measures for methods for adults and other measuring metrics that exists for measuring method for children include:

Work by Read, Macfarlane and Casey, who looked at the suitability of text input methods for children using the task carried out with the method by measuring for effectiveness, efficiency and satisfaction (Read *et al.*, 2001). In their work, they described these criteria in the context of their work and is not clearly articulated to reflect general effectiveness measures.

Markopoulos & Bekker, (2003b) produced a framework for assessing the usability testing methods for children. Their framework involve the following criteria:

- Robustness - Will they be able to apply a particular UTM for their problem?
- Effectiveness - How good results does this UTM produce?
- Efficiency - How expensive it is to apply a UTM in terms of time or other resources?

Some other criteria were also analysed in their work. For example

Reliability - The reliability of a usability testing method pertains to whether factors external to the method influence its outcome when the same testing procedure is used for the same product.

Validity - Pertains to whether the problems it helps uncover are actually usability problems or not

Thoroughness aims to describe the proportion of all usability problems of a product that are found through a test

3.5.1 Defining Method Success for this Thesis

The aim of this research is *to investigate whether children can perform an effective inspection method evaluation (IMCH) on technologies designed for them based on their values.* As a result, effectiveness is one criteria that will be assessed to determine the success of the method. In addition since the method (IMCH) is developed from the HE method original designed for adults, the ease of use of the Inspection Method for Children (IMCH) for the target group (children) will be ascertained. Using this (Ease of Use) as one criteria for assessing a usability method where children are the evaluators is not new in CCI as the Fun Toolkit was validated with this as one of the criteria. This is logical as these methods are adapted from methods for adult and important that the adapted method is easy for children to use if it is targeted at them. ‘Ease of Use’ from Read’s work refer to children’s ability to understand and use the method as it should be used (Read, 2007). Also children have been used in a study by MacFarlane *et al.*, (2005) where ‘ease of use’ was one of the constructs (criterion) being determined from the study. They have described the criterion (ease of use) as children’s ability to easily use a product. Many criteria for assessing usability methods are presented as quantifiable (Hartson *et al*., 2001; Sears, 1997) and some have made textual description (Markopoulos & Bekker, 2003b). From these styles the latter (textual description) has been successfully used for methods tended to children therefore, will be used for the method in this research.

'Ease of Use' is reflected in Read's work, where this is described as children's ability to understand and use the method right. Therefore, in this research, **ease of use** will be determined by children's ability to understand the instructional language, tools used (e.g. severity scale) and finally are able to use the method as described.

Effectiveness described by other researchers as reported previously is a combination of validity and thoroughness where validity was described as either a measure of how well a method does what it is intended to do or whether the problems a method helps uncover are actually usability problems or not. Thoroughness is described as either showing the proportion of real problems found using a particular evaluation method to the number of real problems existing in the target interaction design or the proportion of all usability problems of a product that are found through a test. Although, Markopoulos and Bekker (2003b) described effectiveness as how good the results produced by a UTM will be. **Effectiveness** for this thesis will be determined by the child evaluators' ability to use the method correctly (in the right order) to find real usability problems.

In order to measure suitability, observational data will be collected to determine whether language and tools used within the method and the method process were easy and understandable for the children to follow. While effectiveness will be determined from the observational data and usability data produced by the children.

3.6 Conclusion

This chapter has been able to review the evaluation method literature in HCI and CCI focusing on usability inspection methods; specifically on heuristic evaluation method. It shows that though several research exist for usability evaluation method with adults as evaluators, there is scarcity in literature on inspection methods with children as evaluators. Although several attempts have been made in heuristic evaluation (e.g. MacFarlane & Pasiali, 2005; Salian *et al*., 2013; Wodike *et al*., 2014) but conclusion drawn from these research shows there is still no validated inspection method with children as evaluators. The review of literature on UEMs in HCI and CCI, the attempted HE studies with children and the conclusion drawn from the studies motivated the decision to use ‘HE’ as the base method for this research (this will be better discussed in the next chapter – chapter four).

The chapter highlighted issues that face evaluation methods and especially UIM: Evaluator Effect, Validity and determining the effectiveness of the UIM method. Literature on these subjects were also review and the chapter concluded by highlighting the criteria for measuring method effectiveness and success for this research. The next chapter (four) will highlight and discuss the methodologies, methods and techniques adopted for this method.

3.6.1 Contribution of This Chapter to This Thesis

An understanding of evaluation methods and currently validated evaluation method in CCI was gathered with the review of literature carried out in this chapter. The review of literature from this chapter made it clear that there is a gap in inspection method, as there is currently no validated inspection method for children; the chapter also concretized the reason for this research. Finally, it was evident from this chapter that it is not enough to use a single criterion when evaluating or assessing the effectiveness of UEMs.

4 CHAPTER FOUR: RESEARCH METHODOLOGY AND DESIGN

4.1 Introduction

The aim of this thesis is *to investigate whether children can perform an effective inspection method evaluation on technologies designed for them based on their values*. In order to achieve this aim, approaches have been followed, and several methods and techniques have been used to collect and analyse data based on the research questions of this thesis. This was done in stages with full details shown in figure 1.1 but an extract is presented below, see fig 4.1

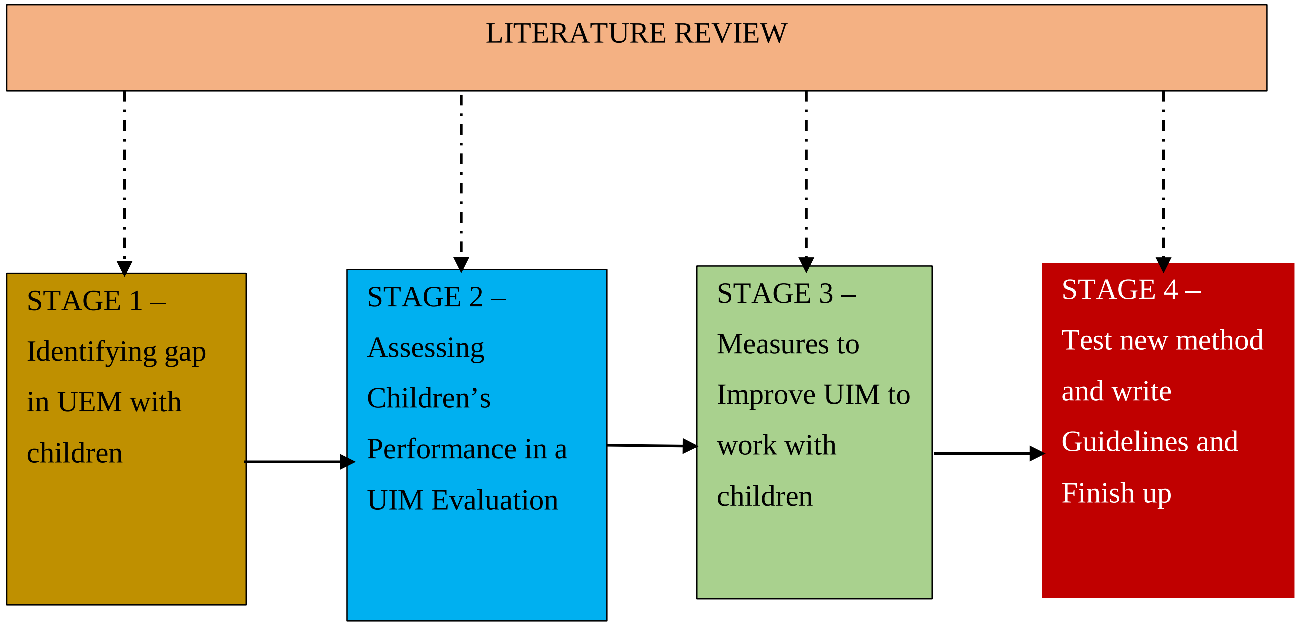

Figure 4.1Stages of this Research

In order to tackle the research aim and answer the research question, it was necessary to investigate, review and understand the target group and the base subject of this research (UIM); this has been carried out in chapter 2 and 3 respectively. In addition, to gather sufficient data necessary to interpret events within this research and produce a viable usability method within CCI and HCI, it was decided that a user centred approach and a mixed method approach will be adopted for this research. Following this decision, this chapter aims to discuss the research approaches, methods, and techniques applied in this research. In order to achieve this aim, the following objectives was undertaken:

- To identify research methodology available for this type of research and determine the one (s) applied in this research.

- To further identify methods and techniques used within the chosen research methodology.
- To report on the ethical implications, reliability and validity of methods used

The research was carried out following two approaches (from HCI and Social sciences) and used different methods and techniques which will be discussed according to the stages of the research design. Therefore, the chapter is presented in three sections: section 4.2 discussed the HCI design approach used. Section 4.3 discussed the general research approaches available for this type of research and identifies the approach used. Section 4.4 discussed other concerns of the research. Section 4.5 outlines the stages of the research and reports the method and techniques used within these stages and 4.6 concludes the chapter.

4.2 User Centred (UC) Approach

This research is rooted within HCI but specifically in CCI and Interaction Design. The research designed an evaluation method (UIM) that will be used by children to inspect the usability of interactive technologies designed for them. In order to achieve this, several studies were carried out using the user centred (UC) approach. This UC is a concept in HCI that allows the involvement of the real user group in the design and evaluation of interactive products (Barendregt *et al.*, 2007; Iversen *et al.*, 2003; Markopoulos & Bekker, 2003a). It is believed that through this means the researcher or designer will have better understanding of the needs, wants and requirements of the user group. In CCI following a UC approach, children's involvement have been classified in different roles: as users, testers, informants and design partners (Druin, 2002). Barendregt *et al.,(2006)* argues that the level of involvement of the children differ with the different roles but they all involve evaluation with children as the evaluators (Barendregt *et al.*, 2007; Barendregt, 2006).

4.2.1 UC within this Research

Within this research, children were involved in evaluations and studies where they acted as the evaluators and participants respectively. In chapter 5, 8 and 9 children played the role of study evaluators and in chapter 7 children acted as the design participants to design their own severity scale.

4.3 Research Approaches

Research is a systematic investigation to find answers to a problem (Burns, 2000). Research investigations are usually carried out following defined approaches which are the principles or philosophies that guides the research purpose (Dawson, 2002). In literature, three approaches have been identified thus:

- Quantitative approach
- Qualitative approach
- Mixed method approach

So, research can be qualitative, quantitative or mixed based on the research aim and objectives. In the CCI community, studies that seek to understand constructs and phenomenon in evaluation with children, often adopt the use of qualitative and quantitative approach to collect data in a single study. For example, Read et al., (2001) in their study with children used both qualitative (observation) and quantitative (questionnaire and key press times) technique to collect data. Methods were used based on what is being measured. Similarly, Barendregt et al., (2007) in their study with children to test the problem identification picture cards (PIPC) method, facilitated the children to identify and describe problems using pictures provided (This could be described as qualitative method of data collection). Also, at the end of test session, children were provided with a questionnaire (which is a quantitative method of data collection) to gather children's preferences of the game under study. It is believed that the more approach employed, the more chance of getting rich findings. Therefore, for the purpose of this research aim (*to investigate whether children can perform an effective inspection method evaluation on technologies designed for them based on their values*), a mixed method research approach was adopted.

4.3.1 Mixed Method Approach

The mixed method research approach emerged as an alternative to the contrast of qualitative and quantitative approaches (Teddlie & Tashakkori, 2009). It has been defined as a type of research design where qualitative and quantitative approaches were used in the types of questions, research methods, data collection and analysis procedures and inferences (Tashakkori & Teddlie, 2003). Following this approach there is need for quantitative and qualitative data collection to help understand phenomenon, interpret and produce rich findings, than if just one data type was collected. A mixed method research approach can be divided into two categories: true mixed or quasi mixed methods (Teddlie & Tashakkori, 2009), this is put a little differently by Tashakkori and Teddlie that it can be divided into mixed method research and mixed model research (Tashakkori & Teddlie, 2003).

- True Mixed Method Approach involves the collection and analysis of data, integrating findings and drawing inferences using both qualitative and quantitative methods in a single study (Tashakkori & Creswell, 2007).
- Quasi Mixed Method Approach involves the collection of two types of data with little or no integration of the two types of findings or inferences from the study (Teddlie & Tashakkori, 2009).

For this research both mixed method approaches were adopted. For example, the studies reported in chapters 5, 8 and 9 took a quasi-mixed method approach where qualitative data (evaluators' usability problems and observational data) which were collected from the participants (evaluators and observers) and quantitative data (from the severity rating scales and the heuristic set in the chapter 5 study) were also retrieved. In the heuristic studies the heuristic numbers were used to determine if the evaluators understood the heuristic sets by their correct linking of the heuristic to usability problems found; this was the basis of decision making alongside the observational data collected. Though if the observers' data were unavailable, this conclusion could also be drawn.

In chapter 7, a little part of the research (developing the severity scale) used a true mixed method approach where qualitative data (severity rating scales drawn by the children) was collected. Also a traffic light severity scale which was in a 3 point Likert scale format (quantitative approach) was used together with the children's severity drawing to determine children's ability to understand the severity scale.

4.3.1.1 Mixed Method Research Design

There are different designs or ways to carry out the mixed method research identified in literature. Morse (2003) identified two research designs within the mixed method, which includes:

- Simultaneous or parallel
- Sequential design.

Teddlie & Tashakkori (2009) first identified two types of mixed method: monostrand mixed method and multistrand mix method. This research is multistrand where research does not employ just one stage like the monostrand but rather involves all three stages of research strand (i.e. conceptualisation stage (e.g. drawing up a research question), experiential stage (e.g. collecting and analysing data) and inferential stage (e.g. making inferences and discussing results)). Therefore, the multistrand design type is the focus of this review. Within this multistrand design type, the following mixed method research designs were identified (Teddlie & Tashakkori, 2009):

- Parallel Mixed Designs
- Conversion Mixed Designs
- Multilevel Mixed Designs
- Fully Integrated Mixed Designs
- Sequential Mixed Designs

4.3.1.1.1 Parallel Design

This is a design that involves carrying out quantitative and qualitative study in a parallel manner either simultaneously (concurrently) or with some time difference. Both planned and implemented quantitative and qualitative phases answer related aspects of the same questions. This type of design is complicated, therefore might require different teams of researchers to conduct these studies. This might be most appropriate for a collaborative research work.

4.3.1.1.2 Conversion Design

This is also a type of parallel design, though in this design one type of data (e.g. quantitative or qualitative) is transformed (i.e. the qualitative is transformed into quantitative or vice versa) and analysed both quantitatively and qualitatively (Tashakkori & Teddlie, 2003). Data is gathered using one method, and further transformed and analysed using the other method. The design answers related aspects of the same questions.

4.3.1.1.3 Multilevel Design

This design can either be parallel or sequential where one type of data is collected at one level of the analysis and another type of data is collected at another level of the analysis stage in a parallel or sequential pattern. Both types of data are analysed accordingly, and results are used to make further multiple inferences.

4.3.1.1.4 Fully Integrated Design

The fully integrated design is a parallel design where the mix in the qualitative and quantitative approach occurs in an interactive or reciprocal way at all stages of the study. At each stage, one approach affects the formulation of the other, and multiple types of implementation processes could occur within this design.

4.3.1.1.5 Sequential Design

In this type of design, the mixing occurs across chronological phases (Quantitative, Qualitative) of the study. The results from previous type of study forms the design components or procedure of subsequent studies and the final result and conclusions are drawn from both types of studies.

Subsequent study (s) is usually carried out to confirm or disapprove or make further explanations of the results of the previous study. In other words, it answers exploratory and or confirmatory questions. The design is more beneficial for solo researchers compared to the parallel type designs.

4.3.1.2 Typologies of Mixed Method Designs

In following any of the above stated research designs, there are different mix method typologies that could be applied. Creswell & Plano Clark (2007) identified four typologies:

- Triangulation = QUAL + QUAN or QUAN + QUAL (Equal amounts of data is collected using both qualitative and quantitative methods at the same time)
- Embedded = quan → QUAL or qual → QUAN (a small amount of data (quan or qual) is collected using one of the approaches then (which the arrow (→ represents) more data (QUAL or QUAN) is collected using the other approach)
- Exploratory = QUAN → QUAL or QUAL → QUAN (An amount of data is collected using one approach then (→) the same amount of data is collected using the other approach)
- Explanatory = **QUAN** → QUAL (An amount of data is collected using quantitative approach then (→) the same amount of data is collected using qualitative approach.

Since this research involves the adaptation of a method originally designed for adults to create a new method that will be suited to children, there was the need to gain an understanding of the suitability of the method in its current state with children. Therefore, this research will involve the use of triangulation typology to gather data from the children, which will further be used to gather information from teachers (who are experienced in designing learning activities for children) on how to design the new method. Then an exploratory typology will be adopted to develop the new method and further explore it with children to ascertain the viability of the method with children. In view of this research intention, only the triangulation and the exploratory design typologies will be explained in more detail.

4.3.1.2.1 Triangulation Design Typology

In this type of design, equal type of quantitative and qualitative data is usually collected concurrently (at the same time) and merged during data interpretation or analysis stage. In a triangulation, one data type could be transformed to the other in order to interrelate the different types of data about a research problem. Also information from one data type could be used to validate the result of the other. For example, in chapter 5, a heuristic evaluation was carried out with children where usability problems, heuristic and severity numbers were collected and used to determine children's understanding of performing a heuristic evaluation. Findings from this study was further taken to teachers (in chapter 6) to gather ideas on how to design a better method for children. In chapter 7 of this research, part of the teachers' ideas were explored with children where qualitative data (children severity drawing was collected) and quantitative data (Children's severity judgement of problems using colours which was converted into quantitative data (numbers) were collected from children to determine children's understanding of the severity rating scale. More findings from the teachers (in chapter 6) and from the children (in chapter 7) were further triangulated to develop the method which is first reported in chapter 7 and explored in chapters 8 and 9.

4.3.1.2.2 Exploratory Design Typologies

Exploratory usually involves the sequential collection of qualitative data and then quantitative data where the mix of both data type is done between the two phases (i.e. by the development of an instrument or by quantitative testing based on the findings or outcome of the first phase). This type of design is used for instrument or taxonomy or theory development on the topic of research interest. An example of exploratory design typology applied in this research could be seen from the studies reported from chapters 5 to 9. Where each study carried out is sequential to the previous and the result of the previous always informed the decision of the next until the new method was designed, and tested with the children.

4.4 Other Research Concerns

When carrying out research, there are other issues that needs be considered, for example:

- Who will I recruit for my research? (Sampling)
- How am I going to involve them that is ethically approved? (Ethics)
- What type of data will I collect (Data Type)
- What measures can I take to produce reliable and valid data? (Reliability and Validity of Data)

4.4.1 Sampling

Creswell & Plano Clark (2011), defines sampling as a procedure that involves determining the location (site) and participants (type, number and method of recruitment) of the research. Teddlie and Tashakkori made a slightly different definition that it is the selection of the units of analysis (e.g. people, groups, settings) of the research (Teddlie & Tashakkori, 2009). This also means the small segment or sample of a bigger population on which the research is focused

(Davies & Hughes, 2014). The type of sampling is dependent on the type of mixed method design applied in the research; that is if the design is a sequential or concurrent (simultaneous) design type (Creswell & Plano Clark, 2007). Since this research used a sequential design approach, sampling will be reported as such.

4.4.1.1 Considerations for Sequential Sampling

In order to decide sampling for a sequential mixed method design, it is essential also to consider:

- The aim and research question of the research
- The design typology or purpose, see 4.3.1.2 (i.e. if the purpose of subsequent data collection is to explore or explain findings from the previous or it is to answer a secondary (embedded) question) For instance in an exploratory design it is recommended that different participants are recruited but in an explanatory the same participants could be recruited for the follow up (Creswell & Plano Clark, 2007)
- The type of data (qualitative or quantitative) that will be collected. For instance, if the intended data is quantitative more participants who are randomly chosen (giving equal selection opportunity to all individuals) will be required (as inference made could be generalised) but if it is a qualitative data or method then fewer than the quantitative number could be purposefully selected (Creswell, 2009) as the qualitative sample is not for generalisation but rather to gather in-depth understanding of the sample group.

4.4.1.2 Types of Mixed Method Sampling

There are two major sampling types identified for quantitative and qualitative researches:

- Purposeful Sampling (as in qualitative) – this is when the researcher intentionally recruit participants because they have knowledge or experience of the key concept being explored.
- Probability Sampling (as in quantitative) – This involves the random selection of participants who are representative of the population or who represents a segment of the population (Creswell & Plano Clark, 2011).

However, there are different classifications of mixed method sampling and five of these have been identified by Creswell, (2009) and Teddlie & Tashakkori (2009) as:

- Basic Sampling Strategy
- Sequential Sampling
- Parallel or Concurrent Sampling

- Multilevel Sampling
- Sampling using multiple sampling strategies

4.4.1.2.1 Basic Sampling

This sampling strategy involves the purposeful and probability or random sampling technique. It could also be seen as purposeful or purposive random sampling, where the researcher randomly selects a small number of units from a larger group. The random or stratified selection is the probability characteristic, while the small number selected is characteristic of purposeful or purposive sampling.

4.4.1.2.2 Sequential Sampling

This is when the sampling or sampling result from a phase informs the sampling for subsequent phase. It can also be a sampling technique where the sampling of a phase determines the sampling for subsequent phase. For example, a researcher randomly selects a number of participants and further uses them to purposefully make participant selection based on certain criteria for the subsequent study (e.g. the researcher selects 5 people and then asked them to each select 5 males who have experience in design, are undergraduates, and have a sibling, who will participate in a survey that intends to assess the life style of undergraduate design students who have siblings). The sequence can either go from qualitative to quantitative or vice versa based on the research interest or focus.

4.4.1.2.3 Parallel or Concurrent Sampling

In this technique, a probability or random technique is used to generate quantitative data while purposeful technique is used to generate qualitative data. In this instance the sampling techniques are done independent at the same time. It could also mean sample that is generated from the joint application of purposeful and probability technique.

4.4.1.2.4 Multilevel Sampling

This is when different sample (unit of analysis) are nested within one another. In this instance, researchers are interested in questions that relates to more levels or units of analysis.

4.4.1.2.5 Sampling using multiple sampling strategies

This is a sampling technique that involves the combination of any previously mentioned sampling strategy.

4.4.1.3 Sampling for this Research

The child-computer interaction (ChiCI) group in UCLAN have affiliation with some primary schools and since this research aims to *investigate whether children can perform an effective inspection method evaluation on technologies designed for them based on their values* and considering the age group of this research (see chapter 2). The decision was made to use school children within the research age group. Also, since the context of the research targets educational games for children, it was decided at one point to gather input from teachers being stake holders in the development of children's learning material.

To ensure the required data (i.e. to access how the HE works in practise and to assess children's performance in an HE study) is gathered, people who are knowledgeable in the heuristic evaluation method being explored were purposefully recruited as HE pilot study evaluators (see chapter 5). Also people who are knowledgeable in working with children and children's technology were also purposefully recruited as study observers.

Due to the nature and needs of the research and the expected outcome (the aim and objectives) multiple strategy sampling (see section 4.4.1.2.5) was adopted. As in some instances, a basic sampling strategy (see section 4.4.1.2.2) was used (e.g. studies reported in chapters 5 to 9) and in some other instances a sequential sampling strategy (section 4.4.1.2.3) was adopted (as in the studies reported in chapter 7). These sampling strategies were decided based on the nature and needs of the research at different stages. For example, the basic sampling was decided where children aged 7 to 11 were the participants selected, because the new method being designed is suited to this age group of children. The participants were chosen by their class teacher who called the names of children that will participate at a given time and at some point asked which child had not participated and sends them to participate. It could therefore be argued that the sampling technique used here is purposive (since the teacher specifically picked the children that should participate at a given time) and on the other hand it could also be argued to be random (since every child was given the opportunity to participate). The sequential sampling was chosen because the result from one studies always informed subsequent studies sampling. For example, the outcome of the HE study with children reported in chapter 5 informed the decision for using teachers as participants reported in chapter 6, as it is believed that they will be the most appropriate participants to give the needed answers for the next stage.

For the purpose of ethical reasons, convenience, and availability, it was decided that studies will be run during school times within the school premises. Since, it was difficult to recruit schools, only four schools were recruited for the entire research and these were schools which already have relationship with the ChiCI research group.

4.4.2 Data Type

In view of the research aim, two types of data were gathered from this research: Qualitative and Quantitative, this will be discussed in view of this research.

- Qualitative data: consists of open-ended information, information that is supplied based on participants' perception and view about the concept of study. This can either be verbal, written or pictorial (i.e. image) (Creswell & Plano Clark, 2007). Within this research the following qualitative data was gathered:
    - Usability problems reported by evaluators (child and adults) – Problem the encountered during game play that we prevent or obstruct the use of the game or cause dissatisfaction (chapters 5, 8 and 9)
    - Observation data on problems children encountered during the study, facilitation issues and any other issues that occurred during the studies (chapters 5, 8 and 9)
    - Teachers' views on how to improve the method for children that informed the design of the new method (chapters 6 and 7)
    - Children's views on what will make a good game (chapters 7 to 9)
    - Children's drawings of severity scale (chapter 7)
    - Researcher's note gathered during studies to explain any occurrence during the study (Chapters 8 and 9).
- Quantitative Data: are closed ended information usually found on altitude or behavioural instruments (Creswell & Plano Clark, 2007). In other words, they are responses to close ended questions. Within this research the quantitative data retrieved are:
    - Responses for the severity rating scales (chapters 5, 8 and 9)
    - Children age and classes (all studies)
    - Heuristic numbers (chapter 5) which is attached to heuristics that describes the kind of problems found during evaluation.

4.4.3 Reliability

Reliability is the ability for responses or result to stay consistent and stable over time (Creswell & Plano Clark, 2011). It is also viewed as the researcher's approach being consistent across different researcher and project (Creswell, 2009). In HCI it is viewed as the ability for an experiment (study) to be replicable by other researchers in other locations and can still yield consistent, dependable and stable result (Lazar, Feng, & Hochheiser, 2010). In the context of this work the focus of reliability is in the following:

- Methods used: to carry out literature on methods chosen and ensure the procedure is well understood and applied correctly.
- Data Analysis: to ensure that data analysis is carried out with the appropriate method (s) and also review and pick an appropriate strategy or strategies that could improve the reliability of the result.
- Newly Designed Method: to ensure that the new method (IMCH) produced is designed following standard measures as documented in literature for designing evaluation methods for children and inspection methods in order to produce a reliable method. This will also include adapting a method for assessing UEMs to improve reliability as reviewed and reported in section 3.5.

In literature, several methods for improving reliability is identified in literature as it relates to qualitative and quantitative data:

Qualitative: The reliability of qualitative data is more relating to the reliability of multiple coders to reach on codes and themes applied during analysis (Creswell & Plano Clark, 2011). To ensure reliability, Krippendorff (1980, p. 130-132) suggests three strategies such as:

- **Stability** (when a coder codes data in the same way over time and gets the same result)
- **Reproducibility** also referred to as **inter-rater or inter-coder** reliability (where multiple coders code data and get the same result)
- **Accuracy** (when panel of experts assess the codes against a predetermined standard or when codes are known from previous studies).

Quantitative: Statistical measures could be used to check for reliability coefficient, internal consistency and test-retest comparison could be carried out while exploring data (Creswell & Plano Clark, 2011). It is suggested that the reliability of result be determined before the assessment of their validity (Creswell & Plano Clark, 2011).

4.4.4 Validity

In Quantitative context, validity means that the scores or responses received from the participants are meaningful indicators of the construct being measured, this is measured against standards drawn from external sources like statistical procedures or external experts. While qualitatively it is viewed as the accuracy, trustworthiness and credibility of the researcher's and participants' account, or result produced. In other words, checking for validity means assessing whether information obtained through the qualitative data collection is accurate.

Quantitative validity could be measure with the following evidence (Creswell & Plano Clark, 2011):

Content Validity – This is to assess whether the items or questions used are representative of possible items.

Criterion-Related Validity – Whether the findings from the study relates to some external standards e.g. scores on similar instruments.

Construct Validity – Whether it measures what it intends to measure.

There is also the conclusion validity which is either internal or external:

Internal Validity – Is the extent to which the researcher can conclude that there is a cause and effect relationship among variables.

External Validity – External validity is when conclusion can be made that the result derived could apply to a larger population.

Qualitative validity could be determined by the following (Creswell & Plano Clark, 2011; Creswell, 2009):

Member-Check Validity – The researcher takes the final result or summaries or themes back to the participants to inquire of them whether the themes are an accurate reflection of their input.

Triangulation: When data from multiple sources is used to confirm findings previously retrieved. Taking data from transcripts and pictures or from multiple sources. Usually this is used to build justification for themes or codes derived.

Peer Debriefing: This involves using an experienced person who in the qualitative study who will review and ask questions about the study to resonate the account from the view point of other researchers other than the actual researcher.

Other methods include: Rich and Thick descriptions, reporting disconfirming evidence, Use of external auditor, spending prolonged time in the field.

4.4.5 Ethics

Although it is the goal of the research to derive credible result and get answers to research questions. However, it is important to ensure the well-being of participants (Teddlie & Tashakkori, 2009). According to Clough & Nutbrown (2012) researchers must obtain ethical approval from the institutions to ensure it meets minimal required ethical standards. Such standards could include:

- Provision of the best possible protection for researchers and their participants
- Ensuring data are collected with informed consent of participants
- Protection of participants details, well-being and identities

The ethical committee might differ from one institution to another, however they pursue a central purpose of ensuring research is done with utmost safety for the participants and that the research is done with morals. In addition, they uphold the national ethics framework that involves data protection, human right, the freedom of information act, only to mention a few. In UCLAN the body in charge of research ethics are called the research ethics committee. They enforce ethical measures to ensure researches carried out have been ethical considered. Within the ChiCI research group, the checktool (Read et al., 2013) have been developed to ethically guide members of the group on participatory design research carried out with children. The following are ethical measures that have been taken for this research.

In view of the primary participants (children) for this research, the researcher and everyone who worked at any point with children as part of the research obtained a DBS (Disclosure and Barring Service, formally called CRB (Criminal Records Bureau)) clearance via the university ethics committee in fulfilment to UK laws for working with children. In addition, to ensure right practice in accordance with the university's ethics and data protection act obligation, all studies and activities for this research was submitted as pro-forma to the University Ethics Committee who reviewed, cleared and approved the research.

Prior to children's participation in any study, full information about the study was provided to the head of participating school to obtain consent from the school and also from parents or guardian of participating children via the school. Since all the children had consent to participate, every child was given equal opportunity to participate in each study avoiding the tendency of depriving any child; that could have caused inferiority complex for the deprived child. In the case of adult participants, information (that covers full content of the study and forms to obtain content) was sent to the adults prior to the study allowing them time to prepare and give their consent. Participation was always voluntary so participants were always verbally informed of their freedom to withdraw from the study before and during the study.

No video recording or personal data were collected. In the cases where photographs were essentially needed, consent was obtained from responsible persons of the participants (e.g. school, teachers). Though in such cases participants were photographed face off (i.e. in positions where participants' faces were not captured) and participants who had no photography consent or who were not happy with being photographed were not covered. Data collected during the study was kept on the university's network and hard copies produced were kept in a locked room within the university's premises or on authorised person at all times deterring access to unauthorised persons. When result is obtained, conclusion is drawn and data is no longer needful, it is destroyed via the university's data disposal scheme.

During studies the needs and comfort of the participants and the school was always considered first before the needs of the research. Safety measures were always taken for each study environment and participants were always informed of safety facilities. Also participants were allowed to use toilet facilities and have refreshment during studies.

4.5 Methods for the Research Stages

4.5.1 Literature Review

Prior to stage one and throughout the study, analysis of literature was carried out which focused on HCI, CCI, Evaluation Methods, and Children. Most literature that pertains to HCI, CCI and Evaluation Methods were retrieved from HCI and CCI, published conference papers, journals, and textbooks, in some cases HCI thesis were used as informed guide for further reading. Research that pertains to children were retrieved from sociology, psychology and education journals, and textbooks and some were also from the CCI literatures. Resources which were electronically retrieved were collected using key terms on academic search engines, platforms and repository such as:

- google scholar,
- the ACM digital library,
- Science Direct,
- EBSCO,
- Springer, (only to mention a few)

4.5.2 Stage one – Identifying a gap in UEM with children

In this stage literature review was carried out following the method described above (section 4.5.1). This review was focused on HCI, evaluation methods and evaluation methods for and with children. With this review of literature it was determined that there is a gap in inspection method with children, so the decision was made to gain access into children's ability to perform a usability evaluation to assess their performance in a chosen UIM

4.5.3 Stage two – Assessing Children’s Performance in a UIM Evaluation

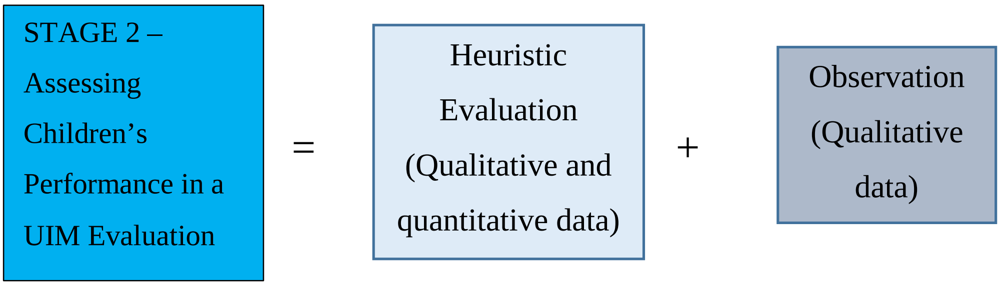

Figure 4.2 Stage 2 - Methods used and Type of Data Collected

The UIM chosen from stage one was the heuristic evaluation (HE) method, this method was decided because it is shown in literature as the widely used inspection method with adults (Woolrych and Cockton, 2002) and it is the only inspection method that has been tried out with children. The method involves the recruitment of evaluators to inspect an application based on set guidelines referred to as heuristics, to find usability problems (see 4.4.2) and rate the problems using a severity scale (see section 3.2.1.3 for full detail of the HE method). In order to assess the performance of evaluator and judge the study success, an observation technique was used to collect observational data (also see 4.4.2). This observation method is explained below:

4.5.3.1 Observation Method

Observation is a typical characteristic of ethnography qualitative research (Gobo, 2011) but has now been adopted as an established method within HCI (Rode *et al.*, 2012). The observational method in HCI is popularly used in user based studies (Barendregt *et al.*, 2007) types of usability evaluation method e.g. user testing, co-discovery, think aloud etc. In this categories, the method have been used successfully to collect data from users on their experiences and interactions with given applications (Donker & Reitsma, 2004; Hanna, Neapolitan, & Risden, 2004; Johnstone, Bottsford-Miller, & Thompson, 2006). It has also been used by MacFarlane et al., (2005) and Sim et al., (2006) to observe children interacting with three educational applications, recording signs of engagement and any usability problems encountered. Within this research, it is used as a technique to capture any problem participants might encounter during the study as regards the method procedure, tools being used for evaluation and facilitation instruction.

Data Analysis technique used in this stage is described in the study (chapter 5)

4.5.4 Stage 3 – Measures to improve UIM to work with children

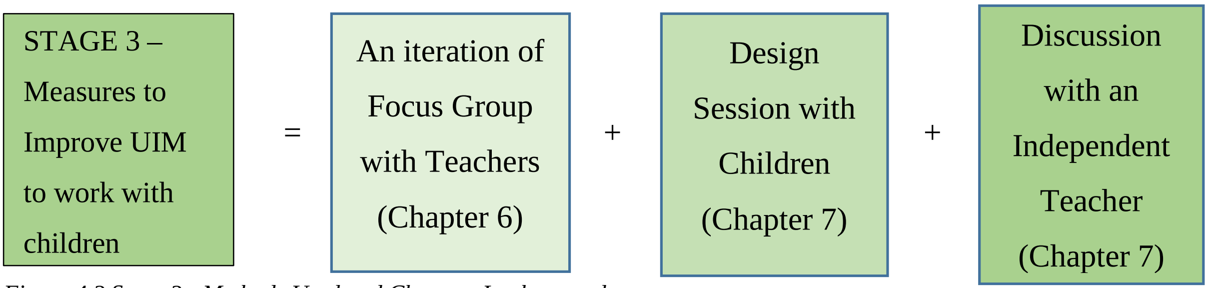

Figure 4.3 Stage 3 - Methods Used and Chapters Implemented

In this stage some stakeholders (teachers and children) in children activities and learning were included to modify HE method to suit children. A sequential exploratory mixed method design was employed for this stage (this design was decided as it is believed to produce rich data needed for the aim of the stage and it is done in sequence because the subsequent step (study) is informed by the previous (study). Activities carried out in this stage is reported in chapters 6 and 7

4.5.4.1 Specific Methods applied in this stage

The stage includes an iteration of focus group carried out with school teachers, then design sessions with children and lastly a discussion with an independent teacher to discuss the newly designed method.

4.5.4.1.1 Focus Group

Focus group is a long and well researched area that have several definitions. However, two different definitions in relation to this work goes thus:

- A carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive and nonthreatening environment (Krueger, 1994).
- Patton (2002) defines it as an interview with a small group of people on a specific topic.

It was decided to use focus group as opposed to standard one on one interviewing technique because of the benefits (advantages) of the focus group as compared to interview. For example, the ability to gather information from more people at one time about a particular topic (see chapter 6 for more of the advantages of the focus group).

4.5.4.2 Design Session with Children

As reported in chapter 2, research in involving children in design within CCI has identified four roles (users, testers, informants and design partners) (Druin, 2002). These roles have been described as different, though it is believed each step has a bit of the role that historically came before them (see fig 4.2).

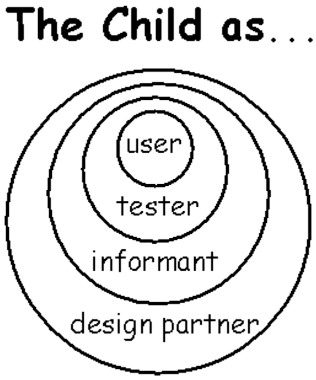

Figure 4.4 The Roles of Children in the Design of Technologies for Them **(Druin, 2002)**

Children as Users: This is reported as the oldest and most commonly used method for gathering design input from children (Druin, 2002; Obrist *et al.*, 2011). In this role, children use a particular technology in a natural setting (Druin, 1999b; Kelly et al., 2006), then adults seek to understand the impact of the technology on the children's experience via different methods e.g. observation, videotaping, think aloud. This is either done at the early, mid or end state of a product development (Druin, 1999b; Kelly et al., 2006) and it is usually carried out for two purposes: to assess general concept (s) that may help inform future technology developers and to gain better understanding of the process of learning which may contribute to future educational practices. The limitation of this role is that children are seen as object and their input towards the design of the technology is minimal (Druin, 2002).

Children as Testers: In this role the initial brainstorming and prototype of a product is already produced by adults. Children are then involved in testing the product and feedback is asked of them by adults via interviews, observations other testing techniques. The purpose of this the children playing the role of a tester is for them to help in shaping new technologies before they are released to the world. In this role children can have more immediate impact on the technology than in the role as users. However, there is the possibility that changes recommended by the children may not be implemented due to adult designers reluctance to implement it, time or fixed budget constraint (Druin, 2002; Kelly *et al*., 2006; Scaife *et al*., 1997).

Children as Informants: In this role, children are consulted as experts (Druin, 2002; Kelly et al., 2006), they inform the design process before any part of the technology is developed. With this role observation of the children is made with existing technologies or children could be asked to make input via paper sketches. When the technology is developed from children's input, it is taken back to children to assess, review and give feedback. Involving the children at different stages of the technology development continues and is done at any stage when the adult developer requires input from the children. This role is beneficiary as children can have an impact on the technology from the conception stage where ideas are gathered on the direction of the technology, they continue to make input on how it should be shaped and ultimately on how it should be evaluated. With this role, children make great ideas of which some might not be taken on board by the adult developer due to pedagogical or time constraint reason. Also this role is limited such that it is difficult for the adult to decide what ideas to say yes or no to (Druin, 2002; Emanuela Mazzone, 2008; Michael Scaife *et al*., 1997; Mike Scaife & Rogers, 1998).

Children as design partners: This role is similar to the informant design role, though in this role children are equal partners with the adult researcher where they make input throughout the entire design process. It is believed that children cannot do everything adults can do but they can be critical of new technologies, have their own special experiences and viewpoints they can contribute to the design process. Children's impact on the technology through this role is enormous and their voice can be heard and implemented. Also children's input as design partners can bring dramatic effect on the technology being developed (Druin, 2002; Kelly *et al*., 2006).

4.5.4.3 Designing the IMCH

As the design of the IMCH followed a quasi-mixed method approach, the design of the severity rating scale for and by the children within the method was designed following an informant design approach, where children were to make input by coming up with their own drawings and also they were given an example severity rating (traffic light) to test their understanding and get their input of such design. It is believed that with this design technique (informant design technique) one could get insight directly into children's thought process on what type of severity scale to design for them and also it will be a scale almost totally designed by children rather than by adults.

4.5.4.4 Testing the newly Designed Method

In order to test the IMCH method, an iterative study of the method was carried out with children. In this case, children were observed as they acted as evaluators of an application designed for them using this method and result from a study was used to design subsequent study and rerun by children, as a result several versions of the method was produced. This is fully reported in chapters 8 and 9 of this thesis.

4.5.5 Stage 4 – Write up Guideline and Finish

In this stage conclusion was drawn from the outcome of studies reported in chapters 8 and 9 to decide on what the method should be which is described in chapter 10. The method guideline is also provided in this chapter (10).

4.6 Conclusion

This chapter has been able to highlight and review available method suitable for this research and stated specifically methods and techniques used within this research. It was stated that this thesis is focused on UIM targeted to children in order to design a suitable UIM for children. Therefore, a UIM (HE) was explored with children which is reported in the next chapter (5). However, below is a structure of the remainder of the (study) chapters and what they contribute to the thesis.

4.6.1 Structure of Chapter Flow towards Creating the New (IMCH) Method

4.6.1.1 Chapter 5

This chapter reports the HE pilot study carried out with adult experts to understand how the HE method works in practise then an HE study was carried out with children to assess their performance in a UIM evaluation. In conclusion of this chapter, it was established that the HE method in its original state is not suitable for children, as there were language issues with tools used within the method and lack of understanding of the method process. The review of literature and via peer debriefing, it was decided that teachers be asked of solutions to these issues and need to adapt the method to make it more suitable which led to the focus group study reported in chapter 6.

4.6.1.2 Chapter 6

Chapter 6 is a report of an iteration of focus group studies carried out with school teachers on issues children encountered with the HE method reported in chapter 5 and how to adapt the HE method to make it more suitable for children. The outcome of this studies suggested the need to make the tools and method process more child friendly and one of the ways suggested is to take the design of the tools to the children and involve children in the redesign process of the method. Having these suggestions (this is well detailed in chapter 6), indulging in peer debriefing, reviewing literature on how children could be involved in design sessions and how evaluation methods for children are developed; the inspiration to include value and use narrative in the method process was considered. This inspired the need for the new method which led to the exploration teachers' ideas on how to design the new method and its tools reported in chapter 7.

4.6.1.3 Chapter 7

Chapter 7 is an exploration of teachers' ideas and ideas gathered from literature. In this chapter there was the review of literature on narrative/storytelling and values which ended up as part of the method producing the first version of the method. Design sessions were carried out with children (younger and older) to design some tools and part of the method exploring and confirming some of the teachers' ideas (from chapter 6), this session resulted in the formulation of IMCH V2. Also the chapter reported the scrutiny of the method version 2 by an independent teacher and the development of the 3rd version of the new method. The chapter ended with how the IMCH V3 works, which was first tested and assessed in chapter 8.

4.6.1.4 Chapter 8

The version 3 of the IMCH was tested to determine the suitability and effectiveness of the method for children in comparison to the HE method study carried out in chapter 5. The result from the study reported here showed children were able to come up with game criteria based on their values, find real usability problems, understand and use the traffic light severity scale and also merge their found problems and have thoughtful discussions on problems found. However, some issues were identified, for which modifications was made to the method producing IMCH V4 which was tested in chapter 9.

4.6.1.5 Chapter 9

The suitability and effectiveness of the Version 4 of the IMCH was tested in this chapter. Result from the study here confirmed some usability problems children reported in chapter 8, it also showed the extension of the severity scale was a useful process. Finally producing the working state of the method as at the conclusion of this thesis. This is reported in chapter 10.

4.6.1.6 Chapter 10

This chapter made an explanation of the IMCH V5 and stated the accompanying guidelines for the method.

4.6.2 Chapter Contribution to Thesis

This chapter shows method available for this research, data collected and the structure on how subsequent chapters contribute to the creation of the new method (IMCH).

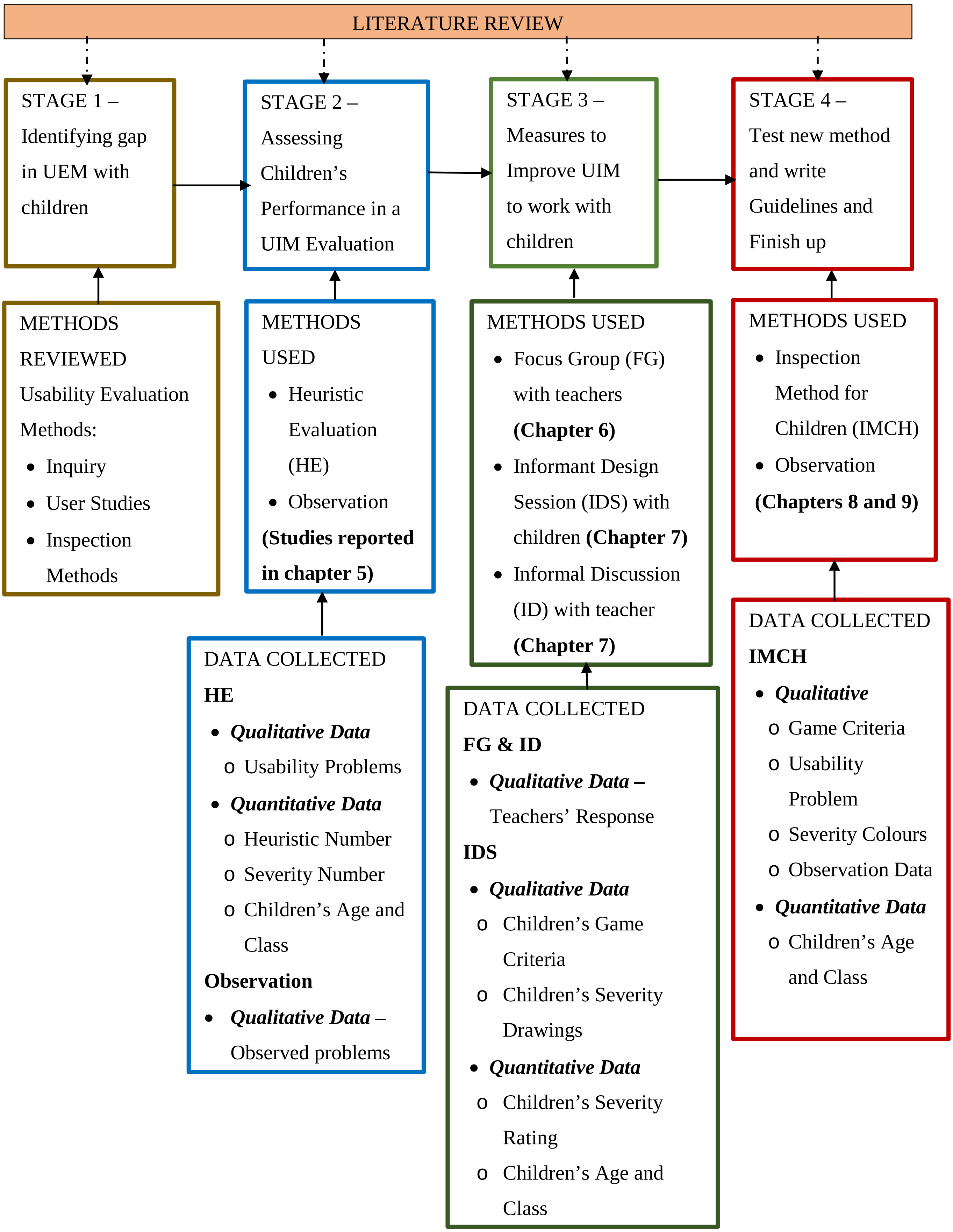

Figure 4.5 Methods Used at Different Stages of the Research and Data Collected

5 CHAPTER 5: PILOTING HEURISTIC EVALUATION METHOD

5.1 Introduction

As reported in section 3.4.1, heuristic evaluation (HE) method (Molich & Nielsen, 1990) originally designed for adults is the most commonly used inspection method with adults (Woolrych & Cockton, 2002). It has also been used with teenagers (MacFarlane & Pasiali, 2005; Wodike *et al.*, 2014) and children (Salian *et al.*, 2013) as the evaluators. The research in this thesis carried out a heuristic evaluation with children to answer RQ2 “How will children’s performance be assessed in a UIM evaluation?” stated in section 1.4.1.

The HE study carried out in this research is very similar to the study by Salian et al (2013) with the intention of confirming the reported problems children encountered in Salian et al’s study. Although, there were variance in both studies; for example, the heuristics set used in this study had more number of heuristics than those in Salian’s study (with the hypothesis that if more heuristics is provided, it will help children find more problems). Also, there were difference in the severity scales used in both studies. For example, Nielsen’s (1995) original severity scale was used in this study, to investigate whether children can understand the text description of each step of the scale meanwhile Salian et al’s (2013) study created and used a severity called “bad scale” with the presumption that children will not understand Nielsen’s scale. Though, there was no report to prove this assumption.

The HE study of this research aims to assess children’s performance in a UIM based method and investigate the ‘ease of use’ and effectiveness of the method for children.

‘Ease of Use’ in this study as stated in chapter 3, section 3.5.1 is the appropriateness of the tools and method process for the children. This is broken down further into the following question:

- Was the procedure easy for children to follow?
- Were the tools provided right for them i.e. were they able to understand instructions and tools provided?

Effectiveness is the extent to which the method was used as designed and able to do what it is proposed to do (find and rate real usability problems).

- Were the children able to find and rate usability problems while carrying out evaluation with the method? Were the usability problems real problems?
- Did the children follow the procedure of the method in the correct order?

Since the researcher has no practical knowledge but rather has only theoretical knowledge on how the heuristic evaluation method works, it was decided that a pilot study with adult experts who have understanding of the method will first be carried out.

As stated above, this chapter aims *to investigate children's performance in a HE study and ascertain the 'ease of use' and effectiveness of the original state of the method for children.* This will be achieved with the following objectives:

- To carry out a HE pilot study with adult experts to gather knowledge on how the method works in practise
- To carry out HE method study with children to assess children's performance
- To compare findings from the pilot study to the study with the children and further compare study to previous study carried out by another researcher.

The chapter is presented in four sections. Section 5.2 reports the pilot study with adult expert evaluators, section 5.3 reports the study carried out with children, section 5.4 discussed both studies and section 5.5 will conclude the chapter stating the way forward for the research based on the results and discussions.

5.2 Pilot Study 1 (Heuristics with adult experts)

The first pilot study was carried out with adult experts to understand the practicability of the HE method with experts. It is also intended to gather insights on how to run such study with children as it will be unethical to run the study with children with no prior knowledge on how the method works in practise. This is because while trying to explain the method, the researcher could use language and tasks that is inappropriate (too advance) for the children's grasp level. Also there could be measures that can bias the children's input or thoughts. These can become problematic as children may not understand the explanation given which could upset them, make them uncomfortable and lead to their unwillingness to participate or carry on with the study. Also unethical use of the method could lead to them producing biased results.

5.2.1 Method

Within this HE study, an observation method was adopted to observe facilitation issues that might arise during the entire study. It is believed that the adult experts (evaluators) know and understand the method process. So if the researcher is unable to provide instructions that they can understand then it will be an issue to explain the method process to the children. Observers are also required to capture issues evaluators might encounter understanding or using the documents and proformas provided for the evaluation, so it could be addressed before carrying out the study with them (children). Finally, to capture possible issues that already exist in literature amongst HE evaluators.

5.2.1.1 Participants

Researchers from the ChiCI group participated for this study. A total of 7 researchers where 4 of these participants are experienced in performing heuristic evaluation, working with children and evaluating children's technologies. Three of the experienced researchers acted as the expert evaluators for the study (since 3 is a recommended number for a HE study (Nielsen, 1993a) and this was only a pilot study) and the fourth experienced researcher acted as the study advisor who gave guidance and recommendations on the appropriate things to do before the study.

The other three out of the 7 ChiCI members were PhD researchers who are experienced in working with children and with children's technology. One of the three is the lead researcher who played the role of the study facilitator while the other 2 acted as the key observers for the study to capture observational data as stated in 5.2.1 above. Though the observers didn't have prior knowledge in making observation especially in a study like this, but with the help of the study advisor, they were given some hours of training on hour to do such observation.

5.2.1.2 Apparatus

Several tools were used during the study which include: the application that would be evaluated (see subsection 5.2.1.2.1), the heuristic set (the guideline) with which evaluators will judge issues found to be problems or not, the severity rating scale used to rate the severity of problems found and various forms were also used.

5.2.1.2.1 Application

To carry out the study, each expert evaluator was provided with an iPad which had the 'Fingu' (a counting) game pre-installed (see figures 5.1). This game was chosen because it was developed by CCI researchers, first presented in IDC 2012 (Barendregt et al., 2012), appropriate for the chosen age target of this research and was played by the researcher and judged to have some usability issues. All evaluators interacted with the iPad via touch as that was the only interactive technique allowed for the study.

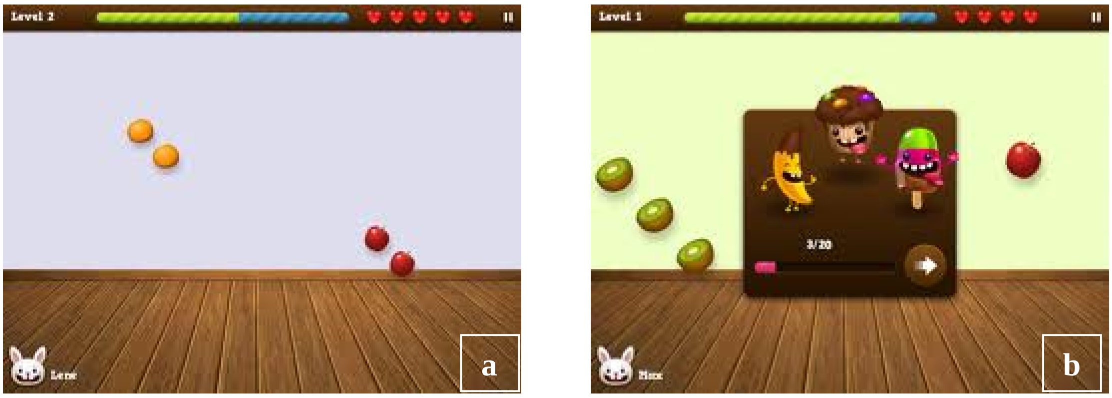

Figure 5.1 Images of iPad Fingu game (a) Actual game play (b) A successful play

5.2.1.2.2 Heuristic Set and Severity Scale

There is no standard heuristic set for evaluating educational games for children. Therefore, Korhonen & Koivisto's (2006) 29 playability heuristic was used (see appendix 1A), because it is originally designed to evaluate games on mobile device (of which the technology used here is a mobile device). It is generic enough to evaluate key aspects of the chosen game given the sections this heuristic set is divided into: mobility, game play and game usability. Though the heuristics here seem too many considering the number in the original heuristics (Molich & Nielsen, 1990; Nielsen, 1993a). However, it is difficult to cut it down without risking the possibility of leaving out important aspect of the heuristics necessary for the evaluation.

Also to help the evaluators rate the severity of each problem found, Nielsen's (1995) severity scale that rates from 0 to 4, where

0 = I don't agree that this is a usability problem at all
1 = Cosmetic problem only: need not be fixed unless extra time is available on project
2 = Minor usability problem: fixing this should be given low priority
3 = Major usability problem: important to fix, so should be given high priority
4 = Usability catastrophe: imperative to fix this before product can be released

In order to ensure evaluators' convenience and ease in the use of tools provided, it was decided that the heuristic set and severity scale should be on a single sheet (see appendix 1).

5.2.1.2.3 Forms

The study had two phases, in phase 1 evaluators played the game, individually found problems and recorded them. To ensure this is done, an individual problem record sheet called the Evaluators' Data Collection Form (EDCF) (see appendix 1) was provided for each participant. This form was decided because it was originally designed by researchers in the ChiCI group and used in a heuristic evaluation study and judged to have been suitable for gathering necessary problem areas in a HE study. In the second phase, the evaluators consolidated their list of problems into an aggregated list in a single sheet or more. Therefore, a problem merging form (see appendix 1) was provided for the evaluators.

Finally, in order for the observers to capture issues that arose from the facilitator during the study and issues encountered by the evaluators, they were provided with the two observers' data collection form (see appendix 1D), which had pre-coded themes to help observers classify problems they would captured. One form was to capture issues during the individual evaluation phase and the second was to record issues that occurred during the merging phase. These themes were decided based on the areas of the evaluation, for example the observer's form for the individual phase had the following aspects needed for the phase: Evaluation Instruction (EI), Heuristic Set (HS), Severity Rating (SR), Data Collection Form (DCF), Task, and others (to record issues that do not relate to the other themes). For the second (problem merging) phase, the following are important: Problem Categorisation Instruction (PCatI), Data Collection Form 2 (DCF2) (i.e. the form in which the evaluators will record their merged problems), the severity scale (this is to indicate whether evaluators had issues agreeing on the severity rates of merged problems) and Others (to record issues that do not relate to the other themes)

5.2.1.3 Procedure

The study was carried out in a lab within the university. The room was set up with three tables in a "U" form to separate the participants from one another, in order to reduce bias. The participants individually arrived at the lab and were directed to sit at a table, each participant was made to sit on their own.

The study advisor (fourth experience researcher) started the study with the following explanation: "you have been invited as usability experts to participate in this heuristic evaluation study and it should take 30minutes of your time. I will leave the facilitator to explain the details of what you are expected to do".

The facilitator gave the following explanation “This evaluation will be in two phases. In the first phase, you will individually evaluate the Fingu game, find problems using the heuristic set, write any individually found problems in your EDCF and attach a severity number to it using the severity rating scale. In the second phase you will come together and consolidate all your individual problems into a single problem sheet and attach an agreed severity number. This will be further explained at the end of the first phase”. Participants were further given verbal instruction: “start and play game to start the first phase”. The introduction and task distribution session lasted 5 minutes while the individual problem finding session lasted for 10minutes.

In the second phase, tables were re-arranged, so evaluators could sit close to each other to do task. The facilitator verbally described the task: “Go through your problems and merge them into a single list. Attach an agreed severity and frequency of the evaluators who found the problems. One person needs to act as the scribe to write down the merged problems”. This phase lasted for 10minutes and the entire evaluation lasted for 25minutes.

5.2.1.4 Data Analysis

As stated previously, the number of evaluators for this study was only 3 which resulted in a group for the problem merging phase. At the end of the evaluation, the evaluators merged all their problems into a single sheet producing a total of 10 problems. Since the problems were by only a group with few number of problems, the researcher carried out analysis on the data by reading through evaluators’ individual problems to determine what area their problems occurred and how many problems each evaluator found and reported. The researcher also read through the merged problem list to investigate whether similarities exist in merged problems reported that could be further merged and categorised into a theme, to aid easy report of results. Four out of the ten merged problems were similar while 6 were unique; therefore, the similar ones were merged into a theme “Unclear Instruction” and the others were also given a theme. Themes emerged from the problems see table 5.1.

The observers’ problems captured were also few, therefore they were put together into a sheet read through to report the findings.

5.2.2 Results

The resulted will be presented in two sections as Evaluators’ Data and Observers’ Data.

5.2.2.1 Evaluators’ Data

Evaluators’ data result will be reported for the individual evaluators and for the merged data.

1. Individual Evaluator' Data

All the evaluators found, reported and rated a total of 16 problems where 2 evaluators each found and reported 5 problems and the third evaluator reported 6 problems, see table 5.1 and appendix 1E for full detail of individual problems. However, summary from the individual problems show that there is an evaluator effect on the usability problems found where most of the problems found were found by one or another and not by all evaluators (see chapter 3, section 3.4.1 for the definition of the evaluator effect).

Table 5.1 Number of Individual problems found

Evaluators	**A**	**B**	**C**	**Total**
Problems found	5	5	6	16

Table 5.2 List of problems and evaluators who reported such problems

S/No	**PROBLEM MERGED**	**Evaluators**			**Total**
		A	**B**	**C**	
1.	No instructions or idea how to play the game	*	*	*	3
2.	Hard to tell how to start the game		*		1
3.	Holding the fingers down unclear		*		1
4.	Unclear starts page			*	1
5.	Lack of feedback/bad feedback	*	*	*	3
6.	Balance in game play			*	1
7.	Level two is the same	*	-	-	2
8.	Time bar disappeared		*		1
9.	Screen did not rotate	*			1
10	Fruit position made touching all fruits at once difficult			*	1

2. Merged Data

A total of ten usability problems were identified for all the evaluators but after analysis (further merging) of the evaluators' merged data, it resulted in a total of 7 merged problems, see table 5.3 below.

Table 5.3 Merged usability problem of adult expert evaluators

S/No	MERGED PROBLEM THEME	PROBLEM MERGED (BEFORE FUTHER MERGING)	FREQUENCY	SEVERITY RATING
1.	Unclear Instruction	**1.** No instructions or idea how to play the game	3	4
		2. Hard to tell how to start the game	1	2
		3. Holding the fingers down unclear	1	3
		4. Unclear starts page	1	2
2.	Bad feedback	**5.** Lack of feedback/bad feedback	3	3
3.	Unbalanced gameplay	**6.** Balance in game play	1	3
4.	Similarity in Levels	**7.** Level two is the same	2	3
5.	Invisible time bar	**8.** Time bar disappeared	1	3
6.	Non Rotate Screen	**9.** Screen did not rotate	1	2
7.	Difficult touch of multiple fruit	**10.** Fruit position made touching all fruits at once difficult	1	2

From the table 5.3 above, two of these problems (no instructions or idea how to play the game and Lack of feedback/bad feedback) were found by all 3 evaluators who rated the severity of these problems as **4 (**Usability catastrophe: imperative to fix this before product can be released) **and 3 (**Major usability problem: important to fix, so should be given high priority) respectively. Also one problem was found by only two evaluators who have rated the severity as 3 and the other seven problems were each found by just one evaluator.

5.2.2.2 Observers' Data

Observers were required to capture issues evaluators might encounter during the evaluation (see 4.2.1.2 for full observational data to be collected). In view of this observers' note gathered is as follows:

Table 5.4 Observers' data collected for individual expert evaluation

FIRST OBSERVER'S (First Phase) Problem List		
Problem List	**Evaluator**	**Related Problem Area**
Spent longer time going through Heuristic Set before game play	A	Heuristic Set
Had to reduce sound to concentrate on filling form	C	Others
Start stage didn't make sense	C	Task
SECOND OBSERVER'S (First Phase) Problem List		
Problem List	**Evaluator**	**Related Problem Area**
Taking a lot of time to finish a stage due to participant inability to simultaneously touch the items on the screen	B	Others

Table 5.4 above showed observers reported a total of 4 problems for all evaluators in the first evaluation phase with 2 problems reported for evaluator C. The other 2 for evaluator A and B (one for each).

Table 5.5 Observer's data for the problem merging phase

First OBSERVER'S Second Phase Problem List	
Problem Category	**Related Problem Area**
Scoring the feedback	Severity Rating
A problem wasn't written but was found	DCF2
Touch Food/ You don't have to or touch anything	PCatI

Table 5.5 show evaluators encountered three problems during the problem merging phase. One problem is the issue of not recording problem found. Another problem was reported bordering on severity rating (scoring the feedback) it is believed the evaluators had issues rating the severity of the problem "Lack of feedback/bad feedback". Also they had initial issues with agreeing on what should become of the problem "Fruit position made touching all fruits at once difficult" which the observer had classified as a PCatI problem, see table 5.5. This could be described as evaluators effect, a fundamental problem in usability evaluations and also in

expert based evaluations (Capra, 2006; Hornbæk & Frøkjær, 2008; N. E. Jacobsen *et al*., 1998; Ling & Salvendy, 2009).

5.2.3 Discussion

Results show evaluators were able to carry out heuristic evaluation and found usability problems on the specified technology (see Figure 5.1 above). Report from the observers did not show that evaluators had issues with the method process or instructions from the facilitator, during the individual evaluation phase. However, it was observed that evaluators encountered some problems commonly reported in usability evaluation literature e.g. evaluators' effect. 'The inability of the evaluators to reach an agreement on what is judged as a problem or on what is the severity of a problem'. This is identified as a threat to the reliability of UEMs (Hertzum & Jacobsen, 2001; Jacobsen *et al*., 1998).

Also that there was omission of problem found, i.e. an evaluator had found a problem but failed to record it. This is confirmable from table 5.2 with the problem "level 2 is the same" (this is highlighted in red in the table) where it was investigated from the individual data that only one (evaluator A) of the 3 evaluators found that problem, however, evaluators had indicated that 2 people found this problem. This will be curtailed in the study with the children as they will be informed and re-informed that only problems recorded, can be merged.

It was also observed from the observers' data that the first observer deviated from what should be captured, that is, usability problems were being captured other than problems that arose from the study as stated in section 5.2.1 above. In view of this it is arguable that other problems would have occurred that wasn't captured by the observers. Therefore going forward more experienced observers will be recruited.

5.2.4 Conclusion

The expert heuristic evaluation was carried out to have an overview of how the method works and also to know possible problems with the method in practise. It is evident from the result that an example of such problem is the evaluators' effect. A known limitation of the usability method evaluation carried out by multiple evaluators.

Also that there is a tendency that evaluators could forget to record problems found as reported in 5.2.3.2, table 5.5. Therefore it is noted for the subsequent study that children need to be encouraged to record problems found.

Though no issue about understanding language was recorded for the evaluators, it was pointed by an evaluator after the study that it would be useful to break down some terminologies used in the study. Example of such are: Frequency used in the merging phase, severity and scribe used in the explanation of the person to write down the merged problems.

It is also important that observers for subsequent studies are trained appropriately such that -- they capture only data that are useful for modifying the method and not deviating.

This research provided means to curtail the last two discussed problem which is reported in the next section: Heuristic Evaluation with Children.

5.3 Heuristic Evaluation with children

Having gained a practical knowledge from the HE pilot study carried out with adult experts noting some issues that occurred from that study, a HE study was carried out with children *to assess their performance and ascertain the suitability and effectiveness of the method with children*. Where suitability will be determined by children's ability to understand the instructional language, and language used in the documents (tools e.g. Heuristic set and severity scale) provided and effectiveness to be determined by their ability to use the method in the correct order and are able to find usability problems with it.
In order to manage some issues that arose (e.g. merging problem that wasn't initially found and forgetting to write down problems found) from previous study with adult experts as discussed in section 5.2.3, it was decided that children will be reminded at intervals to record problems found. Also observers for the subsequent study were chosen early and informed via a document of the actual issues to observe and record. This was provided to them days before the study and on the day of the study this was discussed with the observers.

5.3.1 Method

Method used for this study is similar to that of the previous study reported in section 5.2.1 where heuristic evaluation was the method used; within which an observation method was used to capture issues children will encounter while performing the evaluation.

5.3.1.1 Participants

In total fifteen children from a UK primary school acted as the evaluators for this study. Children were called out of the classroom by their teacher and directed by a researcher to the study area in groups of four, although the last group had only three children.

Three researchers from the ChiCI group, were also present for the entire evaluation process. The author acted as the facilitator of the study and the other two researchers played the role of the observers for the entire study. All the researchers are knowledgeable in heuristic evaluation and have experience working with children and collecting observational data.

5.3.1.2 Apparatus

In order to ensure consistency and in view of the study aim following knowledge gained and lessons learned from pilot study with adult experts, the same apparatus (see 5.2.1.2 above) used in the HE pilot study with adults, was also used in this HE study with children.

5.3.1.3 Procedure

The study was done in a quiet activity room within the school premises and one of the observers called out the evaluators from their classrooms in groups.

Prior to the first phase, facilitator and observers introduced themselves to the children and the facilitator made an explanation of: the HE method, study procedure and evaluators' roles. Facilitator also mentioned the game to be evaluated and gave children a verbal task (start and play game). This introductory explanation lasted for 2 minutes

In the first phase, evaluators played and individually evaluated the Fingu (same as fig 5.1). They wrote down predicted usability problems into the data collection form attaching the heuristic number the predicted problem violates and a severity number (see appendix 1B). Afterwards, the observers retrieved the application while the facilitator thanked the evaluators and explained the second phase. This session lasted for 10minutes

In the second phase, evaluators consolidated individual problem list into a single merged list attaching an agreed severity rating and frequency (number of evaluators who predicted the problem) to each problem. In order to carry out the merging process in an organised manner, the facilitator collected a problem merging form, sat at a table and invited the evaluators to sit round the table with their written problem list and heuristic set/severity scale sheet, then the merging process started.

To merge the problems found, facilitator asked evaluators to read their problems one after the other, after an evaluator read a problem, the others went through their list to see if they had the same or similar problem. When they judged a problem as similar, they read it out to the group who decided if it is the same and should be merged. When agreed as same and merged, the evaluator who read the problem and others with similar problems ticked it off in their sheets to avoid reading the problem again then the facilitator wrote it down. Afterwards, a severity rating from the severity ratings attached to the individual problems was agreed by the group and the facilitator also wrote it down together with the frequency into the problem merging form. If problems are not agreed as same, it was treated as separate. This process continued until all problems were merged then the facilitator thanked the evaluators and collected the sheets. This merging process is a little different from the pilot study merging process with adults, as the facilitator stood in as the scribe to prevent too much writing for the children as suggested by Read (2007) and this was also the technique used by Salian *et al* (2013) who reported this worked well for the merging process. This entire study lasted for 20 minutes at the maximum. Although 20 minutes seem a short time to carry out the evaluation, as children could either play the game in a hurry, and miss out some problems that exist or they will forget to write down problems encountered. However, in order to carry out the study for all groups within the school allocated time and also to ensure children are engaged throughout the evaluation process, this was the time used. Also, literature suggests that session with children should be kept short (Horton, 2012).

5.3.1.4 Data Analysis

Due to availability, convenience and in order to ensure analysis is done thoroughly with no mix up of data, the analysis was performed over two days from the original study by the researcher and a second researcher (who is more experienced in analysing data gathered from children). The evaluators' data was analysed on the first day then the observers' data on the second day. This process is further explained in two sections as Evaluators' data and Observers' data.

5.3.1.4.1 Evaluators' Data

The evaluators' data was analysed in two stages. The first stage was to determine whether the children understood the meaning of the heuristic. In order to do this, the problems found were cross-checked to the meaning of the heuristic attached to ensure appropriateness.

In the second stage, the open card sort technique was used to analyse the data because it is an established method for analysing qualitative data in HCI and CCI (Read *et al*., 2013; Salian *et al*., 2013; Sim *et al*., 2013). Problems were cut into slips where each slip had just one problem and all the problems were placed on a table for easy move around. Since child evaluators consolidated their individual usability problems to produce a categorized list of problems, the analysis was carried out on the merged data. Therefore, the individual data was only used for reference in cases of unclear or vague statements.

Each usability problem was read by both analysts, to identify and agree on an appropriate main theme each problem should be. The analysts' agreement of a theme was reached by discussing and deciding that a more general sentence or phrase from a particular problem be captured as the main theme of that problem and similar problems (see example in next paragraph).

Upon agreement of a theme, it was recorded by one of the analysts and in cases where more than a single problem was agreed as the same, it was classified under the same main theme. For example, problems such as "it doesn't have instructions", "Don't know how to play the game", "I don't know what to do" and similar problems were classified under the main theme "Lack of instructions on how to play the game". Afterwards, the themes were read through and when themes were judged as similar they were further merged. At the end of the analyses a total of two main themes emerged from the problems: "Lack of instruction on how to play the game" and "Game froze".

5.3.1.4.2 Observers' Data

The same researchers carried out the analysis of the observers' data also in two stages using the open and close card sort technique and thematic analysis method. Observed problems for the first (individual) phase of the children's HE were analysed before that of the second (merging) phase and the same analysis process was used for both.

During the analysis, observed problems with pre-coded themes were placed on a single piece of paper. First, problems with the same theme were put together and counted. In cases where a problem was categorized under two pre-coded themes, e.g. the problem "No severity rating attached to form" observed for evaluator 'A' was classified as a EDCF & Severity Rating (SR) by a single observer, the analysts created a merged pre-coded theme EDCF/SR and reclassified the problem because they judged it as relating to two concepts and fit to be in a merged pre-coded theme than in one or the other. In cases where problems were similar within a pre-coded theme, they were merged thus eliminating repetitions. For example, the problems "Unsure of severity rating" and "Did not get the severity rating order" in the 'Severity Rating (SR)' pre-coded theme were merged to produce the problem "Did not understand the severity rating" and left in the same theme.

If a problem which relates to a single concept appear between pre-coded themes, i.e. one observer classified it in one pre-coded theme and the other observer puts it in another theme, a decision was made on its most appropriate pre-coded theme and it was eliminated from the other. For example, the problem "Needed assistance with the game play" was classified as a 'task' problem by one observer and the second observer reported it as "Need help playing game" and classified as an 'other' problem. This was agreed by the analysts as relating more to the 'task' pre-coded theme than as an 'other' problem, therefore was classified as such. Finally sub themes were created from the actual problems in each pre-coded theme and more general themes were created from the sub themes to re-categorize sub-themes (see appendix 1).

5.3.2 Results

The result will be presented for the evaluators and for the observers.

5.3.2.1 Result from Evaluators' Data Analysis

In total there were only 2 unique problems identified (1. Lack of instruction on how to play the game, 2. Game froze) after further scrutiny and merging of similar problems within the 6 initially identified merged problems, see table 5.6 and 5.7. 14 of the child evaluators predicted the first problem and 1 from the 14 evaluators predicted the second; however, 1 child (the 15th evaluator) found 0 problem.

Table 5.6 Number of Problems Reported by the Children

Problems	**Groups**															**Total**
	1				**2**			**3**				**4**				
	A	**B**	**C**	**D**	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**	
Individual	1	1	1	1	2	1	0	1	1	2	1	1	1	1	1	16
Merged	1				2			2				1				6

Two children did not attach a heuristic number to the problem they found. In addition, some children did not understand the heuristic set, for example the heuristic attached to some problems were not in context with the problems found. For example, one of the evaluators predicted a problem "Hard to use (maybe put instruction)" and attached it to the heuristic which says "The player does not lose any hard-won possessions".

Table 5.7 Summary of Evaluators Merged Problems Reported

Evaluators Identified Merged Problems	Final problem themes	Frequency of Evaluators who found problems
It doesn't have any instruction	Lack of instruction on how to play game	14
Don't know how to play the game		
Hard to use (Maybe put instructions)		
I don’t know what to do		
It doesn't show you how to do the game		
Game froze at level two	Game froze	1

5.3.2.2 Result from Observers’ Data Analysis

The observers’ notes reported 46 issues encountered for the first phase and 11 issues for the merging phase of the HE process. After analysis 10 general themes (7 in the first and 3 in the second but two general themes repeated in the second phase so it was merged to that of the first, see table 5.8) and 22 sub themes were identified for both phases (18 for the first phase and 4 for the second phase, see table 5.8).

Table 5.8 Number of Observed Problems: General themes, Sub themes and Actual Problems in Phases

Themes	Phases			
	First (individual)		Second (Merging)	
	Sub Themes	Actual Problems	Sub Themes	Actual Problems
Forms	6	16	0	0
Children	1	1	0	0
Understanding	5	8	1	3
Not Recording Problem	2	7	0	0
Gameplay	2	12	0	0
Missing Information	1	1	0	0
Communication	0	0	2	5
Bias	1	1	1	3
Total	18	46	4	11

Result also show some sub themes were repeated in a general theme because observers have original classified them into different predefined themes, see table 5.9 and full detail in appendix 1F.

Table 5.9 Example of Repeated Sub Theme in a General Theme

S/NO	Main themes identified from sub themes	Predefined Themes (categories)	Sub themes identified within predefined themes	Actual individual problem in problem identified within predefined themes
5.	Not recording problem	Other	Evaluator encountered problem but did not write down (3)	1. Not writing down problem encountered 2. Prompted to write the problem down 3. Ignore issues such as clock and blank screen
		Task	Evaluator encountered problem but did not write down (4)	1. Found problem not put on form 2. Problem not written down 3. Turned off did not write it down as a problem 4. Need prompting to write problems down

The table shows the actual problem reported by the observers from which the sub themes were identified. In addition, the main theme emerged from the sub theme while the predefined themes were the pre-coded themes that guarded the observers' classification of the problems observed during the evaluation.

5.3.3 Discussion

Though children were able to perform a heuristic evaluation but only a small number of problems were identified. This may have been because the game was fully functional and had been thoroughly evaluated as part of the development process or the task they were asked to perform was vague or that they did not really understand the method process, which involves them recording problems they encountered. It is therefore arguable that the latter seem more accurate as it was observed that they did not write down some other problems they encountered, see table 5.9.

With the result from the observers' data, the children encountered a number of issues that are believed to have affected their performance. Examples of such issues include:

Missing Information: During the HE using undescriptive task may have restricted children from exploring the application and find more usability problems. E.g. 'start and play game'. This type of task would have restricted children from exploring the game and finding problems as stated by the DAAR model that discovery analysis (e.g. in this case the tasks for finding the problems) needs to be creative and open while the analysis resources should be cautious and thorough (Woolrych & Cockton, 2002). Although it wasn't clear how much creativity and openness is needed. It was reported severally that 'children didn't either understand how to play game or asked assistance to play game'. Also it was observed that the facilitator did not show up heuristic list during an explanation. This was analysed to have been reported for just one group, therefore difficult to make a claim that this could have affected all the children in finding problems but may have affected the group of children who were being explained to.

Language: Some terminologies used in the heuristic set and severity scale were difficult for children to understand as observers' report state that they repeatedly asked for the meaning of some terms. This is confirmed by Salian *et al*., (2013) who stated that children had issues allocating heuristics beyond the sixth heuristic and also had issues attaching severity to problems found. They further recommend a rephrasing of the heuristic set and an alternative severity scale.

Children: This theme would be discussed in light of its subheadings to highlight the issues identified in it.

- *Bias:* It was observed that when a child verbally stated a problem during the session, the other children wrote down the problem.
- *Engrossed in Game:* Children at some point were immersed in the game, forgetting they ought to find problems until the facilitator or observers had to prompt them to remember what other thing they should do aside playing the game.
- *Not Recording Problems:* Children were not trained for this evaluation. It is therefore possible this affected their ability to effectively record problems. Observers noted some evaluators failed to record problems encountered e.g. an observer's note reads "couldn't create player (not put on form as a problem)". However, this could be a possible issue with this type of evaluation as it was reported in 5.2.4 (the pilot HE study with adult expert) that an evaluator also forgot to record problem found.

In order for the HE method to be effective for use with children, these issues and challenges need to be addressed. These problems may have contributed to the low yield in the number of usability problems reported and therefore it could be stated that this HE study failed because according to Woolrych & Cockton, (2002) there are three failure modes of a HE study or any predictive method evaluation of which some that exist in this study are: misses (failure to report problems that exist) and inappropriate use of method (when a real problem is reported but an inappropriate heuristic is applied to explain the problem).

5.4 Conclusion

This chapter has successfully reported pilot study carried out in heuristic evaluation with adult experts and HE with children. With the study reported in this chapter, children's performance in a UIM study was assessed fulfilling RO2 and shows that the study was a fail as discussed in section 5.3.3 above. The HE method in its state is problematic for use with children due to the issues they encountered whilst performing the evaluation. This difficulty with the method was also evident with the number of times they asked for assistance. They also found it difficult to understand some terminologies in the heuristic set and severity scale.

This confirms Salian *et al*.'s (2013) findings and was speculated, given the children used similar evaluation tools (heuristics and likert type severity scale) unless that the game evaluated in Salian's work is different from the one evaluated in this study and their severity scale had only 3 points while the one used here had 5 points. Also, some observed problems revealed within this study were not reported by Salian *et al's* study e.g. issues like children biasing themselves during the individual stage and facilitator biasing children in the problem merging phase. Children not discussing and having issues agreeing on a severity during the problem merging phase. Therefore, it was useful to have carried out this study which showed issues that needs to be addressed or investigated with subsequent studies.

In this study, the merging phase for the children's evaluation differ a little from that of the adult evaluators because it was believed children will encounter difficulties handling the merging phase all by themselves; so for the children's merging phase, the facilitator acted as a scribe unlike in the case of the adult where they handled the merging phase all by themselves. Although an observer's data reported the facilitator biased the children during this process, there was no report of it affecting the evaluation process or the children's ability to perform the evaluation. However going forward it will be useful to take extra caution not to bias them to avoid influence on their response or change any result.

It is also evident from this study that the method in its current state, is not yet easy to use for children, considering the inappropriateness of applying the method and inability to record problems encountered. Also the lack of success in using the method could be seen in the comparison of the results from both studies (pilot studies with adult experts and HE with children), which shows that adults found more problems than children did and that children were encountering problems but didn't realise it was a problem. This might have been due to several factors e.g. the pilot study was carried out with adults who had better understanding of problems and its consequence when they encountered it. It could also be that the time allocated to the introductory explanation which had the task instruction was insufficient for the children to understand what they were asked to do (given that they kept asking for assistance on what to do). Also, since the adults are experts in performing this type of evaluation, their prior knowledge could have been an advantage over the children who could be classified as novices in this type of evaluation or task. However the usual UIM evaluation issue (evaluator effect) when evaluating with multiple experts was not reported in the study with children. Therefore, it is arguable that this issue can be managed easily with children. It is also possible that this is the case because children are not sure and does not know what to do in this type of evaluation. So they were not fully expressive of their thoughts/ opinion as was the case of the adult evaluators.

As stated by Salian *et al*., (2013), MacFarlane & Pasiali (2005), and Wodike *et al.*, (2014) and with evidence from this study, it is believed that the method is not currently easy for children but needs adaptation and with the appropriate modification and adaptation, children's performance could improve when performing a heuristic evaluation. It is also believed the adaptation process could result in a new usability inspection method for children.

Work reported from chapter 6 to chapter 10 details the processes of adaptations and the final method that was developed.

5.4.1 Contribution of Chapter to thesis

Studies carried out in this chapter informed the research on the suitability of heuristic evaluation method with children as the evaluators. With the study in this chapter, it was clear on certain issues that arose from performing a heuristic evaluation with children as the evaluators. For example:

- Children biased and influenced themselves during the exercise

- The facilitator also biased the children during the merging phase

It also showed that children found it difficult to understand the heuristic evaluation method in its original state and the tools since it was not adapted yet to suit children. For example:

- They couldn’t understand some terminologies in the heuristic set and severity scale.
- They also found it difficult to attach heuristics to problems found.

Finally it was concluded from the outcome of this chapter that a well described task could have improved children’s performance.

All these findings showed areas of focus for adaptation which eventually led to the creation of the new method.

6 CHAPTER SIX: GATHERING TEACHERS' INPUT

6.1 Introduction

Having identified issues children encountered while carrying out the HE study with children as reported in chapter 5, section 5.3.3, a peer debriefing was made with other experienced researchers which drove the decision of gathering input from teachers on how to modify the HE method to suit children. Since they are part of the stakeholders in the design of learning materials for children and have experience in working with children. Also literature records the successful involvement of teachers in the design (Scaife & Rogers, 1998) and evaluation of technologies for children (Pardo, et al. 2006). Although teachers could be considered as novices on how usability evaluation method works, their knowledge of the children, how they learn and how they could be facilitated in a task based activity could be useful to the session. Also, given the proof from literature on teachers' involvement in design and evaluation with children, it is conjectured that an explanation of the method procedure and provision of documents to describe certain aspects of the method (e.g. tools of the method) could provide them with enough information required to make useful contribution towards the topic of discussion.

In addition the review of literature inspired the decision to gather these inputs via focus group (FG) studies. This is because of the benefits of the FG method reported in chapter 4, section 4.5.4.1.1 but particularly for the reasons stated in section 6.2 below. Therefore, this chapter will report the iterative FG studies carried out to *investigate whether input from teachers can inform the redesign of the HE method to make it suitable for children.*

The chapter is divided into 6 parts (4 sections and 2 subsections where the first section has two subsections). Section 6.2 describes the method used, 6.2.1 reports the first focus group study, 6.2.2 reports the second focus group, section 6.3 will discuss the findings from both studies and 6.4 discusses the findings from both studies and 6.5 concludes the chapter.

6.2 Method

As stated earlier, the FG method (where multiple teachers are involved in a group interview. See section 4.5.4.1 for full description of a focus group) is employed for these studies because it is unique and advantageous for its group dynamism (Krueger & Casey, 2000), where the contribution of one participant can inspire ideas in another. The informal structure of the method referred to as socially-oriented (Morgan, 1997) provides an avenue for capturing real life data and will allow teachers the opportunity to form collective ideas (Blanchet-Cohen &

Reilly, 2013). Furthermore, with the FG method multiple data could be collected from more than one person per time unlike a standard one on one interviewing method.

Two focus group studies were carried out iteratively with school teachers therefore the experiment will be reported in two sections to reflect both studies.

6.2.1 First Focus Group (FG) Study with School Teachers

This first study gathered the first inputs from teachers which will inform the subsequent study.

6.2.1.1 Participants

The total of 6 school teachers who taught years 2 to 6 (ages 6 – 11) from a UK primary school participated in this first FG study. In addition two researchers from the ChiCI group (the author and another researcher) played the role of the group moderators (moderator 1 and moderator 2 respectively). Both researchers are knowledgeable on heuristic evaluation method and have practical knowledge of running focus groups.

6.2.1.2 Apparatus

Prior to starting the study, participants were provided with documentations which included a description of the heuristic method, a copy of a heuristic set by Korhonen & Koivisto (2006), the severity scale by Nielsen (1995) (see appendix 1Afor heuristics and severity scale), and evaluators' data collection form (see appendix 1B) all used during the HE study with children (see chapter 5). They also signed a consent form (see appendix 2A) permitting the researchers to record the conversation which was done with 2 Dictaphones.

The researchers also had copies of the documentation in order to explain the process to the teachers. Moderator 1 also had a set of questions (see section 6.2.1.2.1 question design below for actual questions used) with which the study was moderated, writing pad and a pen to capture key points and a wrist watch to time the study, ensuring it doesn't exceed the planned time.

6.2.1.2.1 Question Design

A set of open ended questions were used for both studies. These questions were constructed with careful consideration of Krueger's (1994) guidelines on how to ask questions that would yield useful information in a focus group study. The content of the question was derived from the issues children faced during the HE pilot study reported in section 5.2.3.

6.2.1.2.1.1 *Actual Questions*

FGQ1. How would you recommend we teach year six children to carry out the heuristic evaluation?

FGQ2. How should the heuristic set be presented and explained that children would understand?

FGQ3. How would the children be told to do this without being biased by their peers?

FGQ4. How should the severity scale be designed that will be appealing to Children?

FGQ5. How should the severity scale be explained that will be understandable for children?

FGQ6. What do you think about the format of the evaluators' data collection form?

FGQ7. How do we prevent the children from being engrossed in the game during the evaluation process?

FGQ8. How can children be taught to empathize with younger children whilst carrying out the evaluation process?

FGQ9. How should we explain the problem merging process that will be engaging and fun?

FGQ10. What do you think about the order of the evaluation process?

FGQ11. How can we make the whole process engaging and fun for the children?

6.2.1.3 Procedure

This first focus group study was carried out in a quiet classroom, prior to the study participants were given the documentation in advance to read and become familiar with the content but it was discovered the document was not read. Therefore, the documentation was provided to the participants at the start of the focus group before the introduction (the documentation include: The study consent form, the description of the HE method, and the heuristic set and severity rating sheet which the children had used for the HE study reported in chapter 5). The study started with the moderators introducing themselves, then moderator 2 explained the aims of the study and informed the participants of the study ethics; stating that though the discussion will be audio recorded and transcribed unabridged, teachers' voices and names will be made anonymous and names mentioned will be changed.

The teachers were allowed some time to read, complete and return the consent forms (see appendix 2A) to moderator 1(first author). They were further asked to read the documentation that contained the heuristic method (see appendix 2B for document). In order to ensure the teachers understood the method, moderator 2 went further to explain how the method works and restated that input gathered from the study will inform the redesign of the heuristic method where children aged 9 to 11year olds will act as the HE evaluators. Therefore teachers were encouraged to critique any part of the method and be honest with their opinion. At the end of the explanation that lasted for 5 minutes, moderator1 asked the teachers the first question, which started the FG discussion:

"As teachers of these children, how would you recommend we teach year six children to carry out this heuristic evaluation?"

During the entire discussion, both moderators used the set questions (see above) to moderate and gather more input from the teachers. In order to ensure that the appropriate and correct input is collected, teachers' comments or responses that were unclear was repeated to the teacher (s), who responded in agreement if right or repeated what they meant if not correct. All the participants responded orderly, speaking one after the other. One teacher never really made a clear suggestion but rather nodded and consented to what other participants were saying. Meanwhile, another participant was very expressive compared to the others and seem to always raise a point and pass across her ideas, but this participant went quiet to listen to others when they spoke. At the end of the discussion, moderators thanked the teachers and stopped the audio recorder. The discussion lasted for approximately 27 minutes.

6.2.1.4 Data Analysis

Data analysis was carried out by 2 researchers (the moderators) using a qualitative content analysis approach (Mayring, 2000) in 'NVIVO'. One researcher (the author) did the first analysis and the second researcher (moderator 2) inter coded the data as a reliability process.

Prior to data analysis, audio data collected from the study was transcribed unabridged, though sounds like 'uhm', 'ehm' 'hmm' were removed. The transcript was subsequently compared with the moderators' notes taken during the studies to minimize errors during transcription. The new transcript was read through, organized in sections and labelled with headers that related to the questions asked during the focus group study. The organized data was read through over and over again to gain mastery of the data and bring together sections that were disjointed.

The first set of themes or nodes (as referred to in NVIVO) which formed the main themes were automatically created with the headers e.g. heuristic set, severity scale, merging process, curtailing bias, prevent game engrossment. However, after the inter coder reliability process and coders' discussion on code variations, it was evident that some of these themes are generic i.e. it can apply to other evaluation methods with children while some were specific to the heuristic evaluation. Therefore, two new themes were created to reflect this view. The generic view became the main theme 'wider evaluation context' while the specific view became a sub theme ‘heuristic evaluation method’ and categorized underneath the main theme. In addition, previously coded main themes were restructured to reflect these changes.

6.2.1.4.1 Qualitative Content Analysis (QCA) Process/ Themes emergence

As stated earlier, Mayring's (2000) approach of QCA was used for analysing the data whereby each section of the data (discussion) was analysed and coded step by step. Words that occurred frequently and valid words or phrases (those that have meaning in context of the research question) were coded as sub-sub themes in some cases and sub-sub-sub themes in some other cases. These sub, sub-sub and even sub-sub-sub themes were coded into their appropriate associated higher themes. Responses that would support each theme (either sub or sub-sub or sub-sub-sub) were coded into them appropriately to aid data interpretation.

6.2.1.4.2 Data Reliability

After inter coder activities, it appeared analysts had similar themes but with little variations. However, upon analysts’ discussion in line with the research question and objectives, analysts concluded reaching a consensus with agreed codes.

6.2.2 Second Focus Group Study with School Teachers

This second focus group was carried out to confirm findings from the first focus group and to explore the opinion of more teachers.

6.2.2.1 Participants

The total of four school teachers who taught years 3 to 6 (ages 7 – 11) from another UK primary school participated in this second FG study. Similar to the first study two researchers from the ChiCI group (the author and another researcher) played the role of the group moderators (moderator 1 and moderator 2 respectively). Both researchers are knowledgeable on heuristic evaluation method and have knowledge of running focus groups.

6.2.2.2 Apparatus

The same apparatus used in the first study, were the same used in the second. However additional questions which were derived from the result of the first focus group was used here.

6.2.2.2.1 Questions for Second Focus Group

The actual questions used in the first FG were the same for the second FG. However, some suggestions were made by the teachers of the first focus group which needed confirmation (validation). So as a validation technique recommended in (Patton, 2002) and to explore the second study teachers’ opinions on suggestions made in the first FG, questions were drafted from the themes identified after first FG data analysis. In order to moderate this second study and make the validations, these newly drafted questions were used as questions and prompts in addition to the actual questions (questions used in FG1). Below are examples of these questions:

- What do you think about using scattered graph or tally chart for the severity scale?
- What is your view about the children using Venn diagram to represent common problems?

6.2.2.3 Procedure

The second focus group which held in a staff room during lunch time on a school day followed the same structure as the first study. Though in this study the additional (validation) questions were asked after all the actual questions have been asked. As stated in the previous section, this strategy intends to validate and further explore these themes. All the participants made contributions to all areas of the discussion and the study lasted for approximately 26 minutes.

6.2.2.4 Second Focus Group Data Analysis

The second focus group data was transcribed abridged whereby only questions and responses to questions were typed. However, the other processes were the same as with that of the first focus group data analysis.

6.3 Result for FG1 and 2

Without any experience of performing the actual evaluation, based upon documents provided by the researchers and an explanation of how the method works the teachers were able to understand the heuristic evaluation process and contribute to the discussion. All the teachers within the two focus groups provided suggestions, support for ideas and could be critical of the current process. The discussion confirmed issues that were already known based upon the literature and offered new insights. The results are presented within a wider evaluation context and in the context of the heuristic evaluation process.

6.3.1 Input within a wider evaluation context

6.3.1.1 Teaching children to carry out the evaluation process

After analysis of teachers' input, nine workable themes emerged. Some themes are established ideas in the CCI community when designing evaluation methods for and with children. For example, the fact that the facilitator or experimenter should 'explain the process and show the software', 'start with the product' and 'give them an idea' is seen in the PIPC method (Barendregt *et al*., 2007). The need to 'use child friendly language', 'simpler settings' and 'easy rules for them to follow' are standard approaches already in use. However, there were some other themes that are unique, interesting to explore and that will be useful to the CCI community, and they are:

Having them 'do the process as a discussion' is not be new in CCI but its usage has been in a collaborative design process with children (e.g. cooperative inquiry (Druin, 1999a)) and in a user testing (e.g. constructive interaction (Nielsen, 1993b)). However, it has not been used in usability inspection method with children which will be a worthwhile process to carry out. This might be a useful process to help children be confident to talk about problems they found, especially for the merging process. As reported in (PIPC study) children are more comfortable to talk to people they are familiar with and therefore this collaborative discussion may aid the aggregation process. It was suggested that 'doing role play' could be a viable option to teach the process, their reason was, this will get the children involved in the activity, making it fun and engaging for them. This in the long term might enhance their performance in carrying out the evaluation method. Another contribution was 'Using a booklet or a written document in a child language'. The use of a booklet in an evaluation process is not new since it was used in structured expert evaluation method (SEEM). Their rational for this point is that in the absent of an adult facilitator, children could discuss amongst themselves using this booklet and carry out the method. These points if feasible to implement could help produce an effective evaluation method for children.

6.3.1.2 Preventing bias from their peers

In curtailing bias, teachers contributed eight points (sub-sub themes) of which three appeared to be useful. Teachers stated a need to **making an application appealing to both parties** (boys and girls) to manage or rule out bias. Their claim is if girls are knowledgeable of the application and the process, they will not ask the boys for their opinion and vice versa. They also think when the issue is presented as being serious (**making it a big thing**), then children tend to put in their best and adhere to every instruction given. Finally, the **use of role play** to rule out bias was also discussed. Each child would have a specific role and act as the role rather and this may help alleviate potential bias from individual group members.

6.3.1.3 Preventing children from being engrossed in the game during the evaluation process

To prevent children from being engrossed in the game as reported in the pilot study, teachers said the experimenter or facilitator could discuss the game with the children, however this might not be useful for the evaluation process as the process require the children to individually explore and find usability problems in the game. They also commented that children could be steered with prompts e.g. “what do you think about the controllers”. This is also in the category of biasing the children. There are other interesting input with which they drafted the process on how to use the inputs together, this could be explored. The process involves “letting them play individually”, “giving them a time scale” afterwards you can draw them back to the heuristic list. Another interesting and novel input made was **using role play** and have them record problems on **post it note**. Their argument is that using post it and role play will get the children involve in the whole process rather than just in the game. However, it is difficult to comprehend how children will individually play the game and find problem in a role play activity. Therefore, it would be useful to explore these approaches to produce a heuristic method for children such that they won’t be engrossed in the game.

6.3.1.4 Teaching children to empathize with younger children whilst carrying out the evaluation process

Teachers believe **older children can empathise with younger children** and are the best people to do so. However, they believe children who have younger siblings might do better though to help children who have no younger sibling, **they could be told or shown the ability and capabilities of younger children**. In general, they believe children could be helped to empathise by telling to **have the intended users (younger children) in mind**.

6.3.1.5 Making the whole process engaging and fun for the children

In making the whole process engaging and fun, teachers’ responses yielded 6 categorised input:

- Use role play
- do it in small groups
- tell them it's asking their opinion
- do it as a visitor
- do the process early on with them
- Use engaging or interesting game

Teachers believe when a game is interesting then it will make the process fun for the children, however it is possible this could result in them being engrossed in the game. Unless the approach for preventing engrossment works. They also suggest that doing it in small group will be fun for the children, this is interesting but they did not buttress further on how this would be fun.

Meanwhile, other input made is quite interesting because it means children could enjoy an evaluation process when they know it's asking their opinion. They also believe the time of the day the process is carried out could affect the fun of the evaluation process. Finally, it is interesting to know that facilitating the process as a visitor will make it engaging and fun for them. This contradicts the view of the PIPC authors who stated that children are willing to speak when they are with familiar people, although it could be in the case of thinking aloud.

6.3.2 Input in the context of the heuristic evaluation method

6.3.2.1 Presenting and explaining the heuristic set so that children would understand

Discussions based on the heuristic set produced four already known ways to design guidelines or criteria for children and these included "Use child language or simplify the heuristic language", "cut down heuristic number", "break into sections" and "explain the heuristic set". However, the teachers made other suggestions that could be useful in redesigning the heuristic set to make it more appealing and interesting for use with children. These include: "**Making it colourful**" and "**giving examples for each heuristic**" or "**provide comment box for each heuristic**" so children could interpret the heuristic in their terms and further use it to evaluate the game. Making things colourful for children is not novel but making a heuristic set colourful is something new and interesting. Providing an example for each heuristic in relation to the application being evaluated could aid the children's understanding. The provision of a comment box will also enable the facilitator to clarify the children's understanding of the heuristic but may add to the duration of the study

6.3.2.2 Severity scales that will be appealing to children

In designing a suitable severity scale for children, teachers recommended that the scale should be reduced from a 5 point scale to 3 point scale. Their rational for this was that the 5 point scale might be cognitively demanding for children. This suggestion might be slightly contradictory as the Fun Toolkit uses a 5 point scale within the Smileyometer and is successfully being used. The teachers did add that the severity scale should be presented as smiley faces as this will make it easy and quick for children to decide on what rating is appropriate for problem found. Teachers also mentioned that the language within the severity ratings should be changed, this is well documented in the CCI literature that language used for children should be simple and child friendly. Another suggestion was the **use of scattered graph, the use of tally chart and the use of colours**. The first two points were raised in the first focus group but was criticised in the second focus group. Their claim was that using a scattered graph and a tally chart could be a new technique for the child, they would need to learn this in order to carry out the evaluation. In conducting a heuristic evaluation, it will be more effective if familiar tools and techniques are used. They also stated that children at that age (9 to 11years) will find it difficult to plot scattered graph, it may not be a fun activity and may ruin the entire process. This fact is similar to the claims made by the authors of Memo-Line, that children at this age lack the knowledge to draw a mathematical graph with negative values. Both focus groups were positive about the **use of colours** stating that children are familiar with using colours to represent things and situations. They both thought that severity ratings could be expressed through colour and could be in form of UK traffic lights with red green and amber.

6.3.2.3 Explaining severity scales to children

Regarding how to explain severity scales to children the teachers' contributed four ideas. "Put them together and ask them", "get them to discuss it", "they'd get it but it needs simplify" and simplify language. They believed that children understand the scales if the words are simplified and in child friendly language. The idea of discussing the scales with the children was done in the pilot study heuristic evaluation study with children, but observers' notes identified that children still struggled to understand the scale. However, their claim that children discussing it amongst themselves will help them understand it better is a viable option. As this approach is used within peer tutoring, this method has been used in some other method with children (Druin, 1999a).

6.3.2.4 The format of the evaluators' data collection form

Response in this section will be split into two: form design and form completion. Teachers input on designing the form produced a novel approach that involves **using colours** to make it attractive or jazzed up (in teachers' term) and child friendly. Colours are often being used with children in designing or evaluation but in designing a form that they will complete is scarce in literature. They also made input on not having lot on form and using table format, this is valid and could be used but is not new. In aiding children completing the form, teachers suggested examples should be given, language should be child friendly and children should be discouraged from writing too much. These are approaches already being applied with the fun toolkit method.

6.3.2.5 Problem merging process that will be engaging and fun

Participants for study I and 2 believes children could discuss and do the merging process but to make it engaging and fun, **role play could be used with one person acting as the scribe**. They also suggest that **children could use post it note and white board** where by one person will start by posting their problem which is already on a post it to a white board so all can see it, subsequently those having similar of the same problem could post theirs as well. Then someone will act as the scribe to write the problem down before moving on to the next problem. This sounds interesting and workable. They also suggest **children could use colouring** but this was not clearly stated on how this could be done. Another interesting input was the use of Venn diagram suggested by study 1 participants, however, participants in study 2 believe the use of **Venn diagram** might not work because again that would be the children learning new skill and that Venn diagram for children, works well with three items at the most. In usability studies a large corpus of problems are usually generated and thus this may cause children additional problems.

6.3.2.6 The order of the evaluation process

Teachers suggested 4 order in which the evaluation process should be: “Find problem and fill form as they go along”, “Game first then activities”, **“Game, heuristic, severity rating and form”** and **Game, heuristic set, form and severity scale**. However, the first two options are vague unlike the last two but in all teachers believe the process should always start with the application (game). They believe the game will be captivating to make the children continue the process.

6.4 Discussion

Can input from teachers inform the development of training material and protocol for children in the context of UIM?

In an overview teachers input as to how this method could be adopted situates around using child friendly language and attractive materials to make the method engaging and fun for children. Examples of such input was the use of colours to make it attractive and appealing for children, using simple terms and child language for the heuristic set and severity scale, avoid too much writing for filling the form. These would be discussed with previous evaluation methods for children.

MEMOLINE In this method, children used colours to indicate different situations on their long-term experience with an application (Vissers *et al.*, 2013). For example, a green colour indicates positive experience, red colour indicates negative experience and grey colour for non-usage.

FUN TOOLKIT In the fun toolkit, the fun sorter is used to collect the child's opinion about a technology or an activity that will be used to measure the child's engagement, is done by ranking using note description but in most cases, is done using picture cards with the aim to discourage children from too much writing (Read, 2007). It is therefore arguable that discouraging too much writing which teachers highlighted, is an acceptable phenomenon in the CCI community.

Teachers could also make critical suggestions to generate viable input. For example, in study 1, it was suggested that children use Venn diagram for the merging process. However, in study 2, participants believe using Venn diagram will be too complicated for the children when more than 3 items are involved also they think children having to use Venn diagram for this process, will require them to learn a new skill and for this kind of evaluation, it is recommended that they use familiar tools and methods.

6.5 Conclusion

This work carried out 2 focus group studies with school teachers and the result show that though teachers made novel and not novel input, their input could inform the redesign of the HE evaluation process and also be used within a wider evaluation context. The study also proves that despite being trained, teachers understood the explanation given for the evaluation process and made sensible discussion in context of the subject matter. Finally, as stakeholders in children's cognitive development and as experts in developing learning materials for children, they know the learning and cognitive capabilities of these children and could make critical arguments to support their discussion. Therefore, their input could become viable and effective in informing the redesign of the heuristic evaluation method for children. This method could also be adopted in designing other cognitive based activities for children.

Some suggestions / recommendations made by teachers from this study was further explored with children in a design session which is reported in the next chapter (7), example of such recommendations include:

- Having the children come up with their own game criteria

- Have them discuss the game criteria they came up with, to have better understanding of each one
- Let the children discuss the severity
- The traffic light colours or colours could be used for the severity scale
- A three point rather than a five point severity scale should be used
- Post it note should be used to record problems found and the post it note on a white board could be used for the merging phase
- Role play could be used during the evaluation to make it fun

6.5.1 Contribution of Chapter to Thesis

Findings from the studies in this chapter informed the decision to create the new method. These input were explored and formed part of the content of the method. The following are recommendations from teachers and contributions from this study and chapter:

- Teachers suggested that children should discuss the heuristic set so they can better understand it before using it during the evaluation. However, this might be challenging in a study drafted into a small amount of time.
- Teachers suggested the use of traffic light colours for the severity rating scale. Stating that children love colours and that the use of colours during the evaluation will capture children's attention and will be fun for them.
- It was suggested that children use post it note for recording problems found to make it fun and to make the merging phase easier. It was also suggested that the post it note will make it easier for the children to fill in or report problems found than they having to complete a mapped out huge form.
- It was also suggested that role playing will make the process fun, however it is unclear as to how this will fit well into an evaluation exercise.

From the suggestions teachers made, some ideas were developed that will be tried out with children for example:

Instead of asking children to discuss existing heuristics which was originally designed for adults, children will be encouraged to produce their own criteria with which they can find and contextualise problems found. With the review of literature on children's cognitive ability reported in chapter 2 of this thesis, and designing and evaluating with children reported in chapter 3, it is evident that children are experts of the world around them. They are able to state what they want to see in their intended application or technology. Therefore, it is hypothesized that they are capable of producing their own game criteria based on their values that could guide them to evaluate their own applications.

Also instead of involving the children in a role play, storytelling could be used to inspire the children to perform the evaluation (More detail is available in chapter 7).

7 CHAPTER SEVEN: EXPLORING TEACHERS' IDEAS TO DESIGN THE NEW METHOD

7.1 Introduction

It is an established principle in HCI to involve users in the development of products and a common trend in the CCI community to involve children in the development (e.g. design (Druin & Solomon, 1995; Druin, 2002) and testing (Baauw & Markopoulous, 2004; Kano, Horton, & Read, 2010; Read et al., 2002)) of tools or methods (Read *et al*., 2002; Vissers *et al*., 2013; Zaman, 2007, 2009a) tended to them (Read & Bekker, 2011). The results of the focus group studies with teachers reported in chapter 6 show several suggestions which included the suggestion of consulting children in the development of a suitable heuristics and severity scale for them, and the simplifying of the method for the children (see section 6.3). Therefore, it was decided that a new inspection method will be created that will be easy to use for children to carry out evaluation on technologies designed for them. After the review of related literature on the development of evaluation methods suited to children (e.g. Barendregt *et al*., 2007; Read *et al*., 2002; Vissers *et al*., 2013; Zaman, 2009a), the decision was made to create a method where storytelling instead of role play (as suggested by the teachers in chapter 6) will be used to facilitate the evaluation process. This is because it seemed far-fetched to logically include role play in an evaluation method within the scope of this research, as there is currently no report of role play in CCI in this context. However, storytelling is a concept that have been explored within CCI and have been reported successful when designing with children. Although this is not in evaluation, but it is believed that this could work well in facilitating the evaluation process where children will carry out the evaluation based on their values.

Therefore, this chapter aims to *explore storytelling as a tool for gathering input from users (children) and value as basis for users' decision making*. It reports the review of related literature on the techniques (storytelling and values) employed for the new method. It also illustrates the involvement of children in exploring teachers' ideas on the development of evaluation tools tended to them and reports the review of the first version of the method by an independent teacher.

The chapter is divided into three parts: Part A (Understanding some techniques that will form part of the newly designed method), Part B (Exploring children's ability to produce their own game criteria and their ability to understand the concept of the severity rating scale) and Part C (The review of the early version of the new method).

Part A reviews related literature on storytelling as a method and how it has been used with children in design. This part also defines value and describes the concept within HCI and CCI. Each part reports lessons learned and its use within the proposed new inspection method. Part B reports studies carried out with younger children aged 6 to 7 years old and older children 9 to 11 years old, where narrative was used as a technique to investigate children's ability to produce their own game criteria and design their own severity scale based on their values; and rate usability problems using the traffic light system (TLS). Part C discusses the review of the first version of the method by an independent teacher, what was removed after the teacher's review and what was included.

7.2 PART A: Creation of IMCH version 1

This part of this chapter reviews literature on Storytelling and Values and made discussions on how this is implemented in the new method proposed by this research.

7.2.1 Storytelling

Oxford English dictionary (OED) defines storytelling as the action of telling stories (Oxford-University-Press, 2015b). Definitions that exist by other researchers are: Storytelling is the effort to communicate or the conveying of events in words, images and sounds often by improvisation or embellishment (Haigh & Hardy, 2011; Mokhtar & Kamarulzaman, 2011). It is the oral or written form of our stories that we share with others (Chaitin, 2003). The national storytelling network defined it as the interactive art of using words and actions to reveal the elements and images of a story while encouraging the listener's imagination (National Story telling, n.d.). This work will like to define storytelling as the art of using words, images or sounds to communicate events or stories either by improvisation or embellishment to encourage or inspire the imagination or creativity of the listening audience.

Stories play a huge role in the way we view the world around us (Grimaldi *et al.*, 2013). Much of our lives as humans are conducted via stories. It captures the richness and differences of meaning in everyday existence and gives insight into the complexity of our experiences (Garrett, 2006). It is the vehicle that we use to summarise and remember experiences, and further communicate them in a variety of ways or instances to our chosen audiences (Forlizzi

& Ford, 2000). Stories are reflective, creative and value laden, usually revealing something important about the human condition (Haigh & Hardy, 2011).

Winterson in Haigh and Hardy's work states that "stories are always true; it's the facts that mislead" (Haigh & Hardy, 2011; p. 409). This statement could be true, however a true story is determined in the way it is told. According to Chaitin (2003), the content of a story is not judged to be true or false, solely with respect to its adherence to the empirical fact, but with respect to narrative criteria such as believability and coherence. Therefore, it is arguable that storytelling has a skill (Haigh & Hardy, 2011) that the teller ought to have.

Although Grimaldi et al, Forlizi and Ford, believe that storytelling is the most natural way for passing on information (Forlizzi & Ford, 2000; Grimaldi et al., 2013). Kearney argues that it is as basic to human beings as eating, something that makes our lives worth living and makes our condition human (Kearney, 2002). Haigh and Hardy (2011) argues reflecting the stance of Silver (2001) that storytelling does not require literacy, equipment or energy supply and Mokhtar *et al* states the same that "storytelling needs no special equipment. However, Mokhtar *et al's* continuing statement states that what storytelling require is not beyond imagination and the power of listening and speaking to create artistic images" (Mokhtar & Kamarulzaman, 2011; p. 164) where that power could be argued as the skill required to make a good story. Also, as gathered from literature (Garrett, 2006), Boje (1991) in (Haigh & Hardy, 2011) states that it is important that the story being told 'makes sense' such that it is believable to the audience and can engage them in the facts of the story and participate where necessary.

Storytelling has been reasonably researched and explored in different ways; for example in healthcare it has been used as a way of generating trust between nurses and patients (older adults) and as a means of signposting access to health resources (Haigh & Hardy, 2011). In organisational learning, it has been used to promote a multiplicity of viewpoints which is grounded in a reality that is recognisable to students (Boyce, 1996). In a language classroom, as an effective teaching tool and a communicative skill enhancement (Mokhtar & Kamarulzaman, 2011). In teaching and learning, it has been used as a strategy to foster critical and inclusive ways of teaching PE students who are girls (Garrett, 2006).

It has also been explored within the design field in different ways: Forlizi and Ford as part of their framework, identified narrative or storytelling as a useful tool for designing better user experiences (Forlizzi & Ford, 2000). It is also believed to be an integral part of game design (Duh & Chen, 2010).

Over the years when designing for children, story has played and still plays an important role for entertainment, engagement and learning (Druin *et al*., 1997; Wang *et al*., 2008) where a story telling system is either designed to tell children stories or allow children to create their own stories with the opportunity to learn and or be entertained (Montemayor *et al*., 1999). Also within design the shared narrative space has also been explored in gathering user requirement when designing for children (Dindler *et al*., 2005). However, for the design of usability evaluation tools, there is scarcity in literature on the exploration of this concept. It has been established that stories are prevalent in children's everyday life, it is useful to teach and communicate with children and also encourage them to express their mood (Montemayor *et al*., 1999). Therefore, this work will explore the use of storytelling to engage, motivate and inspire children to come up with criteria that will make a good game (good game criteria) based on their values.

7.2.2 Values

There are dictionary definitions given for the term "values" e.g. WebFinance defines values as an important and lasting beliefs or ideals shared by the members of a culture about what is good or bad and desirable or undesirable. It went further to state that values have major influence on a person's behaviour and attitude and serve as broad guidelines in all situations (WebFinance, 2015). The Oxford dictionary defines it as principles or standards of behaviour or one's judgement of what is important in life (Oxford-University-Press, 2015c). Several researchers have also defined the term (e.g. Gutman, 1991; Iversen et al., 2010; Posner & Munson, 1979; Schwartz, 1994) but as stated in section 1.2 of this thesis, this work will define values following the definition by Iversen *et al*., (2010) which says it is something that a person or group of people consider to be important in life.

As seen in the earlier definitions above, values have the potential to influence a person's behaviour and attitude and can influence their judgement of situations. In the CCI community 'value' is a concept that is adopted within design for and with children. It is an established trend to involve children as users, testers, informants or design partners when designing technologies for them. Literature show that using these methods, children have conveniently been used to design educational technologies (Druin *et al*., 1997; Read *et al*., 2013; Scaife *et al*., 1997) for them. The theory behind the conscious involvement of children is the belief that children are experts on the things concerning them and are able to tell what they perceive as important in the now and future technologies for them. This could be considered as children giving requirement based on their values. Though this is a take on in design with children, it is yet to be a practise in inspection method especially when children are the inspectors (evaluators).

Since this research has made a decision to use storytelling to facilitate the evaluation exercise such that children can easily relate to what is required of them and make input based on their values, the scripting of the story is therefore important.

7.2.3 Scripting the Storyline

In order to ensure the story is understandable, believable and coherent, it was kept simple. Also to encourage engagement and seriousness in children's participation, a story that involved an absent third party was told. Although the story told was unreal, it was ethically checked with another researcher (who understands ethics that concern working with children) to ensure it didn't have any misleading facts or negative impact on the children. An example of a story that was used is:
I have a friend who works with computers and games. My friend is looking to design a good game for children like yourselves and younger children too but he is not sure about what will make a good game for you. He asked me, but I wasn't sure myself so I told him I would come and ask you.

After considering and reviewing the literature in "storytelling" and "value", the first version of the method was produced thus:

IMCH V1 (see fig 7.1 for schematic representation of first version)

- Narrate a story to children about a game in a particular genre similar to what they will eventually evaluate (e.g. I know a friend who builds games for children and he wants to build a spelling game for children like yourselves but he is not sure on what will make a good spelling game… (see section 7.2.3 above for more detailed story) and ask them to come up with what will make a good game (in the stated genre i.e. a spelling game) based on their values

- Give them a spelling game to play and evaluate using the game criteria, find problems, record it on a post it note and further rate the problems with a traffic light system (TLS) severity scale.
- Have them merge their problems by discussing their found problems and put together post it notes that had similar problems
- Agree on a severity rate (colour e.g. green or yellow or red) for the merged problem to give the final severity rating for each merged problem.

Inspection **M**ethod For **CH**ildren (IMCH) Version 1

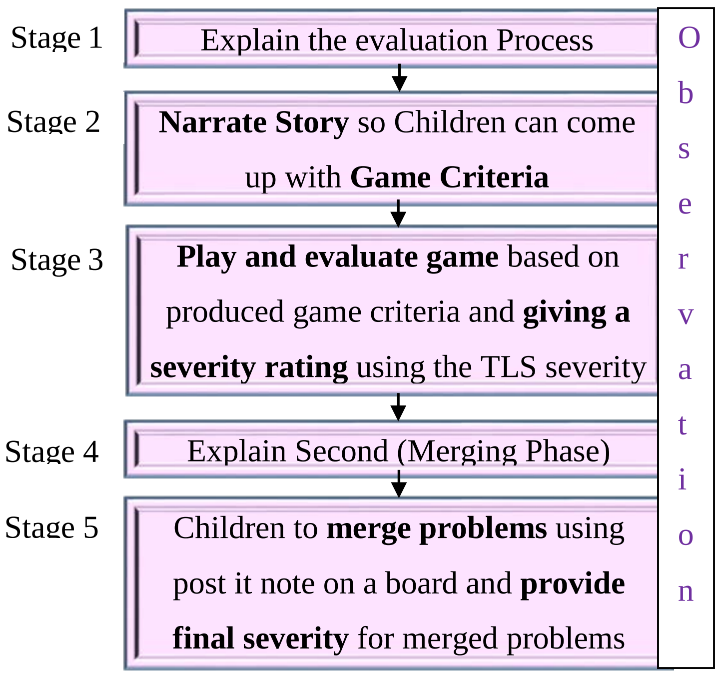

Figure 7.1 The IMCH Method Version 1

Though the above method breakdown is the proposed process for the version one of the IMCH method, it was not yet fully ascertained that it will work with the children. For example, it was not certain that children could produce game criteria based on their values after the narration of the story. It was not also certain that they will understand the TLS severity scale for rating problems found. Therefore, it was decided that these parts of the method should be tested with children to determine how the method should go. Full details of testing is provided in section 7.3 below.

7.3 PART B Testing Parts of the Method with Children

In this session, narrative was used to facilitate the process where children were encouraged to come up with their own game criteria based on their values, illustrate their understanding of the concept of severity by drawing their own severity scale and also to test children's ability to use the traffic light system as a severity rating scale.

The aim of these studies is to test the appropriateness of the intended techniques (narrative and values; which will form part of the new method (IMCH)) when gathering input from children. In order to achieve this aim ensuring wider input, two studies were carried out with children aged between 6 and 11years old: the first study was carried out as a pilot with younger children and the second was a main study with older children.

7.3.1 Study 1 (Younger Children)

7.3.1.1 Method

An informant design study (as described in 4.5.4.2) was carried out with younger children (aged 6 and 7). The holistic aim of this first study *is to pilot the method and have an insight to the capabilities of the actual target group (children aged 9 to 11 years old) for this research*. Although it would have been more appropriate piloting the method process with the target group of the research. However, the school in which the research was carried out was only able to provide access to this age group even though this was contrary to an initial request and since it is difficult to recruit schools or turn the children down, the decision was made to carry on the pilot with this age group. With the speculation that if these younger children are able to understand the method and produce reasonable data then older children should do as well or even better. The first part of the study involves children producing their own game criteria, the second part involves them drawing their own severity and in the third session children were to rate the severity of some usability problems identified from literature using the traffic light system.

7.3.1.1.1 Participants

In total 21 children aged 6 – 7 years participated in this study. All the children were from a year 2 class in a UK primary school. The children were directed out of the classroom by their teacher and put in groups for the study. A total of six groups emerged where groups 1 to 4 had four children while groups but 5 and 6 had two children each (this was the other in which the children were sent in for the study). The author was the only researcher present for this study.

7.3.1.1.2 Apparatus

The researcher used a sheet that had the study guide and the narration to facilitate the study. During the first activity (Game criteria) the researcher captured game criteria stated by the children on the game criteria form (see appendix 3). With this form the researcher was able to capture the group identity, game criteria and frequency of children that identified each criterion. In order to capture children's severity drawing children were provided with a sheet which had only the title "the severity rating scale" (see appendix 3), so they will have enough space for their drawing.

For the third (final) activity, twelve problems and screenshots of its associated games from literature (Baauw *el al.*, 2006; Bekker *et al.*, 2007; Donker & Reitsma, 2004; Salian, 2012; Sim, 2012) were provided on four separate sheets, where two sheets had two problems each, one sheet had three and the last sheet had five problems. The first problem sheet to be coloured had the age, class, group and identity (e.g. A, B, C or D) spaces for the child to fill in while the others had only group and identity spaces to be completed. The games for which screenshots were captured are JamMo: A Music making Game (see figure 7.1), Leescircus: A beginning reading game (see figure 7.2), Milo and the Magical Stone (An adventurous game) (see figure 7.3), and the Anti-heuristic Space Invader Game (see figure 7.4).

Figure 7.2 Jamming Music (JamMo): A music making game

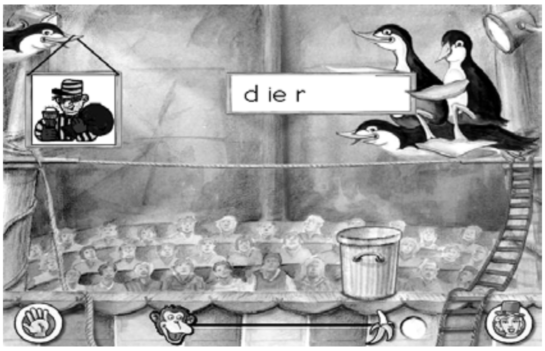

Figure 7.3 LeesCircus: A beginning reading game -drag letters to trashcan view

Figure 7.4 Milo and the Magical stone screenshot

Figure 7.5 Anti-heuristic space invader game

7.3.1.1.2.1 Description of the Games Used

Jamming Mobile (JamMo) Game (Fig 7.2): This is a music making game targeted at children aged between 3 and 12 years old. The game has two sub games: singing sub game (a sub game that sings) and a composition sub game. In the singing sub game, players are expected to find and play an already stored music while in the composition sub game, players are to compose (create) their own music (Salian, 2012).

The Leescircus Game (Fig 7.3): This is a beginning reading game targeted at Kindergarten and first grade children (that is, for children approximately between the ages of 3 and 7years old). This game has 27 different types of exercises situated in a circus environment; although only two exercises were described in this literature (Donker & Reitsma, 2004):

- Matching pictures that rhyme and
- Dragging a letter that will change the meaning of the written word into a trash can.

Milo and the magical stone (Fig 7.4): This is an adventurous game targeted at children aged 4 to 8years. In this game player (children) have to help Milo and his two mice friends find the magical stone which keeps the two mice warm during winter. The game has 10 sub games for which one of them is for them to catch a flies and feed a frog to move on (Baauw et al., 2006).

The Anti-Heuristic game (Fig 7.5) is based on a space invader time game designed for children aged between 7 and 9 years old. The game is designed for a laptop or PC platform where children will interact with the game using a mouse. It has a storyline that involves aliens invading the earth and Fred the farmer needs to defend his farm by defeating the aliens. In defeating the aliens, he saves the earth. A player has a 3 lives and the game has a total of four levels with the alien speed increasing with every increased level and the last level involves the shooting of the large spaceship. For a player to progress to the next level, they need to shoot 15 alien ships. Apart from shooting, cow parts are thrown to destroy alien ships, this is included to add humour to the game.

7.3.1.1.3 Procedure

On arrival at the study which took place at a play and art activities hall within the school premises, children were directed to sit at a table with a pair of children facing the other pair and the researcher sitting at one end across another table as shown in figure 7.6.

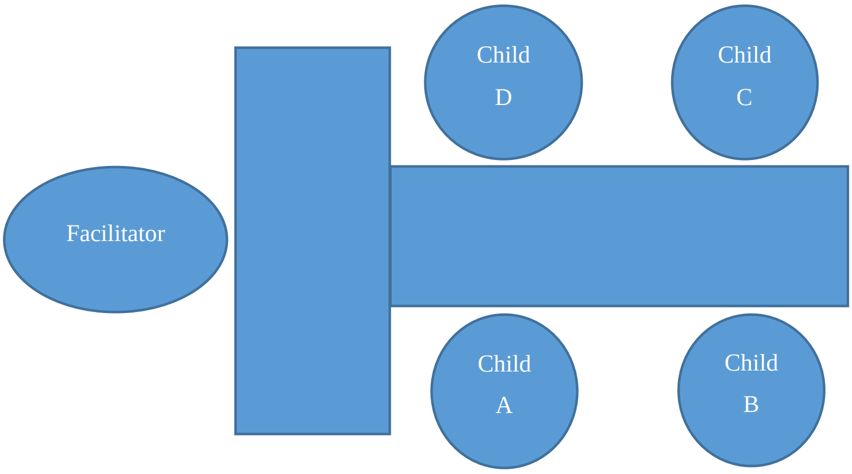

Figure 7.6 Study setting and seating position of participants

The researcher greeted the children, made a self-introduction and asked the children for their names (This was intended for the children to be relaxed and get acquainted with the researcher for a more familiar conversation), this lasted for 2 minutes. As stated earlier, activities were carried out in three sessions: The game criteria session, the drawing session and usability problem rating and colouring session. The researcher started by asking how many children love to play computer games, children answered with the raise of hands (in order to encourage participation and interaction). The researcher carried on by narrating a short story which led to the first question that started the game criteria session:

"I have a friend who works with computers and games. My friend is looking to design a good game for children like yourselves and younger children too but he is not sure about what will make a good game for you. He asked me but I wasn't sure myself so I told him I would come and ask you. So what do you think is the criteria for making a good game?"

As children responded and made points for what makes a good game, the researcher wrote it down. Children who were shy of speaking were individually asked on what makes a good game. This encouraged many quiet children to make input. After 10minutes, the researcher appraised the children by saying "well-done, thank you for these wonderful ideas, I never thought of most of these things”. Then the researcher carried on with the narration which led to the second session:

“After designing the game, my friend intends to test the game with some persons so he could make corrections on it before the final version is released. He believes these persons might find problems with the game and he would like them to rate their problems so he would know what problem to solve first. So he said I should ask you to draw pictures of problem situations, from a situation of ‘not a problem at all’ to a situation of ‘a very bad problem’. He said you could have as many steps as you want and you can draw any picture to represent these situations”.

While children drew pictures, the researcher acted as an observer, writing down things children said and wrote down verbalised statements and discussions made between children. The facilitator also looked at children’s drawings, in cases where a child’s drawing was complicated, the facilitator asked for clarity whereby this is noted in the observation sheet pre-labelled with the child’s group and identity. Children were encouraged to write notes for clarity of their drawing. At the end of 10minutes, the researcher appraised the children again commending their drawings and ideas and carried on the narration that led to the final session: *How many of you remember how the traffic light system work? How does it work?* As the children answered, they were commended then the researcher carried on narrating:

My friend also thinks the traffic light system could be used to rate problems. He gave me lists of some problems for you to colour based on your view of the problem. For example my friend said problems you think are not problems at all should be coloured in green, a medium problem in yellow, and a very bad problem in red, children were asked to echo this for confirmation and asked to indicate by the raise of hands if they had any questions. In cases where children made queries about the game or colouring process, it was explained again until the children indicated they have understood the game and process.

The facilitator handed in the first chat and the coloured pens and explained how the game on the chat works pointing to the problems below as problems that should be coloured. Children started colouring, when the children finished colouring on the first chat, the second game was explained for the second colouring and that was the order until all the games were explained and all the chats were coloured. At the end of the colouring session, children were thanked and asked to return to their classroom.

7.3.1.1.4 Analysis

7.3.1.1.4.1 Game Criteria

In order to investigate whether children have similar or the same values which was determined with the same criteria seen across groups, a merging technique was used to merge data (see appendix 3C). The merging was first done within group then between groups considering different instances: similarity in words, phrases or sentences used. For example criteria that are the same or similar within group were merged e.g.

- In group 5, the criteria "You can choose character that you like" and "You can change your player where some of them are boys and some of them are girls" were merged to be "You can choose character or player that you like where some are boys and some are girls".
- In group 4, the criterion "You learn to bake" was merged to the criterion "You can learn" to produce a merged criterion "You can learn e.g. you can learn to bake".

Criteria that are similar or the same between groups were also merged e.g.

- In groups 3, 4 and 6 it was stated that the game should be fun, these were merged to state criterion: It should be FUN.
- In group 3 a criterion "children can learn" was merged with the criterion "you can learn" from group 4.

In some cases, criteria that have similar or the same meaning were merged e.g. In group 1 a criterion "Moves are hard" was merged with the criteria "You go to new level and it gets harder", "Sometimes it goes easy and sometimes it goes hard" and "It's a challenge" from groups 4, 5 and 6 respectively.

In some other instance merged criteria from between groups are further merged to already merged criterion from between other groups e.g. The criterion "you learn to work your brain" from group 1 was merged to the criterion "it makes your brain healthier" from group 6 to produce the criterion "you learn to work your brain to make it healthier". This was further merged to the merged criterion "you can learn e.g. learn to bake" from group 3 and 4 to produce a bigger merged criterion "you can learn e.g. learn to bake, learn to work your brain to make it healthier".

After the criteria have been appropriately merged, themes were identified from the criterion and merged criteria. Seven themes were identified: Learning, Memorable, Fascinating, Challenging, Progress, Flexible, and Notification.

7.3.1.1.4.2 Severity Drawing

The aim of this pilot study was to investigate children's ability to design their own severity scale. Therefore a basic analysis of the children's drawing was carried out by the researcher to determine what and how children have presented their drawings. It was discovered that one child didn't make any drawing so this was excluded from the analysis while other children drew their severity pictures in steps (levels). The researcher determined how many steps each child had drawn, this was presented in columns and each child's identity was in rows. Each child's picture description was appropriately put against the child's identity and steps as presented by the child see table 7.1 for an example, full table is available in appendix 3G.

Table 7.1 Description of younger children's severity scale drawing

Groups	Identity	Step 1	Step 2	Step 3	Step 4	Step 5
1	A	A sun flower				
	B	A smiley face girl	A sad face girl			
	C	A smiley face girl	A sad face girl	An angry face girl		
	D	A sad face human	A Straight face human	A smiley face human		
2	A	A house with two windows	Another house with 4 windows and smoke coming out of the chimney			

In order to judge children's understanding of the severity rating concept from their drawings, the description of the pictures were read through to identify a theme for each picture. Each picture theme was scrutinised again, any picture theme that was reasonable to relate to severity situations when compared to existing scale from literature (e.g. Nielsen, 1995; Rubin & Chisnell, 2011; Yehuda & McGinn, 2007) were further considered a valid theme but any picture which cannot relate to a severity solution or which is incomprehensible was judged as a random picture and placed under the random theme. A total of 3 themes were produced: Faces (F), Persons (P) and Random (R) pictures.

The review of literature on severity rating scales and other rating scales for children show text is useful to describe each step of the scale. Therefore, the type of text some children provided for each step of their picture, was also analysed and compared within and between groups to ascertain similarities in the type of text produced. If text types are judged as similar or the same, it was merged e.g. in group 4 child D had text for his 2 steps drawing “happy problem” and “sad problem” and in group 5, child A had text for 2 steps drawing also “happy” and “sad”. Both were judged as similar and merged to be “happy problem” and “sad problem” see table 7.5 below.

7.3.1.1.4.3 Traffic Light Severity (TLS) Rating

Each child’s colouring sheet was read through and usability problems rated were compiled into a single sheet, identifying the severity judgement for each child. This is presented in a table (see table 7.2) which helped the researcher determine if children shared the same view on severity judgement and if they understood the use of traffic light as a severity scale, by comparing their severity judgement of each problems to how severe the problem was described in the literature from where the problems were collected.

The title of the game for which the problems emerged, the problems rated, the colours (red, yellow and green) and the groups were represented in the table. In cases where children coloured a problem with more than one colour, this was described as undecided in the same table.

During the analysis, the data of two children were removed from the analysis to avoid inconsistencies in the result, as these children were not present for the entire study, they either joined the study half way through or left before they completed the colouring task.

Table 7.2 Severity representation of problems rated using the traffic light system

Games	Problem	Traffic Light Colours				Groups (No. in group)
		Green	Yellow	Red	Undecided	
JamMo	If you do something wrong it does not tell you			4		1 (4)
				3	1	2 (4)

				4		3 (4)
				4		4 (4)
				2		5 (2)
				1		6 (1)
		-	**-**	**18**	**1**	**Total**
	Dragging items is a bit of a problem		4			1
			3		1	2
			4			3
			4			4
			2			5
				1		6
		-	**17**	**1**	**1**	**Total**

7.3.1.2 Result

7.3.1.2.1 Game Criteria

After merging the same and similar criteria within and between groups as reported in section 7.2.1.1.4 above, it was determined that children identified a total of 23 game criteria with 6 reported in more than one group. Where 1 from the 6 criteria (It's a challenge, moves are hard, it's not easy or it gets harder), is reported in 5 groups (see table 7.3 below), 1 (You can do some stuff, you can do what you want to do i.e. create, make and design e.g. you can change the wall paper) in 4 groups, 3 was reported in 3 groups and 1 in two groups. Each of the other 17 criteria were reported only in a group at a time (see table 7.3 below and full table in appendix 3D).

Table 7.3 Game Criteria and Groups

S/No	CRITERIA	Groups
1	Remind me of when I was small	1
2	Hide and catch people	1
3	I could dance	1
4	It's a challenge, moves are hard, it's not easy or it gets harder	1, 2, 4, 5, 6
5	Copy the tactics	1
6	Game can be on another device (Wii)	1
7	Beat People	1

8	You can learn e.g. Learn to bake and Learn how to work your brain to make your brain healthier	1, 4, 6

Result show that the seven themes derived from this merged criteria had at least 1 criterion and at most 6 criteria classified in them with “memorable” and “challenging” with the least criterion and “Fascinating” and “Flexible” with the highest criteria, see table 7.4 and 7.5 below. Result also show 12 criteria matches existing game heuristics from literature e.g. The criterion: “You get medals and hit high scores” matches the existing game criterion “Players are rewarded” by Federoff (2002), and Desurvire, Caplan, & Toth, (2004), also the criterion “It’s a challenge, moves are hard, it’s not easy or it gets harder” matches existing criteria “the game should be easy to learn but difficult to master” by (Alsumait & Al-Osaimi, 2009; Desurvire *et al*., 2004; Federoff, 2002) (see table 7.5 and in appendix 3E for full table).

Table 7.4 Themes and number of criterion for younger children’s game criteria

Themes	**Memorable**	**Fascinating**	**Challenging**	**Learning**	**Flexible**	**Progress**	**Notification**
No of Criteria	1	6	1	4	6	3	2

Table 7.5 Classification of Game Criteria into Themes and Matching Existing Heuristics/Guidelines

S/No	Themes	Younger Children's Game Criteria	Matching Existing Game Heuristics/Guidelines
1	Memorable	Remind me of when I was small	
2	Fascinating	Hide and catch people	
		I could dance	
		Action and shooting	
		A game with octopus is really interesting	The child is interested in the eLearning program characters because … (2) they are interesting to him, …(Alsumait & Al-Osaimi, 2009)
		It should be fun and funny	
		It has to make people like it and children will want to play with it	The game is enjoyable to replay (Desurvire et al., 2004)
3	Challenging	It's a challenge, moves are hard, it's not easy or it gets harder	A good game should be easy to learn and hard to master (Alsumait & Al-Osaimi, 2009; Desurvire et al., 2004; Federoff, 2002)
4	Learning	Copy the tactics	
		You can learn e.g. Learn how to work your brain to make your brain healthier, you can learn to bake	One reward of playing should be the acquisition of skill (Federoff, 2002) The program supports the child's cognitive curiosity through surprises, paradoxes, humour, and dealing with topics that already interest the child. (Alsumait & Al-Osaimi, 2009)

7.3.1.2.2 Severity Drawing

Result show 3 children had presented their picture in one step, 7 children presented theirs in two steps, 5 children presented it in three steps, 2 presented theirs in four steps and just 1 child presented his picture in five steps (see table 7.6 for a summary and full details in appendix 3G). Result also show only 5 children provided text for each step of their picture and after analysis only 2 text type (one from group 4 and the other from group 5) were similar, therefore was merged leaving a total of 4 text type, see table 7.6 below.

Table 7.6 Summary of younger children's severity drawings

Groups	**1**				**2**			**3**				**4**				**5**		**6**
Identity	**A**	**B**	**C**	**D**	**A**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**C**	**A**
Steps	1	2	3	3	2	1	1	3	2	2	5	4	4	3	2	2	2	3
Picture Theme	R	F	F	F	R	R	R	R	P	P	F	P	P	F	F	F	F	R
Text Provided	N	N	N	N	N	N	N	N	N	N	N	Y	Y	Y	Y	Y	N	N

KEY for Table 7.6 (Picture Theme: R – Random, F – Faces, P – Persons; Text Provided: Y – Yes, N – No)

Table 7.7 Children's text for their severity drawing

identity	**Text**
4A	Not a problem, A bit of problem, A bit more of a problem, A big problem
4B	Happy, Worried, Not good, Very bad
4C	Not a problem, A bit of a problem, A big bad problem
4D, 5A	Happy problem, Sad problem

TLS Rating

Having had the children rate the problems gathered from literature, result show the children shared similar opinion in their judgement of most problems. For example

the problem “If you do something wrong it does not tell you” from the JamMo game was rated as a red problem (a very bad problem) by 18 children out of 19 analysed responses and the 19^{th} child was undecided with the choice of a red and yellow.

Also the problem “The process of restarting the game is long” in the Anti-heuristics space invader game was judged to be a red problem by 15 out of 19 raters. 3 other child raters think it was a yellow problem and 1 thinks it was a green problem.

The problem “It is impossible to follow any tactic when trying to click on the crabs” from the Milo game was rated by 15 children as a red problem, 3 children rated it as a yellow problem and 2 children rate it as a green problem.

However, there are other problems where children’s judgement was spread out almost equally between two levels. For example, the problem “It is difficult to drag items with the mouse” from the LeesCircus game was rated as a red problem by 10 children and as a yellow problem by 9 other children. Also the problem “Explanation was not enough” from the Milo game was rated by 9 children as a red problem and 10 children as a yellow problem (see table 7.8 for summary and appendix 3 for full details).

Table 7.8 Summary of usability problems rated using the traffic light colours

Games	Usability Problems	Traffic Light Colours			Undecided
		Green	Yellow	Red	
JamMo	If you do something wrong it does not tell you	-	-	18	1
	Dragging items is a bit of a problem	-	17	1	1
LeesCircus	Volume on the laptop was too low	1	4	13	1
	It is difficult to drag items with the mouse	-	9	10	-
	Could not click on pictures	2	-	16	1
Milo	Instructions on how to play each sub game were too long	4	3	12	-
	Explanation was not enough	-	10	9	-
	Did not understand how to catch the flies	4	3	12	-
	Did not know how to give the fly to the toad	4	6	6	3
	It is impossible to follow any tactic when trying to click on the crabs	2	3	15	-
Anti-heuristics space-invader	The process of restarting the game is long	1	3	15	-
	Screen turns black with inappropriate feedback	2	5	13	-

7.3.1.3 Discussion/Conclusion

This experiment showed that the story was useful to start an engaging discourse, uphold the engaging scenery as children kept asking questions such as: "How quickly will the researcher's friend (the subject in the story) get the input to create the game?", "Does the friend design games for adults too?". This showed children understood and believed the story following views from literature that the story needs to make sense (Boyce, 1996) and believable (Chaitin, 2003; Garrett, 2006). They were also willing to make input and carry on to the end.

The study also proved that children could come up with useful game criteria that matches existing well referenced game criteria as seen in section 7.3.1.2 table 7.5 above, although some criteria were completely childish e.g. "I could dance", however, this might not be completely ruled out as it shows, a mathematical game that will allow the player dance will be more appealing to this child (ren).

They were also able to rate problems using the traffic light system, despite technically having only two severity scale (yellow and red) as they were asked to use the "green" as "not a problem" and lastly were able to show that they understood the concept of severity with their drawing of a severity scale. Though some children drew random pictures which is not in context of a severity scale e.g. child 2D who drew a single house without any text and child 1A drawing just a sunflower.

Although, the children used for this study are year 2 children aged 6 and 7year olds and according to cognitive development literature (e.g. Piaget, (1971) see literature review in chapter 2), this age group of children are still developing in their thoughts and vocabularies, they are animistic (believing everything have some kind of consciousness (Piaget, 1971)) and finds it difficult to logically work with abstract object. However, basing the process on their values (what they think is important, their judgement) made it much easier for them to make input, this supports Piaget's view that children at this stage are egocentric (i.e. they can only consider things from their own view point). Therefore, it is conjectured that older children will have better understanding.

7.3.2 Study 2 (Older Children)

7.3.2.1 Method

This study will use the same method, techniques and tools reported above to test part of the new method with older children.

7.3.2.1.1 Participants

A total of 24 children aged 9 to 11 years old from years 5 and 6 participated in this study. All the children were from the same UK primary school. The children were called out of the classroom by their teacher and put in groups. A total of six groups emerged where each group had 4 children. Also for this study, the author was the only researcher present for the entire study.

7.3.2.1.2 Apparatus

The researcher carried out the study using the same apparatus that was used in study 1, which is reported in section 7.3.1.1.2 above only in this study the purpose of the traffic light colours was changed, (see procedure section below).

7.3.2.1.3 Procedure

The study was carried out first in an activity hall with the first three groups and the last three groups carried out the study in the school’s library. The Children were called out of their classes by the researcher in groups of four where the year 5 pupils were first called out to participate before the year 6 pupils. This was the case particularly for order and availability of the children.

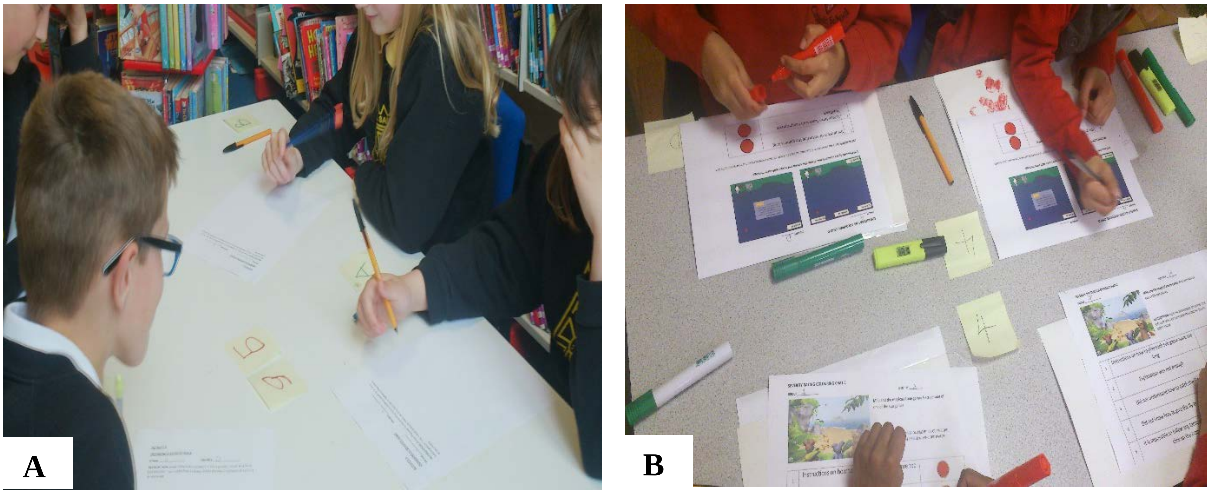

Figure 7.7 Design Session with Older Children A: Children drawing severity scale B: Judging the Severity of Problems

The entire study procedure ran exactly the same as that reported in section 7.3.1.1.3 above, unless that during the severity rating colouring session using the traffic light system, children were asked to colour problems that are little in green, medium problems in yellow and serious problems in red e.g. green: a problem, yellow: a bad problem, red: a very bad problem.

7.3.2.1.4 Analysis

Analysis was carried out by the researcher, who did the analysis separately in order of the different activities. This section will describe and report the analysis by the activities: Game criteria, the traffic light problem rating activity and severity rating pictures (drawings).

7.3.2.1.4.1 Analysing the Game Criteria Data

The analysis of children's game criteria was done in five stages using tables and text:

a. Organising criteria according to children's groups
b. Identifying categories using inductive approach (from the children's game criteria data only) and placing criteria in appropriate categories
c. Eliminating duplicate criterion using merging technique and adding up frequencies
d. Matching criteria to similar existing heuristics/guidelines

a) Organising criteria according to children's group: Children's criteria were organised in table according to their groups (see table 7.9 below for extract of data, full table is available in appendix 3). Criteria were put in rows while the group from which each criteria emanated were in columns. Frequency of children that stated each criterion were also identified and placed in bracket next to the criterion.

Table 7.9 Children's criteria for a good game compared to existing heuristics

S/No	**GAME CRITERIA OF YEARS 5 AND 6 CHILDREN BY THEIR GROUPS**					
	GROUP 1	**GROUP 2**	**GROUP 3**	**GROUP 4**	**GROUP 5**	**GROUP 6**
1.	It should be fun learning not just learning – for a learning game **F1 & L/M/C1**	It should be adventurous **CH2**	Some questions in it like Maths puzzles and stuff **CH3**	Fun like Fun Maths **F4 & L/M/C4**	It should be flexible to be on every game console **F/A/I5**	It has to be an action game **CH6**

b) Identifying categories using inductive approach and placing criteria in appropriate categories: In this stage, researcher identified categories (themes) from children's criteria, and then each criterion was coded with an abbreviation of a suitable category and the number of the group the criterion was stated. For example, in group 2, the criterion "It should be adventurous" was judged to fit the category "challenge" therefore, the abbreviation "CH" was placed in front of the criterion and the number "2" next to CH to indicate category and group. Children's criteria were further put into categories in another table showing groups and frequencies, allowing repetition (the same or similar criterion from within or between groups) for the purpose of data transparency.

Categories were decided by the researcher in view of what the criterion or criteria addresses. For example, any criterion that is associated with difficulty of the game or challenge to the player inspired the category "challenge" e.g. "There should be different levels of difficulty" and "At the end of every 10 levels you should fight a boss". Also any criteria that are associated with game story inspired the category story, e.g. "It should have a good storyline and a plot to the story" and "The game should have a storyline".

A total of 11 categories were identified from the criteria (see extract of data in table 7.10 below, full table is available in appendix 3)

Table 7.10 Criteria Categories and Group

S/No	CATEGORISED CRITERIA	GROUPS
1	**REWARD**	
	When you get it right, you should get reward or coins, you should be able to use the coins or reward to go to the next phase and buy stuff for your character R1	**1**
	You should get rewards for completing a level R3	**3**
	There should be health package for power upgrade R5	**5**
	You should have rewards (coins) for doing things R2	**2**
	You should be able to use your rewards to buy stuff to continue the game R2	**2**
	You get a reward for every correct puzzle you solve in a puzzle game R6	**6**
	You should be able to find things that will give you power up R3	**3**
	You should be able to collect coins or get stars when you are doing good R3	**3**

Key for R# (R – Reward, # - Group number criteria was identified)

c) Eliminating duplicate criterion using merging technique: In this stage, each criterion was linked to the exact group (s). A criterion that is repeated within group were merged to produce only one of such criterion. On the other hand, criterion that is repeated between groups is merged. However, all the groups are identified and written down, see extract in table 7.11 below and full table in appendix 3.

Table 7.11 Identifying exact group for each criterion

S/No	CATEGORISED CRITERIA	GROUPS
1	**REWARD**	
	When you do things or get it right or for correct puzzles or for completing a level, you should get reward or coins or stars, you should be able to use the coins or reward to go to the next phase and buy stuff for your character to continue game	**1, 2, 3, 6**
	There should be health package for power upgrade	**3, 5**

d) Matching criteria in categories to similar existing heuristics or guidelines: Children's unique criterion in their categories was matched to existing game heuristics/design guidelines. Each existing heuristic/guideline was carried verbatim to match with children's criterion, which sometimes are not the same in words used but similar (i.e. matching is determined by the similarities in meaning of phrases or sentences), see table 7.12 below. In cases where the same heuristic is stated by different authors, it was represented as one, though all the authors who stated that heuristic were identified and the sub category for which they have placed that particular heuristic in their work was also represented. For example, the associated existing heuristic "Players are rewarded" was stated by Federoff and Korhonen and Koivisto, this has been categorised as GP3 in Federoff's work and as Game Play in Korhonen and Koivisto's work. All of these were represented during the criterion marching stage. See table 7.12 extract below and full table in appendix 3.

Table 7.12 Matching criterion to existing heuristics/guidelines

S/No	CATEGORISED CRITERIA	ASSOCIATED EXISTING HEURISTIC / GUIDELINE	TYPE OF HEURISTIC / GUIDELINE	REFERENCE
1	**REWARD** When you do things or get it right or for correct puzzles or for completing a level, you should get reward or coins or stars, you should be able to use the coins or reward to go to the next phase and buy stuff for your character to continue game	**GP3 / GAME PLAY** Players are rewarded **GP3 / EUH 3** Rewards are meaningful **GAME PLAY** The game should give rewards that immerse the player more deeply in the game by increasing their capabilities and expanding their abilities to customize	ELearning; Game Play; and Playability Heuristics	(Alsumait & Al-Osaimi, 2009; Federoff, 2002; Korhonen & Koivisto, 2006); (Desurvire et al., 2004)

7.3.2.1.4.2 Analysing the Usability Problem Data Rated using the TLS

Children's rating of usability problems using the traffic light system was analysed in two ways: first, data was first coded where the colours of the traffic light were coded using simple numbers 1, 2 and 3 where colour green is 1, yellow - 2 and red - 3. While the twelve usability problems were coded using UP (which represents the first letter of the two words **U**sability **P**roblem) and further attached to numbers 1 to 12. Therefore, the first usability problem was coded as UP1 and the last usability problem was coded as UP12.

Secondly, in order to test for consistencies in raters' severity judgement, the coded data and children's identities were further put into the SPSS software to test for inter-rater reliability using Cronbach's Alpha reliability test as literature has proven that this is appropriate for testing reliability between multiple raters.

Within SPSS, children's identity were aligned on the column axis and the usability problems on the row axis, then a Cronbach's alpha reliability test was carried out. During this process, the inter-rater correlation was tested and the mean score and standard deviation were also derived.

Analysing Children's Drawings (Severity Pictures)

Children's drawings were analysed by the author in two ways: first, the actual pictures were hand coded following literature to judge the evidence of severity (this is reported in the sub sections below). Secondly, general aspects of each drawing e.g. the number of levels of picture presented, the labels (text) of the pictures, the type of pictures drawn and description of each picture were also analysed.

A. Hand Coding the Severity Drawing

Review of literature shows children's drawing could be coded using different techniques. Xu et al carried out a coding session adopting an approach from literature and also considering the goal of the drawing session (Xu *et al*., 2009). Read et al equally coded children's drawing based on the goal of the study and on set guidelines from literature (Read *et al*., 2013). Codes for this study were derived based on:

- The review of related literature on severity rating scales and other rating scales in HCI and CCI literature,
- The aim of the drawing session: *"can children draw pictures to represent a severity rating scale from the point of not a problem to a very bad problem"* and
- The investigation of the children's drawing.

i. Defining Code Elements

A typical severity rating scale consist of levels with each described by text and/or numeric value (Hocko, 2002). However, Scott suggests that for children under 11, the use of visual stimuli are useful in making a concept in question more concrete than verbal representation alone (Scott, 2000). The aim of this study is to investigate whether children understand severity and can draw pictures to represent a severity scale. Children were asked to draw pictures that represent a situation from 'not a problem at all' to 'a very bad problem'. In view of this, each picture will be coded thus:

1. *Picture that was drawn in levels will have 1 point.*

In line with typical severity scale (as stated above) and as highlighted by Read *et al*. it is advised that the response options of visual analogue scales or multi-choice responses used with children, should be completely labelled to help children produce more reliable responses (Read & Fine, 2005; Read & MacFarlane, 2006). In addition, Barendregt *et al*., (2007) reports that multiple resources (for example, audio and visual though in this case visual and textual) when requiring information from children may make it easier for children to understand the explanation (Barendregt *et al*., 2007). This is also highlighted by Scott (2000) as stated above and supported by Kano *et al*. (2010)

2. *So if a picture go further to have textual description of each level of their picture, then it will have another 1 point.*

Read et al.'s work, aimed to design a tool that will empirically evaluate or measure fun (which is the goal with fun as the main interest). So they targeted the design of their tool to elicit the amount of fun (main interest) that the child experienced having participated in an event (Read *et al.*, 2002). Xu *et al* carried out a coding session adopting an approach from literature and also considering the goal of the drawing session (Xu *et al*., 2009). Read *et al., (2013)* equally coded children's drawing based on the goal of the study and on set guidelines from literature (Read *et al*., 2013).

In this study the goal is to have children's pictures that depict severe instances either in descending or ascending order and that can meaningfully measure the severity of a usability problem. Since severity scales are intended to rate the severity of usability problems, such that resources would be allocated appropriately where the most resources will be allocated to fix the most problems (Molich *et al.*, 2013; Nielsen, 1995). It is expected that each level or step of children's picture is distinct from others and can clearly indicate the severity of usability problem (s) put in that level.

3. *Therefore, if the picture represented in levels is appropriate to judge the severity of usability problem (either the levels are going down to an instance where a usability problem is 'not a problem at all' or going up to when a usability problem is 'a very bad problem'), then it gets an additional 1 point.*

Rating scales in HCI use consistent measures. For example if the first scale is a figure then the rest of the scale will use figures (Dumas & Redish, 1999; Nielsen, 1993a; Rubin & Chisnell, 2011), if the first is a star then the rest will be stars (Yehuda & McGinn, 2007). This is also the case in the CCI community. Where if the first picture is a face as in the Fun toolkit (Read *et al.,* 2002) then all the other level pictures will use faces or if it is a thumb as in the Thumbs Up Scale (Kano *et al*., 2010) then others will be a thumb.

4. *Type of picture*
    a. *If a picture is consistent from the first to the last, then that picture gets a point*

OR

As in the PIPC method (Barendregt *et al*., 2007), pictures were used to represent usability problems. Pictures selected for a usability problem usually reflects the concept of the problem being described. For example, a difficult problem was represented with a picture that has a confused face with a question mark next to the face on the picture. But the next picture might not be the same picture category as the difficult picture but represent another problem concept. For example, a picture that has a baby on it represent the concept of the game being childish.

    b. *If a picture is able to represent the concept of a usability problem situation then it gets a point.*

ii. Actual Coding Process

Four elements have been identified above, these were considered and coded in each drawing for the first phase using binary numbers 1 and 0, following the analysis of children's drawing carried out by Read *et al*., (2013). Where an element was judged as present, it was coded as 1, in the absence of an element it was coded as 0. Therefore, a total score of 4 was expected for a picture to be termed as having a strong evidence of severity.

Categorising Coded Pictures According to Scores

At the end children's ability to understand the concept of severity scale and come up with their own scale was determined by the total score of the picture. For example as stated above, pictures that scored 4 were classified as having strong evidence of severity, pictures that scored 3 were considered as having evidence of severity, pictures that scored 2 as having moderate evidence of severity, pictures that scored 1 had weak evidence of severity and pictures that scored 0 were considered as having no evidence of severity.

B. Analysing the General Aspects of the Drawing

Aside coding the drawings, the researcher decided to analyse different aspects of the children's drawing which was carried out in five stages as detailed below:

i. Description of each drawing
ii. Identifying thematic frequencies according to the type of drawings
iii. Analysing the type of text for each level
iv. Analysing each text, stating text frequency

v. Number of levels and picture frequency

Description of each drawing**:** Children’s drawing was read through and it was identified that children have represented their pictures in levels with the highest level as 10. The researcher therefore placed children’s identity in rows and placed the levels (from 1 to 10) in columns. Each child’s drawing was therefore described with words in the appropriate level. See table 7.13 for extract and full table in appendix 3

Table 7.13 Children’s drawing described

CHILD’S IDENTITY	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	LEVEL 5	LEVEL 6
1A	A big smile with a tooth out on a square face	An average smile on a square face	A small smile on a square face	A smaller smile on a square face	A straight square face	A square face with little sad expression
1B	Four pencils but only one will be circled	Four pencils but only two will be circled	Four Pencils and three will be circled	Four pencils, All four will be circled		

Identifying thematic frequencies for the drawings: In order to analyse the consistencies and similarities in children’s drawing decision, themes were identified from the drawings and each picture was categorised into each theme. The researcher further made a count of drawings categorised into each theme to decide the thematic frequencies. See table 7.19 below for picture themes and frequencies.

Type of text for each level: Since children’s picture levels have been identified with the highest level as 10, the researcher used the same technique used in stage one (Description of children’s drawing) by placing identity in rows and picture levels (1 to 10) in column. The text each child labelled his/her drawing was therefore written down in the appropriate level as the child had done it in their drawing. See table 7.14 for how this has been analysed and represented by the researcher.

Table 7.14 An example of text used to describe each level picture

IDENTITY	**LEVEL 1**	**LEVEL 2**	**LEVEL 3**	**LEVEL 4**	**LEVEL 5**	**LEVEL 6**	**Level 7**	**Level 8**
4C	**Very good**	**Good**	**Ok**	**Bad**	**Very Bad**			

Determining the frequency of the text type: In this stage, children's drawings were scanned through again to identify unique text type. In cases where two children had labelled their drawing with the same type of text, this was merged and represented as one. However, the number of children who have used this type of text was identified in the frequency column and a total of 17 text type was identified. See table 7.19 below for detailed analysis.

Number of levels and picture frequency: In order to understand how many level points or steps children have shown in their drawings, the number of points in each drawing was noted. To identify unique numbers, repetition of numbers were eliminated which resulted in six unique numbers (10, 5, 4, 3, 2, and 1). These were recorded in a column and the frequency of pictures with these number of points were appropriately aligned in another column. See table 7.15 below for schematic description.

Table 7.15 Number of levels used and number of pictures identified for each level

Levels	1	2	3	4	5	10
Number of Pictures	1	3	12	5	2	1

7.3.2.2 Result

7.3.2.2.1 Game Criteria

Children produced a total of 85 game criteria between groups. After first stage data analysis, these 85 criteria were categorised into 11 themes e.g. Reward, Challenge, Age Appropriate, Story, Learnability and Memorability (see appendix 3K for details of themes) with each theme having at least 3 criteria classified in it. After redundancies were removed in each category, data was reduced with themes having at least one criteria classified in it. For example the themes "Story" and "Match to real world" had "2" criteria each, classified into them but after similarities in criteria were merged, it left each theme with just one criterion see table 7.17 below. In addition the "Reward" and "Challenge" themes which had "9" and "21" respectively were reduced to "2" criteria in "reward" and "5" criteria in "challenge".

Table 7.16 Result of merging game criteria

BEFORE MERGING CRITERIA	
Themes and Criteria	**Groups**
Story	
It should have a good storyline and a plot to the storyline	**2**
The game should have a storyline - The last level should be able to end the story where the player battles someone to end	**3**
Match to Real World	
Content of the game should be related to real world e.g. people	**6**
You should be able to compare what is in the game to what is in real world	**3**
AFTER MERGING CRITERIA	
Story It should have a good storyline and a plot to the storyline. The last level should be able to end the story where the player battles someone to end	**2, 3**
Match to Real World Content of the game should be related to real world i.e. it can be compared to real world e.g. people	**3, 6**

A total of 34 game criteria was identified after the merging phase. These 34 criteria from the children were compared to existing heuristics and 29 matched the existing heuristics or guidelines (see appendix 3 for full detail).

It was observed that most criteria that had high frequency matched one or more existing heuristics or guidelines. For example the criterion

"It should be fun learning not just learning – for a learning game e.g. fun like fun Maths" with frequency 12 matched the ELearning heuristics "**EUH 3 MOTIVATION TO LEARN** The e-learning program is enjoyable and interesting" by Alsumait & Al-Osaimi (2009).

Also the criterion "When you do things or get it right or for correct puzzles or for completing a level you should get reward or coins or stars and you should be able to use the coins or reward to go to the next phase and buy stuff for your character to continue game" with frequency 14 matched the heuristics "**GAMEPLAY** Meaningful reward should be given to players. These rewards should immerse the player more deeply in the game by increasing their capabilities and expanding their abilities to customize" by Alsumait & Al-Osaimi (2009); Federoff (2002); Korhonen & Koivisto (2006); and Desurvire *et al*. (2004)

On the other hand, some criteria which had low frequency matched more than one existing heuristics/guideline. For example the criterion

"There should be continuity and connection between games" matched the existing heuristics/guideline/global question "**CONTINUITY** Game does not stagnate. It should provide an asynchronous persistence game world and mechanics that allow the player to feel progress. It should react in a consistent, challenging and exciting way to the player's actions and let the flow of the game meet the expectation" by Alsumait & Al-Osaimi, (2009); Baauw *et al*., (2005); Desurvire *et al*. (2004); Korhonen & Koivisto (2006); and Paavilainen (2010). See appendix 3L for full result.

7.3.2.2.2 TLS Rating Result

The Cronbach's alpha coefficient for the twenty four child raters resulted in a high coefficient of .867, suggesting that the items have relatively high internal consistency. It was also determined that all twenty four raters' judgement does not deviate very much from this over all coefficient (.867). The reliability coefficient score is therefore acceptable given that it is higher than .70.

7.3.2.2.3 Severity Drawing (Pictures) Result

Since children's drawing was analysed following different format and stages, result will be reported in the appropriate forms.

A. Hand Coded Result

Having hand coded the children's drawing, a mean score of 3.4 was derived. After categorising each child's picture score, it showed all the children's drawing had some evidence of severity, given that all the pictures had a score though some had weak evidence of severity, some had moderate evidence but some others had strong evidence of severity (see table 7.17).

Table 7.17 Result of categorising children's severity drawing

Scores	Description of scores	Number of pictures for each level
0	No evidence of severity	**-**
1	Weak evidence of severity	**2**
2	Moderate evidence	**3**
3	Evidence of severity	**3**
4	Strong evidence of severity	**16**

Pictures that had 4 points were classified as having strong evidence of severity which resulted in 16 pictures with strong evidence of severity. 3 pictures were classified as having evidence of severity with 3 points each. 3 other pictures were also classified under moderate evidence of severity with each having scored 2 points and finally 2 pictures which scored 1 point each were classified under the weak evidence of severity with no picture in the category of no evidence of severity.

B. Results on the General Aspect of the Children's Drawings

i. Description of each drawing

After analysing the general aspect of the children's drawing, it is evident that children used different things to represent severity for example they used faces, objects, animals. See table 7.13 above for specific example and table 7.18 for thematic description of things represented in children's drawings.

Table 7.18 Thematic Description of children's drawing and frequency

Picture Type	**Frequency**
A smiley face	11
A stick person	5
Pencil	1
Moon, Earth,	1
Animals	3
Random Pictures	3

ii. Identifying thematic frequencies according to the type of drawings

In identifying themes for children's drawings, a total of 6 themes emerged with each theme having at least one picture categorised into it. The theme with the highest frequency of pictures was the smiley face theme with a total of 11 pictures, while the themes with the lowest frequency of picture are pencil and moon-earth with each having just one picture categorised in them. See table 7.18 above for full details

iii. Result on the type of text and frequencies

Analysis show children used different text for labelling their drawings. Sometimes children used only words or only number and some other times, words and numbers were used. However, a total of seventeen unique text labels were identified though some text labels had been used by more than one child but some others had been used by only one child. For example the labels easy medium and hard was used by 4 children, very good and very bad were used by 2 children while all the others by used by only one child. See table 7.19 for full result.

iv. Number of levels and picture frequency

As children represented their drawings in levels, they came up with different number of levels with 10 as the highest level and 1 as the least number of level. It was also determined that 12 children had represented their severity in 3 levels, which is the level with the highest frequency and 10 and 1 levels with the lowest frequency with each just having one picture categorised into them.

Table 7.19 Picture text type and frequencies

S/No	**Level Texts**	**Frequency of Pictures**
1	Easy, Fairly Easy, Ok Hard	1
2	Good, Ok, Bad	1
3	Very Easy, Not too hard, Hard, Too hard	1
4	Easy, Medium, Hard	4
5	Happy, Not happy, Hard	1
6	Easy, Medium, Hard, Very Hard	1
7	Easy, Medium, Hard, Challenge, Extreme	1
8	A bit of a problem, A bit more of a problem, A little bad Problem, A bad problem	1
9	Very good, Very bad	2
10	Very good, good, ok, bad, very bad	1
11	Very good, Ok, Very Bad	1
12	Easy, Hard, Really Hard	1
13	Small, Big, Bigger, Massive	1
14	Happy, Medium, Sad	1
15	Not a problem, Very bad Problem	1
16	Happy, Not happy, Sad	1
17	1,2,3,4,5,6,7,8,9,10	1

7.3.2.3 Discussion/Conclusion (PRODUCE METHOD V2)

Storytelling was very useful to start the study process and children having to make input based on their values made it easy for them to contribute. However, it was discovered from the result that older children made more input on the game criteria than younger children did. It was also evident that younger children were usually more specific when mentioning game criteria than the older children. For example, younger children stated criteria like “I could dance” and “It should remind me of when I was small” but older children are able to state criteria that is generic and also specific for younger children e.g. “For the little ones, they should be able to learn something for when they are a bit older in an educational game”. It was also observed from the result that most of the older children’s criteria matched existing heuristics and guidelines, as opposed to those of the younger children.

In view of the severity drawing result, though both age ranges of children were able to draw for this session, the older children’s drawing were more contextualised than younger children’s drawing. Although both age groups of children used more of smiley faces to represent their severity scale. More of the older children provided text for their drawing than the younger ones though there were lots of disagreement in the text description with only few text descriptions appearing as the same.

Since the younger children used a different rating when using the traffic light system than that used by the older children, that is, the older children had all three colours as problem levels while the younger children had one as no problem and 2 as problem levels. It is difficult to make comparison on the outcome. Though result for each age group show agreement in judgment within each group.

It could therefore be finalised that children understand the concept of severity rating scale, and are able to come up with their own severity scale and criteria to make a good game. The researcher further implemented the outcome of the experiment in the design of IMCH version 2, where children will come up with their own criteria and then use the traffic light colours to rate the problems they will encounter during the evaluation.

IMCH V2

*The **I**nspection **M**ethod for **CH**ildren (IMCH) is to allow children produce the criteria for a good game based on their values and further use the stated criteria to find usability problems and rate the problems using the traffic light system.* The following is the procedure to perform the evaluation using the IMCH V2:

- Facilitator explains the method and what is required of the children (evaluators) during the entire process.

- The story will be narrated then, children will discuss amongst themselves and come up with game criteria for a game in the same genre as the one to be evaluated.
- Then children will individually evaluate a game, find and record usability problems they will encounter, which they will write on a post-it-note (one problem for one sheet) and colour the post it sheet with a colour to indicate the problem severity using the smiley face traffic light severity scale as a guide.
- Facilitator explains the merging phase.
- Thereafter, they need to orderly post similar problems together (merging them) until all their problems are merged.
- Finally, people who have merged problems together will agree on a final severity for their merged problem. During the final severity stage the severity traffic light colours/ facial scale (see fig 7.8) is placed in front of the children to aid their decision making.

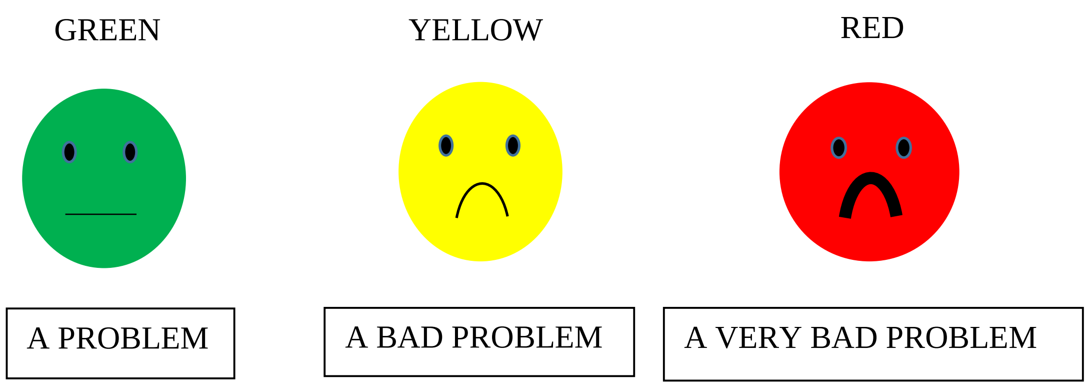

Figure 7.8 Traffic light facial severity scale

According to teacher's suggestions from the focus group (see section), the UK traffic light system could be used as a severity rating scale and the standard number and colours for a UK traffic light are 3: green, red and yellow respectively. Since the essence of a severity scale is to indicate the severity of problems found (Molich et al., 2013; Nielsen, 1995), and the least number of severity scale recorded in literature is 3 (see Salian et al., 2013; Kishian Salian, 2012). Therefore, all the colours in the traffic light severity scale will be used as a step to indicate problem severity. However, since in the real setting "green" colour means "go" and could be counter intuitive to represent problem severity, smiley faces have been added to each colour of the scale following children's drawing of severity rating scale where smiley faces were the most frequently drawn picture having the highest frequency of 11 (see table 7.18).

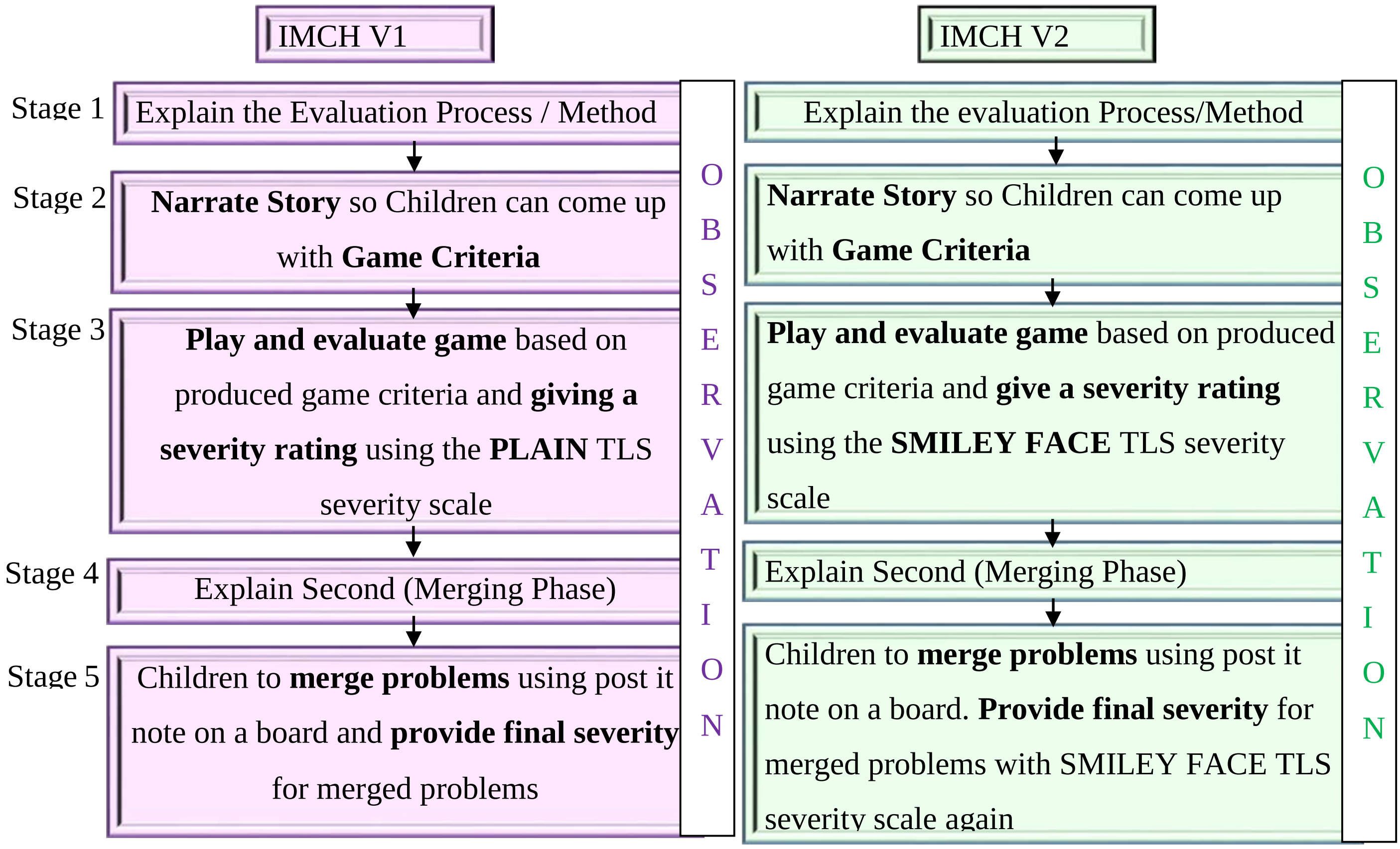

Figure 7.9 Comparing Versions 1 and 2 of the IMCH method

The difference in the V1 and V2 of the IMCH method can be seen in stage 3, where a plain traffic light severity rating scale is used in version 1 while a smiey face traffic light severity rating scale is used for the second version.

7.4 PART C Input from Independent Teacher on IMCH V2

Having come up with the version 2 of the IMCH, it was decided that the method be run by an independent teacher (a teacher who wasn't part of the focus group study reported in chapter 6), someone who has experience in developing learning materials for children and who has experience on how to engage children in a task that will be fun, suitable and engaging.

The aims of this discussion was to:

- Confirm that the tools (e.g. severity scale, post it note) that will be used for the evaluation are usable by children
- Review of the method process to ascertain whether it will be engaging for children
- Get input to improve any area of the method

An independent teacher who taught years 4 to 6 for over 3years was recruited for the discussion and review of IMCH V2. The main researcher acted as the facilitator for the session and a second researcher acted as an observer. This took place in an office space, a convenient environment for discussion and where tools could be properly laid out and demonstrations could be made.

7.4.1 Process

To start the discussion, the participants made a self-introduction and the researcher went further to introduce the discussion, stating what it was about. The IMCH V2 was explained, in this case the researcher ran through the method as it would work in a typical study with the child evaluators. All the tools needed for the typical study was presented e.g. the technology, the intended severity scale, the post it notes, markers and pens.

After verbal explanation of the method, the researcher continued by questioning the independent teacher as follows:

1. How should I present the method to the children that would be easy for them to follow?

Teacher's Response: With verbal explanation, children will be able to follow. However, a "show and tell" will be useful for visual aid to improve their understanding of the method procedure.

2. Children need to come up with game criteria in the same genre as the one to be evaluated, how could all the children be involved in the discussion?

Teacher's Response: When involving the children in a discussion, ensure they are relax with what you are asking them to do. You could allow them to start with the game, after a while retrieve the game from them and ask them to discuss what they think is wrong with the game or what could be added to the game to make it better.

To get everyone to discuss, mention it to them that they all need to make input, if after a while some people are still not saying anything, you could politely pick on them to say something.

3. How should a repeated criterion be sieved out without offending the children?

Teacher's Response: Draw their attention to the similar criterion that is already on the board and encourage that particular child to think of something else since what they mentioned has already been said. This way the child won't feel left out.

4. What about a criterion that is not in context with what is being discussed?

Teacher's Response: Say it out again to the hearing of everyone and repeat the question to the child, you could explain further to help the child understand better and encourage them to think about something in context. This should be said politely but remember to appreciate their effort because children love to be appreciated for efforts they make.

5. Children are expected to write down problems upon finding it, how would they be reminded without distracting them from the evaluation?

Teacher's Response: While they are doing the evaluation, you could say do not forget to write down whatever problem you encounter. This could be repeated reasonably without disrupting their work.

6. After finding their individual problem, they would be asked to merge their problems by each person posting a unique problem on the board so people with similar or same problems could go post theirs on. How should this process be facilitated such that everyone would be involved both the reserved and the extrovert children?

Teacher's Response: Every child should be given the opportunity to post their problem on the board. You could start by telling them what they need to do on the board, then ask them who would like to go first, if no one indicates interest then you could point at someone to go first. In the instance where a child is shy to go to the board, you could allow the extrovert children go first, it is possible that when the extrovert children finish posting all their problems, all the problems in the hands of the reserved child would have been posted so they do not have to go in front. You could also make them go first next time so they don't stay long in front of everyone.

7. Finally the children are expected to agree on severity colour using the traffic light severity scale (this was shown to the teacher) with facial expressions as shown in figure 7.7 above. What do you think about the facial expression being included as part of the scale together with the traffic light colours?

Teacher's Response: I wouldn't say yes or no but I think it shouldn't be a problem if you tell them beforehand what each step should represent, since you have text under each coloured face to describe it, then that should be sufficient.

8. Do you have any other concerns in the method or is there any additions you would like to make that will boost the method suitability for children?

Teacher's Response: It would be useful to have the scale put in front of the children for the whole time, let them start with the game and do the show and tell.

After the discussion, the researcher gathered all the document and said thanks to the teacher noting the input that have been made.

7.4.2 Conclusion (Produced Method V3)

Since this was only to run the method by an independent teacher (neutral but one experienced with working with children), and an informal discussion, the researcher didn't carry out any standard data analysis on data collected. Rather the data was read through and discussed with a second researcher to retrieve inputs that could be added to improve the method's suitability for children. Decisions were made based on the inspection evaluation method literature, studies previously carried out and experience of working with children.

By the end of the session, it was evident that some input made by teachers during the focus group as reported in chapter 6 was repeated. However, since teachers are not experts in usability evaluation methods, some input were not considered but rather left out because they were not in context. For example, in the focus group study, teachers said let them play the game first, then come back to them and ask what issues they have found with the game. Also in this discussion, the independent teacher made similar suggestion of which this type of evaluation could be applicable to a user study and not an inspection method evaluation where evaluators are expected to note problems as they perform the evaluation.

The independent teacher also suggested that the actual game chosen for the evaluation be given to the children (in order to inspire them) before the game criteria session. However, it is believed that this way the children's opinion could be biased. Therefore, this suggestion was tweaked, where a game in the same genre is provided to the children instead of the actual game. Apart from the issues stated above, some other suggestions made were considered. For example the teacher suggested a "show and tell" be included at the beginning to serve as visual aid for the children, as this will improve their performance. This is confirmable in literature that using imitation, children are able to do much more in collective activities or under the guidance of adults (Vygotsky, 1997). Also, the suggested that the severity scale be put in front throughout the evaluation was considered. The IMCH was therefore modified to reflect these additions which produced version 3 of the method.

7.4.2.1 INSPECTION METHOD FOR CHILDREN (IMCH) V3

The IMCH V3 still aimed at doing what V2 should do as stated in section 7.3.2.3, however V3 will proceed thus to include the modifications:

- The facilitator will make an explanation of how the entire process is to run and what is required from the children.
- The children will be provided with the technology, post it notes, colour pen or marker, pen to write and the severity scale will be placed at the front in view of all evaluators. There should either be a board or a sheet with which the game criteria will be written down.
- The Children will be provided a game in the same genre as the one to be evaluated which they will play for 5 minutes
- Children will be stopped, told a story and further asked to come up with what (criteria) they think will make a good game for the genre of game they will evaluate. This session should last for 10 minutes
- The facilitator will then do a show and tell of how the children should perform the entire evaluation from individually finding problems to the merging phase and final severity.
- Children should be shown the actual game for evaluation and be given instructions about the evaluation, then they should individually play and evaluate the game, find problems, write down problems on the post it (one problem for one sheet) and colour the sheet with the marker or coloured pen to rate the problem (s) written down.
- Children will come together to merge their problems where the facilitator will facilitate the process for each child to post their found problems on the board. The same or similar problems are to be posted in the same area.
- Then as the facilitator is still facilitating, everyone who found the problem in a category will agree on a final severity which will be marked on the post it sheets. This will be the process until all the problems are merged.

The following table shows how the method evolved from the first version and the difference in the stages.

Table 7.20 IMCH method Versions and Variation

STAGES OF THE METHOD	IMCH METHOD VERSIONS		
	IMCH V1	IMCH V2	IMCH V3
Stage 1	Explain the Evaluation Process / Method	Explain the evaluation Process / Method	Explain the evaluation Process / Method
Stage 2	**Narrate Story** so Children can come up with **Game Criteria**	**Narrate Story** so Children can come up with **Game Criteria**	Short Game Play
Stage 3	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **PLAIN** TLS severity scale	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale	**Narrate Story** so Children can come up with **Game Criteria**
Stage 4	Explain Second (Merging Phase)	Explain Second (Merging Phase)	Show and Tell Actual Evaluation
Stage 5	Children to **merge problems** using post it note on a board and **provide final severity** for each merged problems	Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY FACE TLS severity scale again	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale
Stage 6			Explain Second (Merging Phase)
Stage 7			Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY FACE TLS severity scale again

The difference in the versions could be seen in the additions or subtractions in the stages. For example, in stages three and five of version 1, the evaluators used a plain traffic light severity rating scale to rate the severity of problems but in the versions 2 and 3, a smiley face traffic light severity scale was used. Then the version 3 differs from the version 2 in stages two where the children were made to play a game for a short while in version 3 rather than the criteria gathering session as seen in the version 2. Also in stage four of version 3 a show and tell session was introduced after being suggested by the independent teacher.

7.4.3 Conclusion of the Chapter

This chapter set out to explore some concepts “storytelling and value” which has already been used in the child-computer interaction community as a useful tool for gathering design requirements from children. Though it has not been used in evaluation method with children, however, these concepts were tested in studies with younger (age 6 and 7years) and older children (age 9 and 10years) to ascertain the possibility of its inclusion in an evaluation method for children. Results gathered showed it could be viable and was included to form the new inspection method for children (IMCH). This new method was further reviewed by an independent teacher who reviewed: the tools used within the method, the method process and also made some input e.g. the need to include a show and tell session at the beginning of the method. Although this input was useful for younger children it seemed like a waste of time for older children after a verbal explanation of the method has been given.

7.4.4 Contribution of Chapter to Thesis

Teachers’ ideas were explored in this chapter this informed the creation of version 1, 2 and 3 of the IMCH method which is the contribution of this chapter.

8 CHAPTER EIGHT: TESTING THE IMCH V3

8.1 Introduction

Having designed the new method (IMCH) with inputs from teachers reported in chapter 6 and other stakeholders (children and an independent teacher) reported in part B and part C of chapter 7. As reported in 7.3.2.3, the method was designed to do the following*:*

- *Allow children to produce a logical value based criteria for a game (in the same genre as the game to be evaluated)*
- *Use the stated game criteria to find real usability problems and*
- *To rate the severity of the problems found and recorded using the traffic light severity scale.*

Therefore, this research carried out an IMCH studies with the aim of *testing the ease of use and effectiveness of the method (IMCH V3) for children*. As stated in section 3.5.1, ease of use will be determined by children's ability to understand the instructional language, tools used (e.g. severity scale) and finally are able to use the method as described while effectiveness will be determined by children's ability to use the method to find real usability problems. These will be determined following two objectives:

- To carry out the study with children to investigate whether they will be able to find real usability problems using the method
- To investigate the data that will be gathered from the study in order to ascertain whether children will encounter the same or similar problems reported in chapter five (the HE study with children) when using this method.

In order to achieve this aim and objectives, studies were carried out over two days in an iterative approach to identify problems children will encounter when acting as the evaluators in an IMCH evaluation which is reported in this chapter.

8.2 Study with IMCH V3

As stated earlier, two studies were carried out for this experiment, these studies will be reported separately. However, the analysis of study one and two were done together where the group identified for study two is classified as a follow on group to the groups identified in study 1; that is, it became group 6 for data analysis purpose.

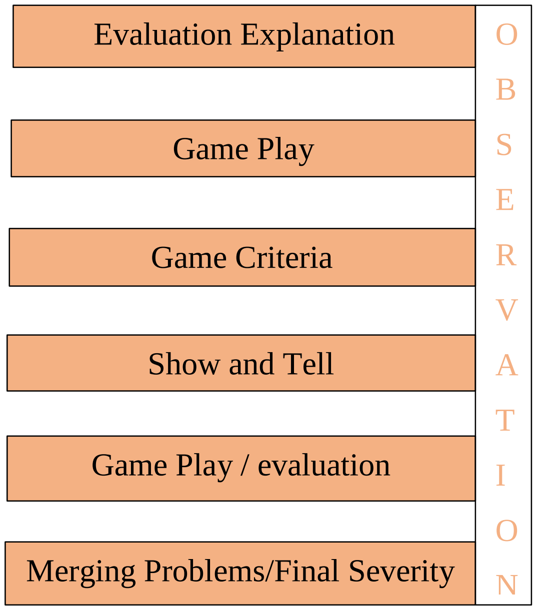

Figure 8.1 IMCH V3 Method Process

The IMCH V3 goes from explaining the method by the facilitator through to children producing game criteria, to playing and individually evaluating the game, giving it a severity to merging similar problems amongst peers and then producing a final severity rate of merged problem. The full details of how each step of the method works is shown below.

Evaluation Explanation: At the start of the evaluation process, the facilitator explains the process

Game Play: Children will be given a game that is in the same genre as the one to be evaluated for them to play for 5 minutes.

Game Criteria: A story will be narrated to the children that will lead them to come up with what they think will make a good game (criteria) in the same genre as the one they have played and will evaluate. In this session children are encouraged to discuss amongst themselves to come up with the game criteria.

Show and Tell Session: A show and tell session will be done, where the facilitator illustrates to the children how to use the tools provided and perform the evaluation on the chosen game (which is in the same genre as the one they played earlier and for which they provided the game criteria), without expressing doing the evaluation, that is without finding the usability problems or biasing their decision.

Game Play / Evaluation: Each child is provided with the game to be evaluated; a pack of Post it Note (to write down any problems encountered, one problem for one post it note sheet), a pen, and 3 coloured (green, yellow and red) markers to mark each recorded problem on the post it note to indicate its severity.

Merging Problems / Final Severity: The application on which the game was installed is retrieved from the children then either a white board or a plain sheet of paper is provided so the children could be facilitated to merge their problems. On one hand, if a white board is used, the different parts of the board could be used to merge different categories of similar problems. On the other hand, if paper sheets are to be used, one paper should be used for one category of similar problems until all problem is identified and classified. Then the agreed severity colour should be used to indicate the final severity.

8.2.1 Method

Due to the findings from chapter 5 that reports that children encountered some problems while performing the heuristic evaluation which is confirmable by literature led to the creation of the IMCH method. This IMCH method is an analytical (inspection based) method suited to children. The method requires children to produce game criteria, play game and find usability problems based on the criteria and on their experience of playing the game. They are also required to rate problems found, merged problems and give a final severity rating for merged problems.

In order to test the ease of use, practicability and effectiveness of the IMCH method for the chosen users (children), they (children) were recruited to play the role of evaluators to inspect an application, find and predict usability problems using the (IMCH) method. During the evaluation, an observation method was adopted to capture issues children might encounter during the evaluation process. It is assumed that the method (IMCH) will work for the target children since most parts of the method was developed with input from children (in the target age) and teachers who are stakeholders in the development of children’s learning aids and materials. However, it is imperative to carry out studies in the method with children to confirm the stated assumption.

8.2.1.1 Participants

In total, twenty five children (aged between 7 and 11years) and three researchers participated in this study that was carried out in two sessions. The children acted as the IMCH evaluators while three researchers acted as facilitator and observers. Though only two researchers were present in each session. All three researchers are experienced with working with children and children's technologies. Though one of the observers has limited knowledge in gathering observational data in this type of study with children.

Twenty year 6 children aged between 10 and 11 years from a UK primary school were the IMCH evaluators for five groups and five year 3-4 children aged between 7 and 8 years from another UK primary school played the role of the IMCH evaluators for group 6. This resulted in a grand total of twenty five children for the entire study.

Child evaluators were put in groups of 4 for the first five groups and a group of 5 for the last group, resulting in a total of 6 groups. The First researcher (author) played the role of the facilitator and a support observer in both sessions while the second researcher acted as the active observer for the first session and the third researcher as the active observer for the second session.

8.2.1.2 Apparatus

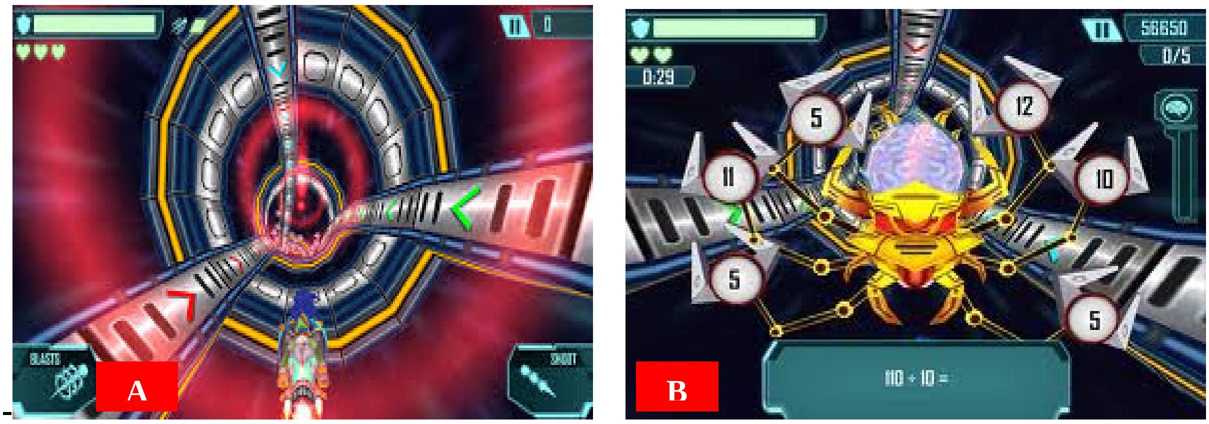

Figure 8.2 Maths Blaster Game a. Game play view b. Completing a level view

To inspire and motivate the children, the free version of the maths blaster game (see fig 8.2) was pre-installed on an iPad and provided to each child. During the study, the children produced game criteria based on their values. With groups one to five, this was written down on a white board using a marker but with group six (the last group), it was written down on a plain sheet of paper using a pen because a white board was not available. For the actual evaluation, children had to play and critic the ice maths ninja game (see fig 8.3) which had different sub games; this was also installed on the same iPad. In order to write down problems encountered, children were each provided with a pack of post it note and pen. To rate problems found and written down, they were each given coloured markers (red, yellow and green), and the traffic light severity scale (coloured chats) (see fig 7.7) were posted on the wall next to the board as a guide for all the children to see. Finally to merge their problems, the white board was used for groups one to five and a plain paper was used for group six; and the facilitator used white board markers (red and green) and yellow highlighter to circle the merge problems as an indication of the final severity. Though a black colour white board marker was used to circle problems that were considered as "not a problem".

Figure 8.3 Ice maths ninja game – a. Sub game selection page b. Actual game play for subzero mines sub game

8.2.1.3 *Procedure*

The facilitator started the process with an introduction of the researchers, and further explained what the study is about and what the children are expected to do. Then children were each provided with an iPad to play the first game (Maths Blaster – a racing simple maths game) for 5 minutes. This game is intended to be a taster game to inspire the children for the game criteria session and make them expectant of subsequent sessions (This was suggested by the teachers in chapter 6 and the independent teacher. See 7.4.1). The next session was the game criteria session where the facilitator narrated a story that motivated the children to come up with game criteria based on their values, believes and having played similar game (game in the same genre) for 5 minutes. Before the narration, the facilitator asked the children if they had played games before, whether they still played games, and if they thought any of the games they had played in the past was a good game. This was to establish previous knowledge of playing a good game to make them think about what a good game should be.

Then the facilitator carried on by telling a story that will guide the children to come up with the game criteria:

"I have a friend who intends to build a maths game for children like yourselves and he is not sure about what will make a good maths game for children. He asked me if I knew what will make a good game for children but I was not sure too. We agreed that I come and ask you

directly what you think will make a good maths game for children like yourselves. So based on the games you have played in the past and with the maths blaster game you just played, WHAT DO YOU THINK WILL MAKE A GOOD MATHS GAME FOR CHILDREN?"

Children were given the opportunity to state criteria for a good game and once the children stated a criterion, the facilitator wrote it on the board for the children to see until the children indicated they had no more. This session lasted for approximately 10 minutes.

Figure 8.4 First IMCH Study with older children A. Game Criteria Being Written Down. B. Merged Problems on the Board with final Severity

The facilitator further explained to the children what the next session entails and spent 3minutes in a show and tell session (where the facilitator acts out what the children should do) to illustrate the exact activity. Though, this was omitted for groups 4 and 5 as the researcher decided after a debrief session with the main observer that the show and tell session left the children idle and bored and was insignificant for that age group. But it was reintroduced for group 6, as they were younger participants. The session required the children to play a second game (ice maths ninja), become critical of the game (based on their values, believes and from the game criteria discussed in the previous session), write issues they might encounter while playing the game on a post it note and rate the game by colouring with the appropriate marker colour using a facial traffic light scale (see fig 7.7) as a guide. The facilitator handed the post it note, coloured markers (red, yellow and green) and pen to the children to do the evaluation. Since the game had different sub games, children were made to play 5 sub games. They were stopped after playing e ach sub game for approximately 2 to 3minutes and reminded to write down and rate any issues before being told the next sub game to play. This was the process until all sub games were played and the session lasted for approximately 12 minutes.

Merging Problems

In the next session the children were facilitated to merge similar problems and agree on a final severity. For this process, a child who indicates an interest of going first is picked to go first. The child posts his/her first problem on the board (for groups 1 to 5) and on a plain sheet (for group 6) and reads it out to the other children, then children with the same or similar problem walks to the board one after the other to post theirs underneath the first problem (see fig 8.5). Afterwards, the children who had those problems were asked what final severity should the problem have and as they agreed on the final severity colour, the facilitator circles the merged problem with the appropriate board marker colour to indicate the final severity (see fig 8.4B above and 8.5 below).

Figure 8.5 Children Merging their Problems

This was the process until all the problems were merged. This session lasted for approximately 7 minutes at the maximum and 4minutes at the minimum; the more problems available to be merged, the more time spent in merging.

8.2.1.4 Data Analysis for the IMCH V3

Child evaluators' data and observers' data were analysed separately. To easily analyse data, child evaluators' data was inputted into excel spread sheet by individual problems, individual severity rating, merged problem, final agreed severity rating and frequency of problem found. Thematic analysis and open coding techniques were the analytical methods used, since these are valid and well referenced qualitative methods in the HCI literature. The analysis was done by the researcher, however the analysis of data went through 3 iterative process between the researcher and a second researcher who is experienced in working with data collected from children. In any instance of discrepancies in researchers' opinion, these were discussed in view of the objective of the study and based on literature to reach agreement.

8.2.1.4.1 Child Evaluators' Data Analysis

In order to determine the effectiveness of the method (correct use of the method and get the result of what the method is supposed to do) child evaluators' data was analysed in four stages:

1. Analysis of the child evaluators' game criteria,
2. Analysis of the child evaluators' usability problems,
3. The investigation of the link between usability problems and game criteria and
4. Determining the severity rate given to usability problems found.

8.2.1.4.1.1 *Stage 1: Analysing Child Evaluators Game Criteria*

The objective of the game criteria session in the IMCH evaluation, is to investigate whether children can come up with their own game criteria based on their values and whether these criteria will inform their decision when judging the usability of the application. Therefore the analysis of the children's game criteria data was carried out in 7 ways:

- Identification of each group's game criteria,
- Merged data within group,
- Coded data according to their groups,
- Linked game criteria to children's usability problems found,
- Merged criteria between groups,
- Found themes for merged criteria and
- Matched criteria to existing heuristics/guidelines

8.2.1.4.1.1.1 *Identification of Each Group's Game Criteria:*

In order to determine whether each group had produced a criteria, the game criteria data was inputted into table by groups (i.e. each group criteria was inputted in a separate column). It was determined that each group produced game criteria.

8.2.1.4.1.1.2 Merged Data Within Group:

To eliminate repetition in order to produce concise list of criteria without losing any criterion or its meaning rather to aid easy interpretation of data, each group's list of criteria was read through more than once to determine similarities in the criteria. In the instance of similarities in criteria, they were merged. For example, in group 2, there were three different criteria

- “It should be harder”,
- “It should be challenging”
- “it should make you think, not too easy but not too hard that you can't do it”.

These were merged into one criterion

“It should be challenging such that it makes you think not too easy but not too hard that you can't do it”.

8.2.1.4.1.1.3 Coded Data According to Their Groups:

In order to aid easy interpretation of the data, each criterion was coded to reflect Game Criteria (GC), Group number (e.g.1) and Alphabets were used to set uniqueness for individual criterion within each group (e.g. A) see table 8.1a, full tables for all the groups are available in appendix 4.

Table 8.1 Detailed Criteria for Each Group

GROUP 1	GROUP 2	GROUP 3
GC1A – It should be fun (1)	**GC2A** – It should be fun doing it	**GC3A** – Have a maze for which you will need to solve Maths problems to get out (10)
GC1B – There should be other maths problems like subtraction and multiplication (7)	**GC2B** – There shouldn't be lots to do before you start playing	**GC3B** – Inside the Maze, do mini maths questions in seconds to go pass monsters (9)
GC1C – It should be played in different world (5)	**GC2C** – It should be creative	**GC3C** – You could use brightness to make game option stand out

8.2.1.4.1.1.4 Linking Game Criteria to Children’s Usability Data:

The criteria were further compared with the usability problems to determine a connection (see stage 3 of the data analysis for explanation of this process).

8.2.1.4.1.1.5 Merging Criteria between Groups:

To achieve part of the objective of the game criteria session, it will be useful to determine whether children had the same or similar value. To determine this, game criteria were compared between groups to determine similarities. In the instance of similarities, they were merged using a merging technique where words or phrases from each criterion was put together preserving the intended meaning for criteria merged and the groups from which criteria emanated were identified. For example GC1D (Get stars for doing it right), GC3E (reward for doing stuff), GC3F (Be able to use reward), GC4H (Be able to collect coins), GC5G (Get a price when you pass a level) and GC6D (Collect coins to unlock weapons and islands) are related to reward. Therefore, these were merged together to produce a big criteria: "You should get rewards for doing task right and you should be able to use your reward for acquiring items during game play" see table 8.2 and full table in appendix 4C

Table 8.2 (Extract) Similarities of Game Criteria between Groups

S/No	Merged Game Criteria between Groups	Groups						Total Group
		1	2	3	4	5	6	
1	It should not only be about maths but it should be fun too	*	*		*			3
2	You should get rewards for doing task right and you should be able to use your reward for acquiring items during game play	*		*	*	*	*	5
3	There should be different bikes and you should be able to change	*						1
4	There should be other maths problems e.g. Addition, subtraction and multiplication	*					*	2
5	There should be different difficulty level or world to complete the game	*			*	*	*	4
6	It should be challenging such that it makes you think not too easy but not too hard that you can't do it	*	*		*	*		4
7	You should be able to play against other characters or persons e.g. aliens or bad guys	*		*		*	*	4

8.2.1.4.1.1.6 Finding Themes for Between Group Merged Criteria:

After all the groups' criteria have been checked and similarities merged, themes were identified from each merged criterion following a thematic analysis approach. Where each criterion was read systematically and themes derived from patterns that occur in the criterion or criteria (i.e. words or phrases). For example, Useful Reward, the criteria is "You should get rewards for doing task right and you should be able to **use** your **reward**". Sometimes themes are derived from the intended meaning of the good game criteria stated. For example the theme "Game Content Preference" has different criteria which are associated with children's preference of game content (e.g. It could be like a guessing game, It should be comparing numbers) see table 8.8 below for extract and appendix 4G for full table. In some other cases, patterns identified from criteria are compared to standards of themes in game heuristics/guideline literature to determine a final theme. For example, The criterion "There should be a story about a person when you get the story you progress" was classified under the theme "Game Story" which is a category that had similar criteria in the literature by Desurvire *et al*., (2004). These were the processes until all the criterion was classified under a theme, see table 8.8 in appendix 4G for all themes.

8.2.1.4.1.1.7 Matching Criteria to Existing Heuristics/Guidelines:

In order to analyse the logic of each criterion, classified criteria were matched against existing game heuristics or guidelines. If a criterion is similar or close in meaning to an existing game heuristics, it is determined as matching that heuristic. See table 8.8 below for extract of identified themes and criteria that matched existing game guidelines, full table is available in appendix 4G.

8.2.1.4.1.2 *Stage 2: Analysing Child Evaluators Usability Problems*

In order to determine if children's usability problems found are real, merging technique (Caracelli & Greene, 1993) was used to analyse the data. Data analysis was carried out in seven ways, six is described below and one already described above under the analysis of the game criteria data:

- Identifying children's merged problems
- Identifying real problems
- Merged problems within group

- Coding real problems in each group
- Linked Real Problems to Game Criteria (Same as that reported under similar heading in the analysis of game criteria)
- Merge real problems between groups and coding
- Determine other categories from real problems

8.2.1.4.1.2.1 *Identifying children's merged problems*

Since child evaluators for each group put their post it note together to indicate merged or categorised problem (as required by the method) and didn't decide on a particular phrase for their categorised problem, the researcher decided a phrase for each categorised problem to aid the analysis process. These phrases were identified as the merged problem. The merged problems were identified following a merging technique in an in vivo coding style. For example in group six (6), 4 out of 5 child evaluators individually indicated in their problem report that the game froze at some point and during the merging phase they had put their "game freeze" post it notes together stating the problems were similar. So in the individual problem column, their problems were inputted sequentially into 4 rows, their individual severity was typed (using letters to indicate the colour of severity) into the next column. However, in the merged column these 4 rows were merged into a single problem phrase "Game froze" (some words which appear in the problem). The final severity the children indicated was inputted into subsequent column (using letters also). Finally, the frequency of evaluators who reported that problem was typed in the last column (see table 8.3).

Table 8.3 Example of categorised problems merged as one

GROUP six – Child evaluators' problem data					
S/No	**Problems identified**	**individual severity**	**Merged Problem**	**Final Severity**	**Frequency**
1	It sometime freezes	Red	Game froze	Yellow/Red	4
2	The problem is when it freezes	Red			
3	When the level finishes, then it make it freeze	Yellow			
4	It went frozen	Yellow			

In cases where a problem was found by one child evaluator, such problems were rewritten in the merged column as the merged problem (See table 8.4 below). This was the order until all the data were inputted and merged see appendix 3 for usability problems for each group.

Table 8.4 Example of single problem reports from group 4

S/No	Problems identified	individual severity	Merged Problem	Final Severity	Frequency
1	Bit Childish	Green	Bit Childish	Green	1
2	At the start it goes on the internet	Yellow	At the start it goes on the internet	Yellow	1

8.2.1.4.1.2.2 *Identifying real problems*

Real problems were ascertained by the researcher after personally playing and evaluating the same game. These problems were presented in tables, see table 8.5a and 8.5b below. Through this medium some other problems were tagged as unreal e.g. in group 5 the problem "When I get a score, it doesn't show", after it was discovered that scores were displayed on the right hand side of the game and after the game ends as evident in fig 8.3 above.

8.2.1.4.1.2.3 *Merged Problems within Group*

The researcher further compared the usability problems within groups to identify similarities in merged problems and merged them further. It was decided that some problems might have occurred because of the state of the game which has been reported as a problem therefore these two were seen as similar and merged e.g. "It won't let you press retry" (which happened because the game froze) and "Game freezes especially when one finishes" from group 2 were merged to produce the problem "Game freezes so it won't let you press retry" see tables 8.5a and 8.5b for extract and full tables in appendix 3.

8.2.1.4.1.2.4 *Coding Usability Problems Identified for Each Group*

In order to gather general understanding of total problem found by the children, to aid data interpretation and to make referencing easier, each merged problem was coded with MP to indicate 'merge problem', a number to indicate the group number from which problem was reported, and an alphabet e.g. A-Z to indicate a unique identity within the group see table 8.4 and 8.5.

Table 8.5 Code and Details of Merged Problems for Each Group (1 to 3)

GROUP 1	GROUP 2	GROUP 3
MP1A – Too Confusing	**MP2A** – Numbers are too close together	**MP3A** – It doesn't stay on the game and it glitches
MP1B – Difficult	**MP2B** – I don't exactly understand as it is too fast	**MP3B** – There isn't a language button if you are from a different country
MP1C – Long Intro	**MP2C** – I put the right answer but it said it was wrong	**MP3C** – When it doesn't click or go onto what I want I have to clear it and go back on the game
MP1D – At the start there's a video, nothing wrong at the start when you choose easy, hard it doesn't show you how to start only says choose game	**MP2D** – It won't let you press retry	**MP3D** – It wasn't clear on what to do

Table 8.6 Code and Details of Merged Problems for Each Group (4 to 6)

GROUP 4	GROUP 5	GROUP 6
MP4A – Bit Childish	**MP5A** – Can't tell which is 9 and which is 6	**MP6A** – Game froze
MP4B – At the start it goes on the internet	**MP5B** – Starting again is annoying	**MP6B** – You can't catch up with the falling ice
MP4C – Game won't restart when you lose neither will it go to home	**MP5C** – It keeps going on to Facebook	

8.2.1.4.1.2.5 *Merged Problems between Groups and Coding*

In order to merge similar problems between group, problems were compared between the groups, where the same or similar problems are merged to produce a further merged problems (FMP) e.g. MP1D, MP2E and MP3D were merged (see table 8.5 and 8.6) together to produce the FMP12 "Instructions were not very clear at the start and later" see table 8.7.

In order to aid easy classification and referencing, the further merged problems were coded with the acronym FMP (indicating they are further merged problem) and a unique number.

Table 8.7 Merging Problems between groups after identifying real problems

S/No	Further Merged (FM) Problems	Groups						Total Group
		1	**2**	**3**	**4**	**5**	**6**	
1	**FMP1** Too Confusing, can't tell which is 9 and which is 6, (Confusing)	*				*		2
2	**FMP2** Difficult (Vague)	*						1
3	**FMP3** Long Intro	*						1
4	**FMP4** Game keeps freezing, especially when one finishes and it won't let you press retry (Obstruction)	*	*	*	*	*	*	6
5	**FMP5** Numbers are too close together (Content Spacing)		*					1
6	**FMP6** I put the right answer but it said it was wrong, It doesn't add right sometimes (inappropriate Scoring)		*	*				2
7	**FMP7** It doesn't stay on the game e.g. it goes on the internet and it glitches (Obstruction)			*	*	*		3
8	**FMP8** There isn't a language button if you are from a different country (Language Inaccessibility)			*				1
9	**FMP9** Bit Childish (Age inappropriate)				*			1

10	**FMP10** Starting again is annoying (vague)					*		1
11	**FMP11** I do not understand as it is too fast; you can't catch up with the falling ice (Fast Pace)		*				*	1
12	**FMP12** Instructions were not very clear at the start and later (Unclear Instructions)	*	*	*				3

8.2.1.4.1.2.6 *Determine other categories for real problems*

After deciding on which problems were real and unreal, the researcher read through the real problems to determine more problem categories. This is intended to clearly inform on the areas children's found problems are situated. The categories identified are:

- Confusion **FMP1**,
- Content Spacing Limitation **FMP5,**
- Vague (**FMP2** and **FMP10**),
- Long Intro **FMP3**,
- Obstruction (**FMP4** and **FMP7**),
- Inappropriate Scoring **FMP6**,
- Language Inaccessibility **FMP8**,
- Age Inappropriate **FMP9**,
- Fast Pace **FMP11**,
- Unclear Instructions (**FMP12**), see table 8.6.

8.2.1.4.1.3 *Stage 3: Investigating the link between found usability problems and Game Criteria*

In order to investigate whether children applied the criteria they produced to problem finding during the evaluation. Each group's problems were crosschecked systematically and compared to the game criteria. This comparison was first carried out within group then later between groups, if a problem violates a criterion or criteria from the same group, the criteria code is written in black against the problem, determining them as a link e.g. in group 4 the problem "At the start it goes on the internet" violates the game criteria "do not put too many glitches in", so these were termed as a link. However if the problem matches the criteria from another group, the game criteria code is written in red against the problem code. This was the order until all the problems found have been compared to all the criteria. See table 8.8 below for problems that match to criteria.

Table 8.8 Linking usability problems to game criteria

GROUP 1		GROUP 2		GROUP 3		GROUP 4		GROUP 5		GROUP 6	
MERGED PROBLEMS	**GAME CRITERIA**	**MERGED PROBLEMS**	**GAME CRITERIA**	**MERGED PROBLEMS**	**GAME CRITERIA**	**MERGED PROBLEMS**	**GAME CRITERIA**	**MERGED PROBLEMS**	**GAME CRITERIA**	**MERGED PROBLEMS**	**GAME CRITERIA**
MP1A	X	MP2A	X	MP3A	GC4G	MP4A	X	MP5A	X	MP6A	GC4G
MP1B	GC2E	MP2B	X	MP3B	X	MP4B	GC4G	MP5B	X	MP6B	X
MP1C	X	MP2C	X	MP3C	X	MP4C	GC4G	MP5C	GC4G		
MP1D	X	MP2D	GC4G	MP3D	X			MP5D	GC4G		
MP1E	GC4G	MP2E	X	MP3E	GC4G			MP5E	X		
MP1F	X	MP2F	GC4G	MP3F	X			MP5F	X		
				MP3G	X						
				MP3H	X						
				MP3I	X						
6	2	**6**	2	**9**	2	**3**	2	**6**	2	**2**	1

Each of the six groups found problems and at the end of data analysis, these merged problems violates some of the game criteria children produced. In some instance, a group's merged problem violates the game criteria from within the same group but in some other instance it violates the game criteria from another group. For example: In group 4, two merged problems (MP4B and MP4C) violates a single criteria (GC4G) from the same group. Meanwhile other groups' merged problem only violates the game criteria from group 4 and a merged problem (MP1B) from group 1 also violates a game criteria (GC2E) from group 2.

Table 8.8 also shows that not all the merged problem from the groups violates the game criteria produced, either within the group or between the group. For example: Groups 1, 2 and 5 each had 6 merged problems but only 2 merged problems from each violates the game criteria produced. Group 3 had 9 merged problems and only two violates a single game criteria. Finally group 6 had two merged problems and only one problem violates a game criteria. This is graphically represented below:

Graphical representation of Table 8.8 – The link of merged problems to Game Criteria

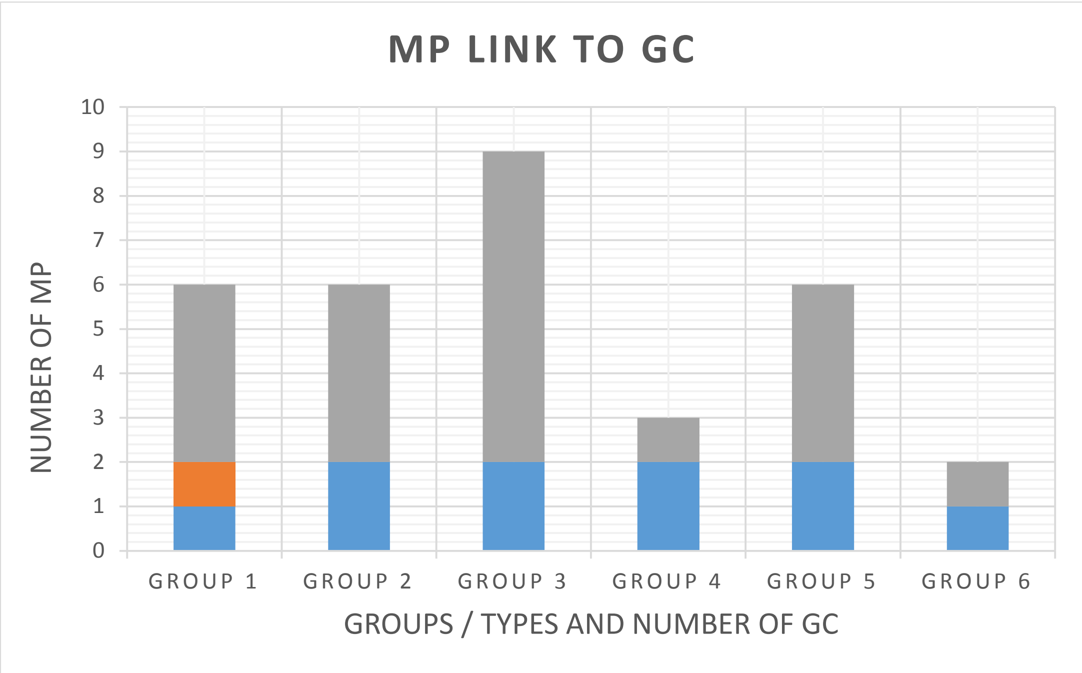

Figure 8.6 Merged Problem Linked to Game Criteria

(KEY: MP – Merged Problem, GC – Game Criteria)

There are 6 groups and these groups produced merged problems with group 3 having nine merged problems, which is the highest number of merged problems and group 6 having only two merged problems, which is the least number of merged problem. This also shows each group has at least one merged problem linked to a game criterion from between group, where group 6 is the only group with one link but other groups have two merged problems linking to game criteria.

The total number of merged problem for each group have been indicated with the grey bar then the type and number of the game criteria each group's merged problem links to or violated have been indicated with the blue and orange bar.

In this chart, merged problems from each group that links to game criteria from within or between groups have been identified, e.g. all the groups have the blue bar which indicates a type of game criteria, this means all merged problem from all the groups except one merged problem in group one violates a single type of game criteria, that is, only group 1 had two merged problems that links to two different game criteria but the other groups links only to one game criterion.

8.2.1.4.1.4 *Stage 4: Identifying Severity Ratings for problems found*

The researcher read through the problems reported for each group to determine the severity rate given. It was observed that children had rated their problems into 6 categories (red, yellow, green, red and yellow, and green and yellow). However, it was observed that a problem was not allocated a severity so this was classified as "Not Rated". In some other instance, children had issues deciding on a final rating for a problem (Game froze) in group 6 as the severity allocated was divided into two severity levels (red and yellow), this was classified as such (red and yellow) see table 8.9 for all severity ratings. In another case, a child in group 5 had identified a problem "How to start a new game" and rated it as "Green and Yellow".

Table 8.7 Merged Problems and their Severity Ratings

Severity Colours	**Merged Usability Problems**	**Total**
Green	MP1C, MP3F, MP4A	3
Yellow	MP1B, MP2C, MP2D, MP2E, MP3D, MP3G, MP3H, MP4B, MP4C, MP5B, MP5E, MP3I	11
Red	MP1A, MP1E, MP1F, MP2A, MP2B, MP2F, MP3A, MP3B, MP3C, MP3E, MP5A, MP5C, MP5D, MP6B	14
Green and Yellow	MP5F	1
Yellow and Red	MP6A	1
Not Rated	MP1D	1

KEY for some problem codes:

Green: **MP1C** – Long intro, **MP4A** – Bit Childish

Yellow: **MP1B** – Difficult, **MP2C** – I put the right answer but it said it was wrong

Red: **MP1A** – Too Confusing, **MP2B** – Numbers are too close together

Green/Yellow: **MP5F** – How to start a new game

Yellow/Red: **MP6A** – Game Froze

Not Rated: **MP1D** – At the start there's a video, nothing wrong at the start when you choose easy, hard it doesn't show you how to start only says choose game

Table 8.8 (Extract) Identifying game criteria themes and matching them to existing heuristics/guidelines

S/No	Themes and Criteria	Groups	Existing Game guideline/Heuristics
MGC1	**Aesthetics/Design**		
A	It should have good theme song	3	Application should react in a consistent, challenging, and exciting way to the child's actions (e.g., appropriate video clips with the music). (Alsumait & Al-Osaimi, 2009) Should use visual and audio effects to arouse interest (Federoff, 2002)
B	It should have good graphics	2	The font choice, colours and sizes are consistent with good child screen design (Alsumait & Al-Osaimi, 2009)
C	You could use brightness to make game option stand out	3	Use noticeable and distinct avatars that have intuitive information mappings (Pinelle et al., 2009) Make effects of the Artificial Intelligence (AI) clearly visible to the player by ensuring they are consistent with the player's reasonable expectations of the AI actor. (Desurvire et al., 2004)

MGC2	**Game Content Preference**		
A	There should be questions and the answer options should come up for you to answer the questions	6	
B	There should be a football shooting at the right sum	5	
C	It could be like a guessing game	5	
D	It should be comparing numbers	6	
E	It should be adventurous	6	
F	You should be able to build stuff	5	Allow players to build content (Federoff, 2002)

8.2.1.4.2 Observers' Data Analysis

In order to determine the suitability of the method and further confirm the effectiveness of the method, the Observers' data was analysed by the researcher using content analytical method described by (Holdford, 2008). Upon scrutiny of the data, it was evident that data was presented as either problems or comments, therefore the analyses was carried out to reflect these (problems and comments). The researcher read through the observers' data more than once to gain full understanding of the data.

Analysis was done serially by groups. In each group, data was put in two character categories (Facilitator and Evaluators) to reflect the subjects being observed and further into two sub categories Problems and Comments. For example under the Facilitator category, there was "problem" and "comment" sub categories to separately capture problems and comments that relate to the facilitator. This was also the case for the evaluators.

To rigorously analyse the data, each sub category had a 5 column table which captured the serial no, problem or comment, the identity of the observed, the problem association (the point of the evaluation session when the observation was made) and a fifth column to code each observed content. See table 8.11 for the extracted data, full detail is available in appendix 4.

Table 8.9(Extract) Group 1's Observed Data for problem sub category

S/No	Problems	Identity	Problem Association	CA Theme
1	Not used to playing therefore, moving slowly and confused	C	Gameplay	Slowed down – lack of game play experience
2	Playing with one hand, therefore moving slowly	B	Gameplay	Slow movement – use of one hand
3	Using one hand to push button and it is taking time because the surface is too large- Moving slowly and confused	A	Game play	Using one hand and confused – Moving slowly

8.2.1.4.2.1 Coding Observers' Data

After all observed content had been put into their appropriate sub categories (problems or comments), each sub category data was carefully and systematically read through, it was discovered that data reported under comment was not useful for this study, so that was not taken further. However, the data identified and put in the problem sub categories were analysed using a qualitative content analytical method where patterns of phrases from the data content were identified as themes, noting frequencies of occurrence and groups see table 8.12 and appendix 4. These themes were attached as a coding unit until all the data have been coded with a theme.

After problem patterns have been identified for each child evaluator in a group. It is believed that some or all problems observed could affect the children's performance during the evaluation. Therefore, in order to identify the ways by which children's performance were affected, observers were informed of the need to capture the effect of any problem observed. These effects have been captured alongside the problem themes reported. Although it was discovered that some problems were captured without effect.

For the analysis, similar problem patterns that occurred for all the evaluators in the group were merged and identified as the theme e.g. in group 1, it was observed that evaluator A and B were "using one hand" and this had an effect on their evaluation by "slowing them down". In this instance, using one hand was reported twice on the group and it was reported that for both times, it slowed down the evaluators, therefore for group 1, "using one hand" was identified as one problem and "slowed down" as a problem effect against the said problem. Sometimes, a problem might have multiple effect, these are captured and put against the problem. For example in group 1, evaluator C was observed as lacking game play experience which had a "slowed down" effect and a "confused" effect on his evaluation. Therefore it was captured that for group 1, lack of game experience had a slowed down and a confused effect on an evaluator see table 8.12. This was the process until all content was coded, problem themes identified with problem effects appropriately allocated within the group.

Further analysis was carried out to compare problem themes and effect reported between groups, in the instance of similarity in problem themes, they were merged to form a single problem theme but if different effect are identified, these are attached as multiple effect to the problem theme e.g. the problem theme "Disregarded Instruction" had two problem effect "Not know how to play" and "played on a different level". These were all attached to the problem and noted as ways in which disregarded instruction affected children's performance see table 8.12 below.

8.2.1.4.2.2 Categorising Observed Problem Areas

After problem themes have been identified from problems, the area during which problems were observed were also classified as associated problem area, see table 8.11 above and individual group table in appendix 4. Problems were further categorised under these problem areas. The areas identified are: Game Play (see table 8.12), Game Criteria (table 8.13), Merging Phase and Severity Rating.

Table 8.10 Problems Observed Around Game Play

S/No	Problem Themes	Groups						Problem Effect
		1	2	3	4	5	6	
1	Use of one hand/Crossed Hand	*			*	*		Slowed Down
2	Confused	*						
3	Problem Finding Level		*					
4	Not Following / Disregarded Instruction	*						Not Finding Problems
		*						Plays another game
		*						Wrote two problems on a single sheet
		*						Not know how to play
		*	*					Plays on different level
				*				Chose same level/game twice
		*			*			Wrote problems at the end
5	Game Froze	*						Game play obstruction
		*						Helping out one another
		*			*			Confused
			*		*			Slowed down
					*			Annoyed
					*			Lacked enjoyment
							*	Kept seeking assistance

6	Unclear Game Instruction			*				Children did not understand instruction
7	Couldn’t Find Sub Game					*		Lags behind
8	Study break						*	Unsure of problems previously encountered
9	Fast Pace Game	*	*					Difficult to find sum
10	Stating out Problems Found	*						Potential peer bias
11	Lack of Game Play Experience	*						Slowed down
		*						Confused
							*	Unsure of action

Table 8.11 Problems Identified for Game Criteria Area

S/No	**Problem Themes**	**Groups**						**Problem Effect**
		1	**2**	**3**	**4**	**5**	**6**	
1	Distracted by Game sound				*			Not listening to narration
2	Shy start					*		-
3	Taking Pictures with IPad	*						Not concentrating
4	First time tablet user	*						Not attentive to explanation
5	No interaction (Facilitator and Children)	*						Bored
6	Not playing game			*				
7	Lack of game play experience						*	Bored, confused and Not Contributing

Table 8.12 Problems around Merging phase and Severity Rating

Problem Area	Problem Theme	Group	Problem Effect
Merging Phase	Not following instruction (1)	1	Delayed the merging phase
Severity Rating	Did not rate problems (1)	1	-
	No group interaction (1)	2	Difficulty agreeing on severity
	Evaluators’ Effect (1)	6	

8.2.1.4.2.3 Analysis of Observed Data for the Facilitator (Researcher)

Observed data gathered for the facilitator was analysed using similar approach employed for the child evaluators’ data, using tables and qualitative content analysis approach. In this case the identity is always the facilitator.

8.2.2 Result

The results are reported separately in view of the children and the observers’ data.

8.2.2.1 Child Evaluators’ Game Criteria

8.2.2.1.1 Identification of Children’s Game Criteria and Merging Within group

The sum of fifty two (52) criteria was identified between groups. However, after merging similar criteria in group 2 (as group 2 was the only group with simiar game criteria within the group), the number of total criteria between groups reduced to 50 Game criteria. Group 2 ended up with the least number of 5 criteria and group 6 with the highest number of 12 criteria. Though each group had a total number of criteria identified in them, some criteria were present in more than one group while some were unique to a particular group. However, one criterion was identified in all the groups, see table 8.2 and 8.10. After merging criteria available in more than one group, a total of 29 criteria was identified for all the groups (21 criteria were separately unique to a group (see table 8.2) while 8 criteria were separately in more than one group e.g. 1 criterion was available in three groups, 3 criteria were each available in two groups, 3 criteria were each available in four groups, and 1 criterion in 5 groups see table 8.2)

Table 8.13 Summary of Children's Game criteria data

Groups	1	2	3	4	5	6	All	Total
Number of Criteria before merging within group	8	7	8	8	9	12	52	52
Number of Criteria Identified after merging within group	8	5	8	8	9	12	50	50
Number of Criteria Unique to group	1	3	3	3	4	7	0	21

Identifying themes for children's game criteria

The sum of 12 themes were identified and game criteria classified into them with the minimum number of classified criteria in a theme as 1 e.g. Inspiring/Imaginative, Game Story, Useful reward, Fun and Multi Player and the maximum number as 7 for "Flexibility" theme, see table 8.10.

Table 8.14 Themes for Children's Game Criteria

S/No	Game Criteria Themes	Number of Criteria for each Theme
1	Aesthetics and Design	3
2	Game Content Preference	6
3	Game Progress	2
4	Multiplayer	1
5	Minimal Frustration	2
6	Fun	1
7	Useful Reward	1
8	Flexibility	7
9	Challenge	4
10	Inspiring and Imaginative	1
11	Game Story	1

8.2.2.1.2 Findings for Matching Children's Game Criteria to Existing Heuristics/Guidelines

After matching the 29 criteria that was produced from the merged criteria between groups to existing game heuristics/ guidelines, it shows 18 of the 29 criteria matched existing heuristics/guideline from 6 literature, see appendix 4.

8.2.2.2 Child Evaluators' Usability Problem Data

Following the data analysis of the child evaluators' data, it shows that each group found and recorded problems with the least recorded group problems as two and the most recorded group problems as nine. Collectively, all six groups reported 32 raw usability problems, see tables in appendix 4 and table 8.17 below, however, only 23 problems were judged as real problems by the researcher. 2 problems were judged as unreal (i.e. it didn't occur e.g. in group 5 the problem claim was there is no score displayed, meanwhile scoreboard is statically available during game play). Two other similar problems (each from group 2 and 3) were judged as possibly unreal because this might have occurred due to the evaluator's error not a game malfunction or glitch but because it was seen by two people it is left as an in between real and unreal (possibly unreal) problem and One problem ("None interduction" from group 1) is classified as incomprehensible as it is difficult to decide on what the child meant and the 4 remainder of the 32 problems have been merged within group since real problems were determined after problems were merged within group see table 8.17 and other tables in appendix 4.

Table 8.15 Summary of Children's usability problems reported

Groups	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Number of raw usability problems	6	6	9	3	6	2	32
Number of real usability problems	5	4	6	3	4	2	23
Number of problem(s) merged with another within group	0	2	3	0	0	0	5
Number of problems merged with other group (s)	2	3	3	2	3	1	14
Number of real usability problems unique to each group	2	1	2	1	1	1	8
Number of unreal problems	0	0	1	0	1	0	2
Number of Possibly Unreal	0	1	1	0	0	0	2
Number of Incomprehensible	1	0	0	0	0	0	1

8.2.2.2.1 Findings of Merging Problems between Groups and Identifying Themes

After merging the similar real problems between the groups, the total number of real problems reduced to 12 problems see table 8.4 above. Findings also show a total of 10 themes identified from the problems with two themes having the maximum number of 2 problems each and the other 8 themes having the minimum number of 1 problem each, see table 8.16. During theme identification and problem classification, 2 problems have been classified as being Vague as it is unclear on what exactly the problem is describing e.g. problem FMP10 states "Starting again is annoying". It is difficult to ascertain why it is annoying, see table 8.5.

Table 8.16 Themes Identified for Children's Problems found

S/No	Further Merged Problem Themes	# of FMPs Identified
1	Confusion	1
2	Vague	2
3	Long Intro	1
4	Obstruction	2
5	Inappropriate Scoring	1
6	Language Inaccessibility	1
7	Age Inappropriate	1
8	Fast Pace Gameplay	1
9	Unclear Instructions	1
10	Content Spacing Limitation	1

8.2.2.2.2 Findings for the Severity Ratings of Problems Reported

Result show that children rated the severity of all the problems found except one (At the start there's a video, nothing wrong at the start when you choose easy, hard it doesn't show you how to start only says choose game). 3 problems were rated as a green problem, 12 problems were rated as a yellow problem, and 14 problems were rated as a red problem. Findings also show that two problems "MP5F" (How to start a new game) and "MP6A" (Game Froze) were rated as a "green and yellow" and "yellow and red" problems respectively, see table 8.9 above.

8.2.2.3 Result: Child Evaluators' Usability Problem link to Game Criteria

Having compared the problems children found and reported, to the game criteria they stated, it shows only group 4 had problems (MP4B – "At the start it goes on the internet" and MP4C

"Game won't restart when you lose neither will it go to home") that violates a game criteria (GC4G "Do not put too many glitches in") stated in the same group. Group 1 had two problems (MP1B "Difficult" and MP1E "Game keeps freezing") that violates game criteria (GC2E "It should be challenging such that it makes you think not too easy but not too hard that you can't do it") from group 2 and group 4 GC4G (stated above) respectively. Group 2 and group 3 had two problems (MP2D "It won't let you press retry (game freezes)" and MP2F "Game freezes especially when one finishes") and (MP3A "It doesn't stay on the game and it glitches" and MP3E "Game freezes every time") respectively that violates one game criteria GC4G from group 4. Group 5's two problems (MP5C "It keeps going on to facebook" and MP5D "Game freezes when trying to restart after failing so don't know how to start a new game") both violates group 4's game criteria GC4G. A problem (MP6A "Game froze") from group 6 violates game criteria GC4G from group 4, see table 8.8.

8.2.2.4 Result from Analysing Observers' Data

8.2.2.4.1 *Observation for Child Evaluators*

8.2.2.4.1.1 Observed Problems for Children

Observers' data show a total of 50 raw problems were reported for the child evaluators. However, these problems were analysed down to 23 problems situated in four areas Game play, Game Criteria, Merging Phase and Severity Rating. Result shows 12 problems which affected the child evaluators in 24 ways were observed for "Game Play". 7 Problems with 6 effect were observed for "Game Criteria", 1 problem with an effect for "Merging Phase" and 3 problems with an effect for "Severity Rating" (see table 8.17). Full details of problem themes observed and their effect is available in table 8.17 and in appendix 4.

Table 8.17 Areas in which problems were observed for Child Evaluators

S/No	Problem Areas	Associated Problems	Number of Ways Problems Affected children
1	Game Play	11	23
2	Game Criteria	7	6
3	Merging Phase	1	1
4	Severity Rating	3	1

8.2.2.4.1.1.1 *Problem findings around Game Play*

Result show that observed problems sometime affected groups in the same way e.g. "Not following instruction" had a "slowed down" effect on groups 2 and 4, also "Fast pace of the game" made it difficult for groups 1 and 2 to find the sum (see table 8.12). On another instance, a problem affected a group in diverse ways e.g. for group 1, the problem "Game froze" had the following effect "obstructed game play", "made them confused" and "made them to start helping one another", see table 8.12 above.

Table 8.18 Problems Observed around Game Play

S/No	Problem Themes	Number of Group	Number of Problem Effect
1	Use of one hand/Crossed Hand	3	1
2	Confused	1	1
3	Problem Finding Level	1	1
4	Not Following / Disregarded Instruction	4	7
5	Game Froze	4	7
6	Unclear Game Instruction	1	1
7	Couldn't Find Sub Game	1	1
8	Study break	1	1
9	Fast Pace Game	2	1
10	Stating out Problems Found	1	1
11	Lack of Game Play Experience	2	2

8.2.2.4.1.1.2 *Observed Problem around Game Criteria*

In the Game Criteria session, 7 problems were observed to have affected the children's performance. The lack of experience problem which was reported 3 times for 2 groups had negative effect on the children in 3 ways. The other 6 problems each reported once for a particular group had an effect on the children's game criteria performance. Though the Shy Start problem had no effect on the children's game criteria performance.

8.2.2.4.1.1.3 *Observed Problem around Merging Phase and Severity Rating*

In the merging phase area, it was observed that one child from group 1 did not follow the instruction therefore delayed the other group members during the merging phase. For the Severity Rating area, 2 problems observed for two groups seemed to have caused a difficulty in child evaluators agreeing on a severity, while a child not rating a problem didn't have an effect at all (see table 8.15).

8.2.2.4.1.2 Observed Comments for Child Evaluators

Apart from reporting problems some observers observed and made comments (things which they believed is advantageous for the evaluation). 7 comments (themes) in two areas: Game play and severity rating where 6 comments (themes) reported 15 times in total relates to Game Play, these comments were observed in groups 1 to 5, while the seventh comment is around severity rating. This was reported once in a group that the children did a good final severity rating when they worked together.

Table 8.19 Comments observed for child evaluators

S/No	Comments (Frequency of comment)	Group Identity	Comment Association
1	Game paly experience aids understanding (1)	1	Game play (6)
2	Game froze – Peer Helping out (2)	1	
3	Girls are more attentive than boys (1)	2	
4	Read and follow instructions (5)	2, 3	
5	Use of two hands aids quick movement (3)	4	
6	Clearer instructions aids relaxation (3)	5	
7	Good severity Rating Agreement (1)	3	Final Severity Rating (1)

8.2.2.4.2 Observation for Facilitator

8.2.2.4.2.1 Observed Problems for Facilitator

During the evaluation the facilitator was also observed and it showed that 3 problem themes were reported from the observation which bothers on Game Criteria ("Lengthy writing causing boredom for children") and Evaluation Instruction ("Lack of clear game instruction" and "Long explanations of severity scale"). These problems were observed for 3 groups (3, 4 and 5), see table 8.22.

Table 8.20 Comments observed for child evaluators

S/No	Problem Theme	Problem Association	Group Problem Occurred
1	Lengthy writing causing boredom for children (3)	Game Criteria	3, 4, 5
2	Lack of clear game instruction (3)	Evaluation Instruction (Game Play)	3, 4, 5
3	Long explanations of severity scale (1)	Evaluation Instruction (Severity Scale)	4

8.2.2.4.2.2 Observed Comments for Facilitator

There were two comments made about the facilitator during the observation which is around Game Criteria and Game Play, these comments were made for two separate groups (see table 7.23).

Table 8.21 Observed comments for the facilitator

S/No	Comment Theme	Comment Association	Group Comment was Reported
1	They might not know what a mat game is	Game Criteria	1
2	Less confusion due to clear instruction	Game Play	5

8.2.3 DISCUSSION

8.2.3.1 Discussing Child Evaluators' Problem Data

Real Problems: To determine what problems were genuine, the researcher played the game and encountered some of the problems. For example "Game freezing", "environment switch to browser page", "sometimes the game pace is fast "(although this is on higher level when the game becomes more challenging), and the "number confusion" (it is difficult to tell the number 6 from the number 9). Some of these problems are confirmable from the observed data. For example it was observed and reported severally that the "game froze" that slowed the children down or left them bored or annoyed. It was also observed and reported that the "game pace was fast". Since evaluating technologies is based on the evaluators experience and usability, it is possible that some other problems were also genuine e.g. Difficult, although this problem is quite vague as it is unclear as to what is difficult, if it is the actual game play, finding sub levels or finding subgames.

Children's Shared Values: Following the definition of "value" by (Iversen et al., 2010) and for this work (what children consider as important), it is evident from the result that children share the same value in certain game criteria, as they had similarities in some game criteria and they also had criteria that were unique to each group. Though sometimes game criteria might be unique to a group seeming children do not share the same value on it but this criteria matched existing heuristics /guidelines. Proving few and even more children could come up with reasonable game criteria.

Stating the Obvious: It was also determined that children stated the obvious as good game criteria which might not be a novel input given the game will originally have that. For example, children stated criteria such as "You will need to solve Maths problems to get out of a place or to move on" when the instruction given was state criteria that will make a good maths game but given they ought to state criteria based on what is important to them, stating such criterion is not out of place and shows they have an understanding of the game context.

Problem versus Criteria: In investigating children's problems found to criteria stated, it shows that only one group had stated problem that violates a criteria from the same group. On one hand, it is therefore arguable that the game criteria children stated did not guide children to find problem and therefore was not useful for the evaluation process even though most of the criteria seem sensible and matched existing heuristics / guidelines. On the other hand it could be argued that criteria children stated could have guided their judgement of the game, however, it may not be sufficient enough as the highest stated criteria by a group was 12 while most existing game criteria has more (Alsumait & Al-Osaimi, 2009; Desurvire *et al.*, 2004; Federoff, 2002; Röcker & Haar, 2006; Villalta et al., 2011)

Evaluator Effect: It was evident from the evaluators' problems found that evaluators had found different problems (though they were using the same method on the same game) and at some point children had issues agreeing on a severity rating, as evaluators insisted on keeping their own rating. This is referred to as Evaluator Effect. Evaluator effect described as differences in usability problems found and severity judgement by multiple evaluators who used the same usability evaluation method on the same interface (Hertzum & Jacobsen, 2003). This is often an issue when more than one evaluator is involved in usability evaluation, though this is often reported for adult evaluators, it is now evident that this is also an issue with child evaluators and also an issue in this method.

More Steps for Severity Scale: A three step traffic light facial severity scale (see figure 7.7 above) was used to rate problems found during the evaluation. It was evident from the data that evaluators had difficulties agreeing on a severity e.g. the game froze theme (see figure 7.9) as discussed in the previous section and a single evaluator used two severity colours for one problem (see table 8.9). It is conjectured that if there were more steps e.g. a step in between yellow and red then evaluators might have a point of agreement and if there was another step between green and yellow then the last evaluator could have a better choice.

More Evaluation Methods: An interesting finding is that a child evaluator had reported that the game is childish, even though the game is designed for that age group. This shows that children can become more critical with this method (IMCH) as there is tendency to evaluate beyond usability. It also gives the need to state that evaluating technologies for just one aspect (e.g. usability) might not be sufficient enough but also the need to evaluate for likeability, fun etc. As Woolrych & Cockton (2002) recommends that it is useful to employ more than one method when evaluating technologies.

Urgency: It was also reported within gameplay area that the use of one hand slowed the children down and was also commented that the use of two hands made children quicker. However, if there is no urgency this won't seem an issue but if urgency is of essence then this will be an issue. Also as multiple evaluators are involved where they need to wait on each other to merge problems it will be important that urgency is considered so other evaluators won't be kept waiting for too long.

Gender Considered: It was also reported that girls were more attentive and followed instruction than boys. Though this is beyond the scope of this research to consider but future work can look in this area.

Game Criteria Link to Problems Found: It was shown from the data analysis and result that children's usability problems did not link to their game criteria which proves that children did not use the game criteria as a guide to find problems. However, this is not farfetched as children were not encouraged to do so, even though the game criteria was left on the board in front. It is also possible that because children were not constantly drawn to the game criteria it will be difficult to retain and memorise as a guide to find problem. There is also the tendency that the number of criteria produced is not sufficient enough or the aspect the criteria covers is not broad enough to cover generic findings. Given that criteria produced seem specific rather than generic. Further studies will therefore make modifications to encourage this.

8.2.3.2 Discussing Observers' Problem Data

Observers' data compared to other data sources: The observers' data showed that game instruction for group 3 was unclear and it was reported in group 3 that it was not clear on what to do during game play. It was observed and reported that game froze and this was also reported by child evaluators. An observed data also reads for the facilitator that lengthy writing on the board without continuous interaction left the children bored, this is an issue that needs to be managed in future studies to keep children engaged throughout the entire process.

Effect of facilitator's Lengthy Explanation: It was also observed that facilitator sometimes made lengthy explanations that causes a switch off and made children to show eagerness to want to play the game. This seem evident as it was also reported that at some point child evaluators were eager to go into the game and do the actual evaluation, especially the evaluator that was a first time tablet user. Although this was not evident in the children's data but the reflective note captured by the researcher stated that some children were bored during the show and tell period (see appendix 4). Therefore, it is evident that some observed data were real and useful for making modifications of the method process in the future.

Speculative Data: Some data were based on speculations e.g. the comment made that "They (children) will not know what a mat game is", as this did not affect the children's performance neither was it reported anywhere as an issue. It is therefore useful to have prior training for observers as data collected could be mixed up such that analysis could become problematic.

IMCH Observed Data Compared to HE Observed Data: As reported in chapter 5, section 5.2.3 there were issues such as: Children not understanding the heuristics and severity rating language, children not recording problems, children being engrossed in game, issues with merging problems found (see section 5.2.3). In the IMCH evaluation, these problems were not reported for the children, it is therefore evident that the IMCH is more simplified to suit the children than the standard HE method

However, some issues were reported for the IMCH evaluation which include:

1. Lengthy explanations that led to boredom for the older children (e.g. with the show and tell method)
2. Unclear game instructions that caused difficulties in doing the task
3. A first time tablet user who is not attentive to instructions but just wants to play with the application.
4. Lack of game play experience that causes difficulty in making contributions during the game criteria session
5. The insufficiency in the steps of the severity scale that made final severity judgement difficult
6. Study break which led to confusion in the problems children reported previously
7. Children's refusal to follow instruction

Though some of the problems stated above could be managed in future studies but it is also believed that some will be difficult to address with the method. For example, the issues of 'lack of game play experience' will be difficult to address as the method does make provisions to provide game experience. Also the issue of the children's refusal to follow instruction might be difficult to manage as it was evident that the children who displayed this behaviour decided to do so; because the observer's note reported that they refused to follow instructions even though they were given the instructions several times. According to child right and ethically the children have the right to choose if they want to do what they are asked to do or not.

8.2.3.3 Judging Ease of Use and Effectiveness of the method

Ease of Use: As defined in this research ease of use is children's ability to understand the instructional language, tools used (e.g. severity scale) and finally are able to use the method as described. As discussed in the section above group 3 which had four evaluators had unclear game instruction and this affected their ability to perform the evaluation as they did not know what to do. Given this instance it could be said that the method was not easy to use at some point for this children. However, 4 child evaluators out of 25 evaluators used for this study is only 16% of the entire evaluators indicating 84% did not have any issues with using the method. Also apart from this report of unclear game instruction, no other report was made concerning children's ability to use the method.

Effectiveness: Is defined as children's ability to use the method to find real usability problems. A total of 32 raw problems were found by all the 25 children and 27 problems were judged as real problems. This is 84.4% of the entire problems reported with 15.6% judged as unreal, possibly unreal or incomprehensible. This shows high effectiveness in the children’s ability to use the method. However, because the method is still being tested it difficult to claim this high effectiveness at this stage.

8.3 Conclusion

This chapter has been able to report the IMCH studies carried out with older and younger children. It shows that children were able to produce game criteria, although game criteria produced for chapter 7 seem more generic than chapter 8. It is believed that this (outcome of chapter 8) could have occurred due to the initial game children played, which is in the same genre as the game to be evaluated.

They were also able to find and rate usability problems. However, some problems were observed which relates more to the game being evaluated, some problems on the game criteria session which states that the children became bored whenever the researcher stops interacting with them to write down stated criteria. It was also stated that the show and tell session was not very useful for the older children but was useful for the younger children. This therefore will be modified such that it will be used in consideration of the age group involved in the evaluation.

This chapter proves that some issues children encountered with the heuristic evaluation method as reported in chpter 5 could be avoided with the IMCH method. For example:

- The issues of understanding terminologies with the heuristics since they will produce their own game criteria
- The issues of understanding terminologies in the severity rating scale since the severity scale presented in this method is just a facial coloured severity scale which is a popular tool for children.

Also the IMCH method seem more successful as the problems they encountered bordered more on the game being evaluated than on the method process. Though there are some issues that need addressing e.g.

1. Lengthy explanations that led to boredom for the older children (e.g. with the show and tell method)
2. Unclear game instructions that caused difficulties in doing the game task

3. A first time tablet user who is not attentive to instructions but just wants to play with the application.
4. The insufficiency in the steps of the severity scale that made final severity judgement difficult
5. Study break which led to confusion in the problems children reported previously

In view of the problems above it could be argued that the method is not fully made suitable for children. On the other hand, children's ability to find and rate usability problems and carry out the process in the correct order to find real usability problems could result in the claim that the method is effective. However, since it arguable from the data that the game criteria children produced did not inform their problem finding process, which is determined as an issue, the effectiveness of the method could also be argued as incomplete; although further modifications will be made to address this.

Following the issues highlighted the following modifications have been made to IMCH V3 to produce IMCH V4:

1. ***Lengthy explanations that led to boredom for the older children (e.g. with the show and tell method)***

In order to ensure an interactive session throughout the evaluation process, the age of the child evaluators will be considered to either include or not to include the show and tell session. If the children are younger children then the show and tell will be included. This is backed up as seen in chapter 2 that younger children aged 4 to 7years need more visual representations to carry out proper task. On the other hand, if they are older children 9 to 11 then this will be excluded to eliminate the issue.

2. ***Unclear game instructions that caused difficulties in doing the game task***

To make the game instruction clearer for subsequent studies, the sub games and levels in which task will be carried out will be clearly explained and possibly shown on the application unless children finding the levels and sub games themselves is part of what they should evaluate. To further ensure clarity, task could be written down and given to the children.

3. ***A first time tablet user who is not attentive to instructions but just wants to play with the application.***

Since the child's inattentiveness is attributed to his eagerness to play with the tablet which had already been given to him before the instructions are being given, this order will be swapped such that the instructions will be given first before the application or technology.

4. ***The insufficiency in the steps of the severity scale that made final severity judgement difficult***

In order to deal with the insufficiency in the steps of the severity scale, an additional two steps will be added to the current severity scale to produce a five step severity scale which will follow conventions of most severity scales (Jakob Nielsen, 1995; Yehuda & McGinn, 2007) which have more than three steps and also rating scales used with children e.g. the fun toolkit (Janet C. Read, 2007) and the thumbs up scale (Kano *et al*., 2010).

5. ***Study break which led to confusion in the problems children reported previously***

To address this problems, it will be ensured that the amount of time needed to complete a study is taken into consideration prior to starting the study and if the time is insufficient then the study will be left to be run in full in the next available sufficient time.

Finally in order to address the ***issue of the children using the produced game criteria to find problems***, this will be tackled in two ways:

- The game session prior to the game criteria session will be removed
- The children will be given opportunity to discuss all produced game criteria before evaluating the chosen game and they will be reminded during the evaluation to look at the game criteria at intervals.

The game session where children played a game for 5 minutes before the game criteria session will not be included to ensure a more generic game criteria is produced. This is because it is believed that the game children played (which is in the same genre as the criteria needed and also in the same genre as the game to be evaluated) biased the game criteria they produced and what they produced inclined more to that game concept making their stated criteria more specific rather than generic. For example, the game children played had a player who rides bike and at some point of the game, the player combatted an alien before moving to the next level. Children stated criteria such as: "There should be different bikes and you should be able to change" and "There should be an alien you have to defeat when you complete all levels". Also, considering game heuristics or guideline literature on how the heuristics are developed, the heuristics produced are usually generic to cater for games in that area. For example, the online multiplayer game heuristics by Pinelle *et al*., (2009) is generic enough to cater for most online multiplayer games in spite of the subject on which the game might be situated. Although this might be problematic if the subject of the game needs evaluation as well, however the use of more than one evaluation method could cater for this. Also, Korhonen & Koivisto (2006) playability mobile game heuristics is designed to cater for the playability aspect of mobile games not just a single type of mobile game.

Therefore, it is believed that a more generic criteria, allowing in depth discussion of all produced game criteria before the evaluation and reminder of the game criteria during the evaluation could help the children's problem link more to their produced game criteria which will further improve the effectiveness of the IMCH V4.

Further study in IMCH V4 will be carried out with children to retest the effectiveness and ease of use of the method (with the possibilities of eliminating observed problems) and to confirm problems children found. This will be reported in the next chapter (9)

8.3.1 Contribution of Chapter to Thesis

The first test of the IMCH (though in its third version) was carried out for the first time for ease of use and effectiveness.

Data from this study produced IMCH V4 as shown below:

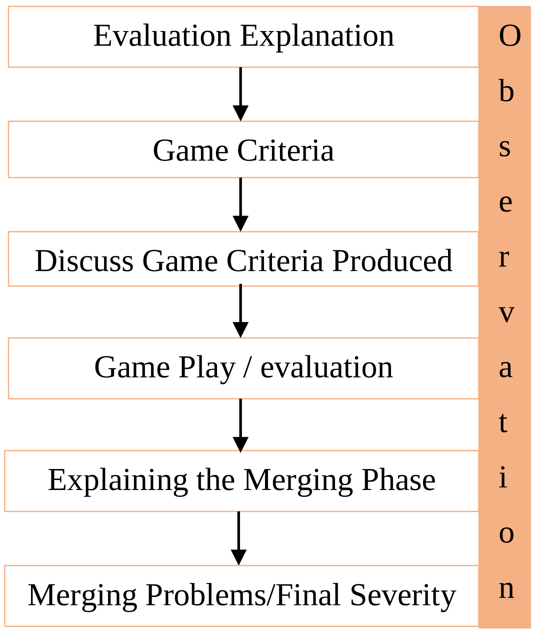

Figure 8.7 Version 4 of the IMCH

Table 8.22 Comparism of the Four Versions of the IMCH

STAGES OF THE METHOD	DIFFERENT VERSIONS OF THE IMCH METHOD				
	IMCH V1	IMCH V2	IMCH V3	IMCH V4	**OBSERVATION**
Stage 1	Explain the Evaluation Process / Method	Explain the Evaluation Process / Method	Explain the Evaluation Process / Method	Explain the Evaluation Process / Method	
Stage 2	**Narrate Story** so Children can come up with **Game Criteria**	**Narrate Story** so Children can come up with **Game Criteria**	Short Game Play	Narrate Story so Children can produce Game Criteria	
Stage 3	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **PLAIN** TLS severity scale	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale	**Narrate Story** so Children can come up with **Game Criteria**	**Discuss Game Criteria Produced**	
Stage 4	Explain Second (Merging) Phase	Explain Second (Merging) Phase	Show and Tell Actual Evaluation	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale	

Stage 5	Children to **merge problems** using post it note on a board and **provide final severity** for each merged problems	Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY FACE TLS severity scale again	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale	Explain Second (Merging) Phase	
Stage 6			Explain Second (Merging Phase)	Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY FACE TLS severity scale again	
Stage 7			Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY FACE TLS severity scale again		

9 CHAPTER NINE: CONFIRMING FINDINGS AND RETESTING IMCH

9.1 Introduction

The previous chapter successfully investigated and reported the suitability of IMCH V3 with children, although it also reported some observed problems children encountered which should be minimised or possibly eliminated to make the method more suitable indicating a partial success of the method. Therefore the study in this chapter aims *to test the effectiveness and ease of use of IMCH having made the modifications stated in section 8.3 and also to confirm issues children found from the study with IMCH V3*. The following are the objectives of this chapter:

- Test the effectiveness (to use the method correctly to find and predict real usability problems) of the method from the data gathered by the children
- Test the ease of use of the method (children's ability to understand the instructional language, tools used (e.g. severity scale) and finally are able to use the method as described) from observational data
- Compare data from IMCH V4 to that of IMCH V3 to confirm problems child evaluators previously reported.

This chapter is divided into 5 sections: section 9.2 will describe the IMCH V4, section 9.3 will state the method process, section 9.4 will report the findings, 9.5 will discuss the chapter, where 9.5.1 is stating the guideline for running this type of evaluation and 9.6 concludes the chapter and highlights what next.

9.2 IMCH V4 Described

9.2.1 What the method should do:

The IMCH V4 a modification of V3 is required to:

- Help children produce game criteria based on their values
- Help children use the game criteria as a tool to be more critical while evaluating the game
- Aid children to find real usability problems with the game criteria as a guide and based on what they consider as important (values)
- Collectively merge usability problems found, having a thoughtful discussion
- Allow children to successfully rate the severity of merged usability problems reported

9.2.2 How the method works:

The description of how the method works is schematically represented but also explained:

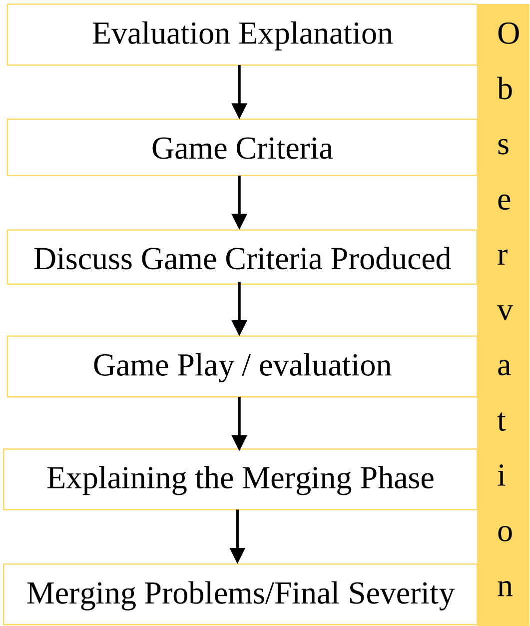

Figure 9.1 IMCH V4 Method Process

The figure above shows the flow of event during the IMCH V4 evaluation process. The method starts with explaining the method process to the children through to the game criteria session (where children will produce their own game criteria) down to them evaluating the system and finding usability problems until they merge their similar problems and give final severity. How each step should be carried out have been explained briefly (see bullet points below).

- Evaluation Explanation: The method will be explained to the children at the beginning of the evaluation
- Produce Game Criteria: A story will be narrated to the children then they will be asked to discuss and come up with what they think will make a good game for the genre of game they will evaluate. Response fro this session should form the game criteria. This session should last for 10 minutes
- Discuss Game Criteria: The game they will be evaluating will be explained to them then they will be asked to discuss their produced game criteria in view of this game.
- Game Play Preparations: The children will be provided with the task of the game, post it notes (to record usability problems), colour pen or marker (to rate problems), pen to

write and the severity scale will be placed at the front in view of all evaluators. There should either be a board or a sheet with which the game criteria will be written down.

- Evaluation Explanation: The game to be evaluated will be stated again, an explanation of the task and instruction for the method will be given.
- Individual Evaluation: The children will be provided with the technology which has the game on it with an iteration of the evaluation instructions, then they are allowed to individually play and evaluate the game, find problems, write down problems on the post it (one problem for one sheet) and colour the sheet with the marker or coloured pen to rate the problem (s) written down.
- Merging Problems: Children will come together to merge their problems where each child will need to post their found problems on the board or plain sheet. The same or similar problem are to be posted in the same area.
- Final Severity: Everyone who found the problem in a category will agree on a final severity which will be marked on the post it sheets. This will be the process until all the problems are merged. The facilitator should facilitate the merging and final severity phase.

9.3 Method

The IMCH V4 will be used in this study by the children to produce a value based game criteria and evaluate a game to find and rate real usability problems. An observational method will also be used to collect any issues children might encounter with the tools of the method or the method process itself; and also issues that might arise from facilitation.

9.3.1 Participant

Children from a UK primary school acted as the IMCH V4 evaluators. In total, 12 children from a year 5 class were recruited for the study. The children were called out of their classrooms by their teacher in groups of 4 which resulted in a total of three groups for the entire study. A researcher from the ChiCI research group (who is experienced in collecting observational data) acted as the study observer and the author facilitated the study.

9.3.2 Apparatus

Figure 9.2 Ice maths ninja game – a. Sub game selection page b. Actual game play for sub-zero mines sub game

The game criteria children produced during the study was written down on a white board using a marker. For the evaluation, children had to play and evaluate the ice maths ninja (arithmetic task) game (see fig 9.2) which had different sub games (same as the game used in chapter 8 since this study intends to confirm findings from the previous study); this was installed on an iPad. Children were each provided with a pack of post it note and pen to record usability problems found.

9.3.2.1 The Severity Scale

It was evident in chapter 5 that children struggled to understand Nielsen's (1995) severity scale which led to the development of the traffic light scale used in the IMCH V3 study reported in chapter 8. This also had issues as the steps (3 steps) provided was not sufficient to classify the severity of problems found. Therefore, it was decided that the steps will be extended, this was implemented in consideration of the following:

- Severity scales from literature
- Rating scales designed for children
- What the children did during the study to manage their need for more steps

9.3.2.1.1 Extending the Scale

In other to extend the points, severity rating scales from literature (Nielsen, 1995; Yehuda & McGinn, 2007) were investigated including the bad scale designed specifically for children (Salian, 2012). Also other rating scales used with children (Kano et al., 2010; Read, 2007) were also investigated and it shows majority of the scales had more than three points, though that excludes the bad scale by Salian, (2012). It is proven in literature that the fun toolkit is an established succesful rating scale with children, showing that children can understand and have the ability to use a five point scale. Therefore, the scale was extended to a five point scale.

9.3.2.1.2 Positioning the added severity points

It was observed from the children's data that when they were not satisfied with a severity point and did not think that the problem should be rated in the next point, they used both colours to create an in between rating (severity point). For example, if children think the problem is a bit worse than a green problem but not as bad as a yellow problem, they used green and yellow markers to indicate an intermediary point between green and yellow e.g. in the IMCH V3, in group 5, the problem "How to start a game" was rated with a green and yellow colour. This was also the case when they have problem that is between yellow and red e.g. in group 6 of the same IMCH V3 study, the problem "Game froze" was rated with a yellow and a red colour. Therefore it was decided that the two severity points created will be placed between the colours, i.e. one between green and yellow and the other between yellow and red. Also the between points were coloured with the colours surrounding them, see fig 9.3 below.

9.3.2.1.3 Deciding Text description for the new scale

In order to decide on the new five point severity scale, the previous text description was kept for the first three steps. The 'bad scale' text was considered which had "Awful" as the worst severity rate (Salian, 2012), as well as Yehuda & McGinn (2007) star scale which also had awful as its worst severity point. This word was therefore compared to the negative text of the fun toolkit which had "Awful" as its worst point too. It therefore confirms that children can understand the term "Awful" as a very bad instance. However, to arrive at the text for the last point, it was decided that Nielsen's text would be adopted only it will be made more general such that children could understand. Therefore the term "a disaster" was decided. Although this text and the colouring will be tested to investigate whether children can understand and use this five point scale and its text description.

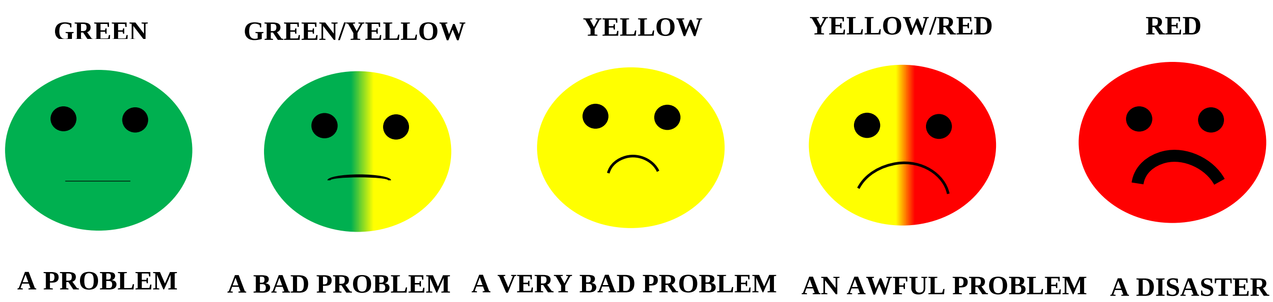

Figure 9.3 Coloured Face Severity Rating Scale

Coloured markers (red, yellow and green) were used by the children to rate problems, and the coloured face severity scale (see fig 9.4) which was posted in front was used as a guide for the appropriate severity colour. Finally children merged their problems on the white board and the facilitator used white board markers (red and green) and yellow highlighter to circle the merged problems as an indication of the final severity. When the evaluators decides and agree that an issue was not a problem at all, the facilitator circles it with a white board (black) marker.

- **Green:** *A Problem* – This is a very little problem that can be ignored if there is no time to fix
- **Green/Yellow:** *A Bad Problem* – This is a little problem that should be solved but with small importance. That is, if it is a problem the player can solve by him/herself without any help.
- **Yellow:** *A Very Bad Problem* – This is a medium problem that should be solved with medium importance. For example, it is a problem that the player needs help or direction to correct.
- **Yellow/Red:** *An Awful Problem* – This is a difficult problem that should be solved with high importance. It is a problem that can stop the player from playing although at this point the game is not yet written off.
- **Red:** *A Disaster* – This is a very difficult problem that must be solved to make it a good application before the game can be released otherwise it can be termed as not usable at all.

In spite of the above classifications made for the severity steps, judgement of the severity of problems are based on individual intuition and perception e.g. what player A judges as a red problem can be seen as a yellow/red problem to player B. Therefore, the severity scale is only a guide not an express same measure for all.

9.3.3 Procedure

The researcher started the study by making a self-introduction, then explained what the study is about and what the children are expected to do. The researcher started the game criteria session by asking the children if they had played games before, whether they still played games, and if they thought any of the games they had played in the past was a good game. The researcher then narrated a story (see story below) which was intended to motivate the children to come up with their own game criteria based on their values. This session lasted for 2 minutes

Story:

"There is a man who build games, he wants to build a maths game for children like yourselves and he is not sure about what will make a good maths game for children. When I thought about it, I realised I wasn't sure about what will make a good maths game too. So I decided to come and ask you directly on what you think will make a good maths game for children like yourselves. Based on the games you have played in the past, WHAT DO YOU THINK WILL MAKE A GOOD MATHS GAME FOR CHILDREN?"

Children were given the opportunity to state criteria for a good game and once the children stated a criterion, the researcher wrote it on an A4 sheet placed on the table before the children until the children indicated they had no more. This session lasted for 9 to 10 minutes.

The researcher explained to the children what the next session entails and provided each child with an iPad which had the game installed on it. The session required the children to play the ice maths ninja game and become critical of the game (based on their values, believes and from the game criteria discussed in the previous session), write issues they might encounter while playing the game or anything they predict to be a problem on a post it note and rate the game by colouring with the appropriate marker colour using a coloured face severity scale (see fig 9.3) as a guide.

Figure 9.4 The Evaluation Session (a) Child Evaluators Performing Evaluation (b) Coloured Face Severity Scale in View

The researcher handed the post it note, coloured markers (red, yellow and green) and pen to the children to do the evaluation. Since the game had different sub games, children were made to play 3 sub games. They were stopped after playing each sub game for approximately 3 to 4 minutes and reminded to write down and rate any issues before being told the next sub game to play. This was the process until all sub games were played and the session lasted for 10 to 12 minutes.

Merging Problems

In the next session the children were facilitated to merge similar problems and agree on a final severity. For this process, a child is picked to read out his/her first problem to the other children and posts it on a plain sheet at the middle of the table, then children with the same or similar problem post theirs one after the other on the same sheet. Furthermore, the children who had those problems were asked what final severity should the problem have and as they agreed on the final severity colour, the facilitator indicates the colour on the merged problems with the appropriate marker colour (see fig 9.5). This was the process until all the problems were merged. This session lasted for 5 minutes at the minimum and 7 minutes at the maximum; the more problems available to be merged, the more time spent in merging.

Figure 9.5 Children's Merged Problems with Final Severity

The picture (fig 9.5) represents the merge of similar problems by child evaluators in group 2. In this picture, the different problems on the five post it notes have been judged as similar by the child evaluators who have also agreed the final severity to be a red problem (the red big 'V' that has been made across the post it).

As a means to determine children's understanding of the coloured face severity scale, children were asked if they understood the step descriptions of the scale and were asked to explain the differences in the steps. This happened at the end of the study and lasted for 4 to 6 minutes depending on how quickly each group was able to make their explanations.

The entire study lasted for 30 minutes at the minimum and 37minutes at the maximum as the set up for the first study took longer than the others.

9.3.4 Data Analysis

Data analysis was carried out by the researcher in the same stage and using the same method as reported in the previous chapter (see section 8.2.1.4). Therefore only variations to the previous data analysis will be discussed.

9.3.4.1 Game Criteria

Game criteria for each group was carefully read through and merged within group, then each game criterion was coded in order to aid smooth and easy linking to usability problems found. The code identified is similar to that of the previous chapter though in this chapter an additional number was given to indicate a difference with the previous therefore the code was given thus: 2 (to indicate second IMCH study), G C (for Game Criteria) same as previous, Group Number (1 or 2 or 3) dependent on the group being coded and Letters (A-Z) were also given to ensure the uniqueness of each criterion (see table 9.1 and appendix 5 for full table).

The coded criteria were then compared to merged problems which had been coded using the same pattern to investigate whether the criteria the children stated informed their decision when they found problems, table to illustrate this is shown in table 9.2 below. These criteria were then merged between groups using the same method as previous (see table 9.3 for merged game criteria) and then themes were identified (this identification of themes is better explained below).

Table 9.1 Coded Game Criteria (EXTRACT)

S/No	GROUP 1	GROUP 2	GROUP 3
1	There should be different games for different maths problems i.e. you can either do addition or subtraction **2GC1A**	There should be different challenge levels **2GC2A**	There should theme **2GC3A**
2	There should be different difficulty levels **2GC1B**	There should be clear instructions **2GC2B**	It should have instructions **2GC3B**
3	You should be able to choose game type either maths only or maths and fun **2GC1C**	Different Maths Problem game (e.g. add, subtraction) **2GC2C**	It should have a background **2GC3C**
4	You need a catchy name **2GC1D**	It should be fun maths not boring maths **2GC2D**	It should have music **2GC3D**

Game Criteria produced by the child evaluators and presented according to their groups. Each game criteria have been coded as described above.

Table 9.2 Linking IMCH V4 Found Problems to Game Criteria

Merged Problem GROUP 1	**Linking Game Criteria**	**Merged Problem GROUP 2**	**Linking Game Criteria**	**Merged Problem GROUP 3**	**Linking Game Criteria**
2MP1A	2GC1M	2MP2A	X	2MP3A	2GC3G
2MP1B	X	2MP2B	2GC2B	2MP3B	X
2MP1C	2GC1B	2MP2C	2GC2B	2MP3C	2GC1N
2MP1D	X	2MP2D	2GC2E	2MP3D	X
2MP1E	2GC1N	2MP2E	2GC1N		
2MP1F	X				

Children’s merged problem being linked to their produced game criteria, game criteria from another group is indicated in red.

<u>KEY FOR SOME PROBLEMS AND LINKING CRITERIA</u>

(MP – Merged Problem and GC – Game Criteria)

Having linked children's problems found to the game criteria they stated it was determined that sometimes children usability problems found matched the game criteria from the same group e.g.

In group one

- **2MP1A:** It is not clear what to do **– 2GC1M**: There should be video tutorial or help button to teach someone who doesn't know how to play
- **2MP1C**: On sub-zero mines, if you get it, more zeros should appear **– 2GC1B**: There should be different difficulty levels

In group two

- **2MP2B**: I don't understand the game **– 2GC2B**: There should be clear instructions
- **2MP2C**: I don't understand why it is called cool 21 – **2GC2B**: There should be clear instructions

In group three

- **2MP3A**: Make sure you can pause it **– 2GC3G**: You should be able to pause the game

Some other times problem from a group didn't match game criteria from that group but rather to game criteria from another group e.g.

A group two problem matched a group one game criteria (**2MP2E**: It freezes after a while that you can't click retry **– 2GC1N**: You should be able to do it all over when you a level) and a group three's merged problem matched a group one's game criteria (**2MP3C**: It won't let you retry **– 2GC1N**: You should be able to do it all over when you a level).

Table 9.3 Game Criteria Merged between Groups

S/No	Merged Game Criteria Between Group	1	2	3	Total Group
1	There should be different games for different maths problems i.e. you can either do addition or subtraction	*	*		2
2	There should be lots of level so you don't finish it ASAP but let the difficulty levels be different, also you should be able to create your own level	*	*	*	3
3	It should be fun Maths not boring maths so you should be able to choose game type either maths only or maths and fun	*	*		2
4	You need a catchy name/theme	*		*	2
5	It should be the same as what is done in school	*			1
6	You should be able to pause the game			*	1
7	You should be able to log in, so you can save and continue later	*			1
8	Be able to challenge others (e.g. Friends)	*	*		2
9	You should be able to play it on different devices	*			1
10	You should be able to win prices (rewards) when you get right, your scores should also turn to prices that you can use to get stuff, change backgrounds and get power ups. And be able to take your reward away	*		*	2
11	It should be in a safe environment so people can't hack into it.			*	1
12	It should not just be questions to answer but it should be a quest i.e. it should have a storyline		*		1
13	It should have music			*	1
14	There should be video tutorial or help button to teach someone who doesn't know how to play and let the instructions be clear	*	*	*	3
15	You should be able to reset or make your own levels when you finish a level so you can do it all over.	*			1
16	There should be a bonus level after you got it right that will last for a limited time	*			1
17	There should be different modes and backgrounds	*		*	2
18	It should be for different years of children i.e. there should be different levels targeted at different age of children		*		1
19	You should be able to change language	*			1
20	It should be for boys and girls	*	*		2
21	You should be able to send what you have done to yourself and friends	*			1

9.3.4.1.1 *Identifying Themes for Game Criteria*

Themes from the previous chapter was used to categorise the criteria. Themes were noted down and each criterion was systematically crosschecked against each theme, where a criterion seem appropriate to go into a theme, this was placed in that theme. However, some criteria did not fit into any theme of the previous study so this was compared to heuristics from literature to identify an appropriate theme. For example, the 'User Control' theme which had two criteria ('You should be able to pause the game' and 'You should be able to log in, so you can save and continue later') classified into it, was identified from Alsumait & Al-Osaimi's (2009) ELearning heuristics although it is a category originally from Nielsen's heuristics. This was the order until all the criteria were classified into a theme (see table 9.4 for Extract and full table in appendix 5). The game criteria were further compared and linked to existing heuristics from literature, see table 9.4 and full table in appendix 5

Table 9.4 Game Criteria and Themes (EXTRACT))

S/No	Theme and Criteria	Group	Existing Heuristics
1	**Flexibility and Accessibility**		
A	There should be different games for different maths problems i.e. you can either do addition or subtraction (flexibility)	1,2	
B	You should be able to play it on different devices (flexibility)	1	The e-learning program may be used on a variety of equipment and platforms such as laptops, PDA. (Alsumait & Al-Osaimi, 2009)
C	You should be able to reset or make your own levels when you finish a level so you can do it all over. (flexibility)	1	Make the game Replayable (Federoff, 2002)
D	It should be for different years of children i.e. there should be different levels targeted at different age of children (Flexible)	2	
E	You should be able to change language (Flexibility and Accessibility)	1	
F	It should be for boys and girls (Flexibility)	1,2	

2	**Challenge**		
A	There should be lots of level so you don't finish it ASAP but let the difficulty levels be different, also you should be able to create your own level (Challenge)	1-3	There should be variable difficulty level (Federoff, 2002) Vary the difficulty level so that the player has greater challenge as they develop mastery. Easy to learn, hard to master. (Desurvire et al., 2004)

9.3.4.2 Child Evaluators' Usability Problems Data

The children's usability problems were analysed using the same technique as with the previous study. However, because it was discovered from reading the observer's data, the observer had not only captured problems children encountered with the method process and tools but has also noted problems children encountered but did not record, giving examples. These examples where then compared to children's data according to the group to confirm this and identify the total number of omitted problems (problems encountered but not written).

Data was read through and coded in order to link the problems to game criteria (see table 9.2). This coding was done exactly the same way the game criteria was done but in this case MP (merged problem) was used to replace GC (Game criteria) to differentiate the merged problem from the game criteria codes (see table 9.6). Data was then merged within and between groups (see table 9.5). Finally children's problems were observed to fit into 3 categories so were separated into these categories: Problem Encountered, Being Critical and False Alarm/Comment (Table 9.7)

Table 9.5 Merged Problems and Associated Groups

S/No	Merged Problems	Groups			Total group
		1	2	3	
1	I don't understand the game, it is not clear what to do	*	*		2
2	Sometimes the numbers get stuck behind the zeros and you can't get the numbers	*			1
3	When you lose, it freezes such that you can't go back to home or retry	*	*	*	3
4	On sub-zero mines, if you get it, more zeros should appear	*			1
5	A bit too easy for older children		*		1
6	When you make it sometimes it won't work as it is hard to drag the ice cubes	*		*	2
7	It keeps going on something else			*	1
8	Try not to let numbers go too fast as the ice cube numbers fall too fast		*		1
9	The music is good but it puts me off	*			1
10	Make sure you can pause it			*	1
11	I don't understand why it is called cool 21		*		1

A between group merge of all children's merged problems and an indication of the problem associated group.

Table 9.6 Coded Merged Problems and Groups

GROUP 1	GROUP 2	GROUP 3
It is not clear what to do **2MP1A**	Try not to let numbers go too fast as the ice cube numbers fall too fast **2MP2A**	Make sure you can pause it **2MP3A (false alarm)**
Sometimes the numbers get stuck behind the zeros and you can't get the numbers **2MP1B**	I don't understand the game **2MP2B**	When you make it sometimes it doesn't work **2MP3B**
On sub-zero mines, if you get it, more zeros should appear **2MP1C**	I don't understand why it is called cool 21 **2MP2C**	It won't let you retry **2MP3C**
The music is good but it puts me off **2MP1D**	A bit too easy for older children **2MP2D**	It keeps going on something else **2MP3D**
When you lose, it freezes such that you can't go back to home **2MP1E**	It freezes after a while that you can't click retry **2MP2E**	
It is hard to drag the ice cubes **2MP1F**		

Table 9.7 Children's Problem Category

Problem Encountered	**Being Critical**	**False Alarm/ Comment**
I don't understand the game, it is not clear what to do	A bit too easy for older children	Make sure you can pause it
Sometimes the numbers get stuck behind the zeros and you can't get the numbers	On sub-zero mines, if you get it, more zeros should appear	
When you lose, it freezes such that you can't go back to home or retry	The music is good but it puts me off	
Try not to let numbers go too fast as the ice cube numbers fall too fast		
It keeps going on something else		
When you make it sometimes it won't work as it is hard to drag the ice cubes		

9.3.4.3 Confirming Problems from Previous Study

In other to confirm problems from previous studies, each group's problem was compared to the result of the problem list of the previous study to identify similar or the same problem. In the instance of similarity this was judged as a confirmed problem (see table 9.8).

Table 9.8 Current Problems Confirming Previous Study Problems

Current Study Problems	**Current Study Group**	**Previous Study Confirmable Problems**
Don't understand as it is not clear what to do	1, 2	Don't Understand
When you finish it freezes and it won't let you press retry	1, 2 & 3	Game froze
It keeps going on something else	3	It keeps going on Facebook
A bit too easy for older children	2	Bit Childish
Try not to let numbers go too fast as the ice cube numbers fall too fast	2	Can't catch up with the falling ice

9.3.4.4 Identifying Severity Ratings for Problems Found

Table 9.9 Final Severity Ratings for Merged Problems

Severity Colours	**Problems Found (using Problem Codes)**	**Total**
Green	2MP1B, 2MP1C, 2MP1D, 2MP2A, 2MP2C, 2MP2D	6
Green/Yellow	2MP1F	1
Yellow	2MP1A, 2MP3A, 2MP3D	3
Yellow/Red	2MP2B, 2MP3C	2
Red	2MP1E, 2MP2E, 2MP3B	3

In order to determine how the child evaluators have rated their problems and if the five step scale made a difference, each merged problem severity was identified and put in the table 9.9

KEY TO THE CODED PROBLEMS

Green - On sub-zero mines, if you get it, more zeros should appear **(2MP1C)**

A bit too easy for older children **(2MP2D)**

Green/Yellow - It is hard to drag the ice cubes **(2MP1F)**

Yellow - It is not clear what to do **(2MP1A)**

It keeps going on something else **(2MP3D)**

Yellow/Red - I don't understand the game **(2MP2B)**

It won't let you retry **(2MP3C)**

Red - It freezes after a while that you can't click retry **(2MP2E)**

When you make it sometimes it doesn't work **(2MP3B)**

9.3.4.4.1 Children Explaining the Severity Scale

In order to determine children's understanding of the severity steps, they were asked to give an explanation of the steps, this is shown in table 9.10 below according to the groups

Table 9.10 Children's Explanation of the Severity Steps

Severity Colours	**Severity Steps**	**Group 1**	**Group 2**	**Group 3**
Green	**A Problem**	You can live with it	You can totally ignore it	This is an overlook
Green/Yellow	**A Bad Problem**	It is not so serious	It is just a tiny problem that the player can solve	It is not serious because it is a small problem
Yellow	**A very Bad Problem**	It is at the middle	It is a middle problem	It is a medium problem
Yellow/Red	**An Awful Problem**	It very bad and important to fix it, though you can use it but only a little	It is a bad problem that can stop the player from playing	It is a very bad problem that needs fixing though the game is not written off yet
Red	**A Disaster**	It is very bad, a big problem that it cannot be used unless it is fixed	It is a very very bad problem like a crash so you have to fix it	This is a very huge problem that will make the game not a good game at all. You have to fix this problem too

9.3.4.5 Observer's Data

After careful reading of the observational data, it showed observer had captured problems children encountered while playing the game as well as problems encountered during the entire study. It was also seen that in the observed problem children encountered while playing the game could be separated out into two categories: problem children recorded and problems they omitted. The latter was captured into a table (see table 9.9) since the forma was already recorded by the child evaluators.

Table 9.11 Observed Problems Encountered by Evaluators but was not Recorded

S/No	Observed Omitted Problem	Groups
1	Couldn't get his sound working	1
2	… IPad screen kept rotating the wrong way	1
3	'B' ended up loading Facebook from within the menu	2
4	Didn't know how to get back to the main menu from sub game start screen	2
5	Struggled to choose game options without being shown	3
6	Frostris "I don't get this"	3
7	Stated they had a problem but did not write it down	3

Also real problems required to assess of the method suitability for children was also captured and put in another table. In order to ensure uniqueness of data, similar data was merged (following the merging technique reported in the previous chapter) but the groups and problem area for which data was captured was noted (see table 9.10).

Table 9.12 The Required Observed Problems

S/No	Observed Problems Required	Problem Area	Group
1	Take a long time to contribute to the conversation	GC	1
2	D made point that has already been made by A	GC	1
3	'A' tried to help 'D' think about what they would like in a game	GC	2
4	'C' stated that it is difficult to think of what would make a good maths game	GC	2
5	Had to be told by A not to touch the Q'S on sub-zero game	GP	3
6	Stated they had a problem but did not write it down	MP	3
7	Struggled talking about SR's but did seem to understand them	SR	3

8	‘C’ had to read and explain the instructions and how to record the problem to ‘D’	GC	2
9	Talked about ratings as how good /easy/ hard the game was rather than about the problems	SR	1
10	Wrote second problem on the same post it note	GP	1, 2
11	‘C’ seemed to ‘control’ what ‘D’ did throughout the study	Other	2
12	C and D Wrote the same problem multiple times for different sub games	MP	2

9.4 Result

9.4.1 Game Criteria (Result)

Findings from the analysis show that children produced the sum of 38 game criteria with group 1 producing the highest number (21) of criteria and group 2 producing the least number (8) and group 3 having 9 criteria. It is also shown that 2 criteria were common to all three groups but some were unique to certain groups (see table 9.11).

Table 9.13 Summary of Children's Game Criteria

Groups	**1**	**2**	**3**	**All**	**Total**
Number of Criteria before merging within group	21	8	9	38	38
Number of Criteria Identified after merging within group	18	8	8	34	34
Number of Criteria Unique to group	10	2	3	2	18
Number of Criteria Unique to group after merging between group	7	2	3	21	33

Themes were identified for the criteria and it resulted in a total sum of 12 themes with ‘Flexibility and Accessibility’ as the theme with the highest number of criteria, Aesthetics and Design with 3, and User Control and Useful Reward, with each having 2 criteria categorised in them. All other themes had just one criterion in each (see table 9.12).

Table 9.14Game Criteria Themes Identified

S/No	**Game Criteria Themes**	**Number of Criteria**
1	Flexibility and Accessibility	6
2	Challenge	1
3	Fun	1
4	Aesthetics and Design	3
5	Match to Real World	1
6	User Control	2
7	Multiplayer	1
8	Useful Reward	2
9	Game Security	1
10	Game Story	1
11	Help and Instruction	1
12	Sharing	1

9.4.2 Child Evaluator's Data (Result)

The aim of this study is to investigate whether children can use the IMCH method again to find real usability problems which will confirm problems from previous study. Result show children actually found 5 problems which confirms some problems from the previous studies. These 5 problems were originally 8 (see table 9.13) but some problems were similar between groups and therefore were merged to produce 5 problems (see table 9.8 for actual problems and groups). A total of 29 individual problems were produced and 15 merged problems. Although, it was observed that children had recorded quite a number of problems, however, they had omitted some other problems encountered (see table 9.13). As an analysis procedure, the merged problems were further merged to produce 10 problems. These were categorised into 3 categories: Usability Problems Encountered, Being Critical and False Alarm. Result show the first had 6 problems, the second had 3 problems and the third category had only one problem (see table 9.7).

Table 9.15 Summary of Child Evaluators' Data

Groups	Individual Problem Reported	Merged Problem	Problems Omitted	Observed Problems	Similar Problem to Previous Study
1	9	6	2	9	2
2	13	5	2	13	4
3	7	4	3	8	2
Total	29	15	7	30	8

9.4.3 Result for Severity Rating

Table 9.9 showed children were able to rate all problems found and recorded, with 6 problems being rated as a green problem (a problem), 1 problem rated as a green/yellow problem (a bad problem), 3 problems being rated as a yellow problem (a very bad problem), 2 problems rated as a yellow/red problem (an awful problem) and 3 problems rated as a red problem (a disaster). In addition, the result from children's explanation of the severity scale showed all the groups were able to provide and explanation for each step of the scale with difference in the explanation (see table 9.10).

9.4.4 Observation Result

The sum of thirty problems were observed for the child bothering on different areas of the evaluation: Game criteria, Game play, Severity Rating, Merging Phase and Other.

Table 9.16 Number of Problems Reported in Different Problem Areas

Problem Area	**Groups**		
	1	**2**	**3**
Game Criteria	2	2	-
Game Play	6	9	6
Severity Rating	1	-	1
Merging Phase	-	1	1
Others	-	1	-

Result also show that after categorising observed problems into 3 categories: Observational Problem Required for the study, Child Evaluators' Omitted Problem and Evaluators' Recorded Problem, 12 of the observed problem which bother on problem areas stated above was reported for all three groups (see table 9.10), 7 Omitted Problems which was only on Game Play area was reported for all three groups (see table 9.9) and the last 11 problems were problems children encountered and reported.

9.5 Discussion

In this study Children were able to provide meaningful game criteria based on their value (what they consider as important) that bothers on learning, challenge, fun, security, interface appearance etc. Some of these criteria did match to existing heuristics though some didn't. However, it speculated that it is so because most heuristics are designed by adults, kept more general than being targeted to children's games. For example, for both studies (chapter 8 and 9) children have always recommended that the game be designed to suit both boys and girls (this was also suggested by the teachers in the focus group study of chapter 6) but this is hardly evident in heuristic literature. This could be argued as what the children consider as important for their type of technology.

Another point is how importantly children consider security to be that they desire that the game should be played in a safe environment that cannot be hacked. This however confirms the views of the CCI community (reported in chapter 2 of this thesis) on how children are changing and are becoming very connected to the changing technology. This is because in the past people use technology without bothering so much on security but these days, technology users are getting well informed about the need for safety (security) when using these technologies.

The procedure of the method, is for children to state good maths game criteria based on their views and afterwards play a maths game and find usability problems. It is believed that by discussing the game criteria, it will help the children become critical of the game where critical defined by the oxford dictionary online (Oxford-University-Press, 2015a) is stating an analysis of the merits and faults of a work of literature, music or art. This should be evident in their found problems being linked to their game criteria. Result show that more problems from this study linked to their game criteria unlike the previous study. It is therefore arguable that the process got them to think because it was stated from their criteria that a good game should have a pause button and one of the problems stated (though classified as false alarm) is that it should have a pause button. However, because the children did not carry out the linking process themselves it is difficult to make a claim on this, as it is also possible that their values which the method permits and their encounter during game play would have made the link possible.

9.5.1 Children's Problems Found

Children were able to find real problems that confirms problems from previous study however it was also evident from the result that a false alarm (it should have a pause button) was stated by a child. It is not certain on what caused the child's mentioning of this problem as the pause button is available and put in a visible area during game play, if the child did not know or understand the pause icon, or the child is affirming (being excited) that this game does have a pause button as is expected of good games. It is therefore revealed that just like other UIMs children have the tendency to produce false alarms using this method.

9.5.1.1 Critical Problem finding

It was evident with this method that children could become critical of their technology based on their values, as it was stated that though the music is good, it has a tendency to put one off, even when told that the volume button of the device could be used to turn down the music, the child evaluator related that such provisions be made within the game. Another child similar to the report given in the chapter 8 study stated that the game is a bit childish for older children. Though this game description and reviews states that the game is appropriate for all age group of children / children aged between 5 and 11 years old. Also a child expressed the dissatisfaction in the challenge of a sub game (sub-zero mines) stating that more zeros should be introduced after clearing a level. This therefore shows that when children are presented the opportunity to evaluate a game based on they consider important (their values) they have the tendency of becoming critical other than just providing them with a set of heuristics or guidelines to find problems.

9.5.1.2 Severity Ratings

Children's explanation of the severity ratings as shown in table 9.10 shows they understand the severity rate description and their rating of problems shows they understood their choice of severity rate and did not just make a random choosing as similar problems were rated with the same severity rate. E.g. in group one, the problem (When you lose, it freezes such that you can't go back to home **(2MP1E)**) is rated as a red problem the same as similar problems reported in group two (It freezes after a while that you can't click retry **(2MP2E)**) and group three (When you make it sometimes it doesn't work **(2MP3B))**

9.5.2 Observations

The observational data showed that though children can find real problems, they also omitted recording real problems encountered. This could either be children did not understand the implications of such problems as argued by MacFarlane & Pasiali, (2005) that the lack in children understanding or experience in usability evaluation could result in them undermining the implications of problems encountered. Also Donker and Reitsma argues that children might ignorantly omit to state problems because they think they are the cause of the problem and therefore are shy to state it (Donker & Reitsma, 2004).

The observational data also shows that children who cannot read or write and seats close to their friends have a tendency of being controlled by their friends therefore unable to produce honest opinion but rather will produce biased opinion.

It was also observed that children attempted to write more than one problem on a post it, which suggests that it is likely they didn't understand that process. Since for the previous study this type of issue was hardly reported unless for children who insisted on writing on one sheet, the show and tell would have made it easier for them to understand that process.

9.5.2.1 Comparing Observed Problems to HE and Previous

No observed problems from the previous study was observed in this study, however in comparison to the study in chapter 5 (HE with children) issue such as a child stating that it was difficult to state what will make a good game criteria could account for some children inability to think about such. However, because children finally produced the game criteria, could articulate and explain what is being discussed it could still be an easier measure compared to them trying to understand a guideline produced by adults in adult language and views.

Also, on one hand it was stated that a child was explaining the severity scale as how good, easy or bad the game was rather than the problems found which could be interpreted as the child's perception of the severity scale in connection to the game criteria section; because on the other hand it was observed that though children struggled to explain the severity scale, their explanation showed that they understood all the levels.

9.5.3 Comparing IMCH V4 Study to IMCH V3 Study

There were issues identified in IMCH V3 as reported in section 8.3 that led to carrying out IMCH V4, this issues include:

1. Lengthy explanations that led to boredom
2. Unclear game instructions that caused difficulties in doing tasks
3. A first time tablet user who is not attentive to instructions
4. Insufficiency in severity steps that made final severity judgement difficult
5. Study break which led to confusion in the problems.
6. There was also the concern that children's problems found didn't match their own group criteria and few between groups' game criteria.

With the IMCH V4 these issues were addressed thus:

Lengthy Explanations were avoided as the evaluation was facilitated with prepared script and the show and tell session was not included in the evaluation. However, for the usefulness of the show and tell session that involves giving the children visual aid of the correct process of carrying out the evaluation. It is recommended that this session could be included to show the children how to record the problems alone and this could be kept short to prevent boredom.

The game instructions were kept concise and child evaluators were shown on an iPad the particular games that was needed for the task.

There was no **first time tablet user** this time although evaluators were first given the instructions before they were provided with the tablets so this prevented the issue of inattentiveness to instructions given.

In order to address the issue of **insufficient severity steps**, an additional two steps were added to the three initial three steps scale producing a five point scale. In this study it was shown that all problems were rated and no observed problems of evaluators disagreeing on level to rate a problem during the merging phase.

Study break was not an issue here because all study was planned to fit in the available time such that no part of the study was broken.

During this study, children were given more time to discuss game criteria and during the evaluation children were reminded at intervals to think about the criteria they have mentioned and find problems using it. Result show children had more **link between their problems and their game criteria**. It could be attributed to the reminder but it is believed that the more time provided for children to discuss the criteria would have helped them articulate these points while looking for problems.

9.5.4 Judging Ease of Use and Effectiveness

Ease of Use: Going by the data gathered for the study, there was no explicit report of children having issues with the method process. However, from the children's data, it showed one child and stated a false alarm (a problem that did not exist), it is therefore possible that this child did not understand what is to be recorded as a problem, or just need to make a point given this was the same point made by the child during the game criteria session. This child will be considered as haven't understood the method process clearly. Another child was observed to explain the severity scale as for rating product other than for rating problems, It is either this child got it missed up while talking as the child found and rated real problems or the child does not clearly understand the method tool (severity scale). Therefore this child will also be considered as haven't clearly understood the method tool. There is also the case of a child stating that it was difficult to state a good game criteria and another child who was helped almost throughout the process. So four children out of 12 children is considered as not had ease with using the method i.e. 33.3% of the entire population while 66.7% was not seen as having issues. Though the percentage of children who is judged as those who had ease with using the method is higher, the percentage is lower compared to previous study, therefore future studies will not only look to mitigate problems that cause the difficulty in method use but also clear data will be collected for the purpose of clearly measuring this construct using more children.

Effectiveness: In this study, a total of 15 problems were reported where one is judged as false alarm therefore will be classified as unreal and 7 problems were observed to have been omitted by the children (misses). This therefore shows that 31.82% of problems that exist were missed by the child evaluators, 4.55% of the problem was an unreal problem and only 63.64% of real problems that exist were reported. Therefore it is recommended that in future children will be better trained to understand what constitutes problems that should be reported.

9.6 Conclusion

The aim of the study in this chapter was to test the effectiveness (children's ability to use the method correctly to find real usability problems) and ease of use (children's ability to understand the method process, instructions, tools used in the method and are easily able to perform the evaluation) of the IMCH V4 method, having made modifications to the V3 reported in chapter 8 giving some issues observed in IMCH V3 study. It was seen from the study that the V4 confirmed findings from V3 e.g. problems children reported as seen in table

9.8 and V4 was more effective i.e. children were able to produce more general game criteria, become more critical/ find more real problems using this version and also rate problems found without difficulties. However, it was observed that children were wrote/ attempted to write more than one problem on a post it, which leads to the conclusion that the show and tell may be useful to correct issues like this though the session should be kept minimal to avoid boring the children. Though this issue is not serious as it didn't obstruct the evaluation process.

This chapter has made report of the IMCH V4 study carried out with children and compared the outcomes to the outcomes of IMCH V3.

9.6.1 Contribution of Chapter to Thesis

This work in this chapter was to tackle issues that arose in chapter 8 to produce a more concrete IMCH with children. Therefore IMCH V4 was produced.

10 CHAPTER TEN: THE METHOD – Inspection Method for Children (IMCH) VERSION 5

10.1 Introduction

The method was designed using a user centred approach and a mixed method approach (as stated in chapter 4) which involved sequential explorative and iterative plan of studies and triangulation of data. This method development involved a total of 97 children who acted as evaluators or design participants, 11 teachers who made suggestions towards the method design, and 9 researchers (who acted as expert evaluators and observers) who are experienced in working with children and children's technologies. The method is designed around children's values such that children's production of game criteria, problem prediction and finding and problem severity judgement is based on their values. Just like any other evaluation method, the method might have its drawback which is the reason it is advised that more than one evaluation method be used to arrive at maximum result. However, the method is capable of guiding children through an inspection evaluation process, find and rate real usability problems (as show in chapters 8 and 9).

The purpose of this chapter is to describe the method and state its accompanying guidelines, which will be used by CCI researchers and technology design practitioners to work the method with children. The chapter is divided into four sections: section 10.2 will state what the method is supposed to do, 10.3 will describe how the method should work, 10.4 will outline the method guidelines and section 10.5 will conclude the chapter.

10.2 What the Method Should Do

The method is to be facilitated by an adult who will guide children to the evaluation where the method is expected to:

- Help children produce game criteria based on their values
- Help children use the game criteria as a tool to be more critical while evaluating the game
- Aid children to find and predict real usability problems with the game criteria as a guide and based on what they consider as important (value)
- Allow children individually rate individual usability problems found using the five steps coloured face severity scale (see fig 10.1)

- Collectively merge usability problems found, having thoughtful discussions while merging problems
- Allow children successfully rate the severity of usability problems found

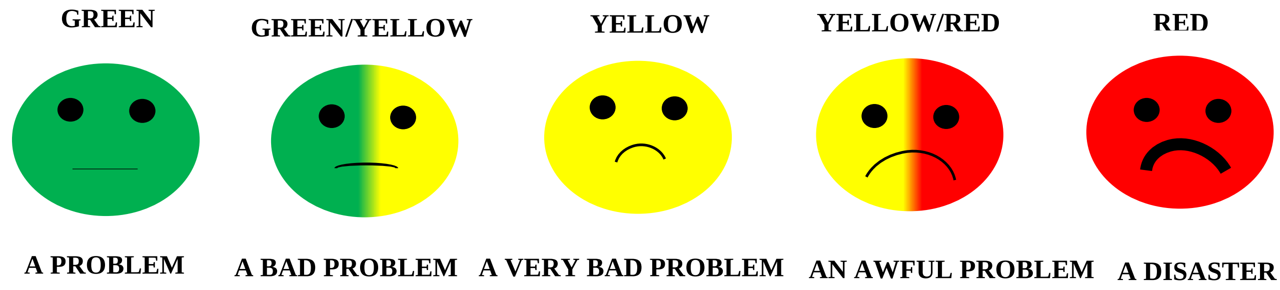

Figure 10.1 Coloured Face Severity Rating Scales

10.3 How the Method Works

The method goes from product familiarisation stage to final severity rating stage as shown in the figure below:

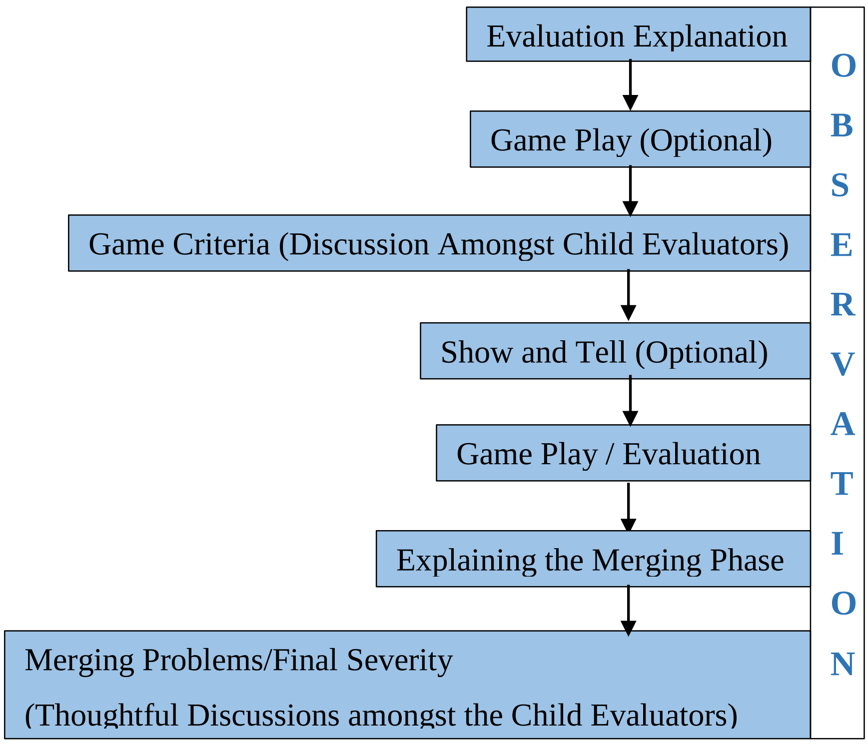

Figure 10.2 The IMCH V5 Process and Order

10.4 Guidelines for Carrying out the IMCH Evaluation with Children

In order to carry out the evaluation, an adult needs to be present to facilitate the session and narrate a story that will direct the children to do the evaluation. Some materials needs to be made available for the evaluation, this includes: 2 games, the application being evaluated, severity scale, coloured markers, post it note or nicely cut sheet that can contain just a problem, blu tacks, pen and sheet or blackboard. Also the evaluation should be carried out in the following stages:

- Explanation of the Method
- Play Game
- Produce Game criteria (in the same genre as the one to be evaluated)
- Show and tell session
- Play and evaluate game, find problems and give it a severity rate
- Merge problems between evaluators
- Give it a final severity rate

Explanation of the Method/Evaluation Process: The facilitator should state what the method is and explain the method process giving the children an overview of what they would do and how long the process is expected to last for.

Play Game: In the session, a good game is presented to the children to play. It is believed that the method is useful both for evaluation and for gathering design criteria from the children. If this is tended to gather design criteria, this session would be useful to inspire the children within a particular game genre. However, if the criteria being gathered intends to be generic then this session might create bias for the children towards a particular game genre or type for example as reported in the final point of section8.3 because children were given a bike based maths game for this session before the game criteria, most of the game criteria produced centred on this particular game. This session should last for 5 minutes at the least but if there are novices amongst the participant allowing them to play for longer may help them gather a little knowledge on the content of such games.

Producing Game Criteria: In this mini session, the facilitator should come up with a believable but not deceptive story that will drive the children to produce good game criteria. An example of a story is: *there is a man that wants to build a maths game for children but he is not sure about what will make a good maths game. When I thought about it, I realised I was not sure myself about what will make a good maths game for children like yourselves. So that's why I am here to ask you what you think will make a good maths game for children.* The story could be made age specific, gender specific or game type specific (a good maths football game) dependent on what is being gathered. The session is intended to make the children think about what a good game should be, preparing them for the game that would be evaluated (which is believed to have usability problems to find). The facilitator should either write down the criteria children will state on a sheet or on the board such that it is visible for the children to see, and should discuss the criteria stated with them as it was observed from the study reported in chapter 9 that children found problems that linked more to their game criteria and it is believed to be a possible result of the discussion. Sometimes time could be a constraint, this can therefore be discussed briefly rather than in details.

Show and Tell Session: In order to guide the children on how to correctly record and rate usability problems they will find during the evaluation, a show and tell session is recommended that should last for 2minutes. This will involve the facilitator doing an actual demonstration of the evaluation (from playing the game to finding problem and rating it using the severity scale) for the child evaluators to see and follow. This is especially useful for younger children because if the session is lengthy older children can become bored of not doing anything and just watching the session. This is evident with studies reported in chapter 8, the show and tell was used with year 6 children aged between 10 and 11 who found the session boring and was disengaged as seen with the observational data. However the session with younger children aged between 7 and 8 engaged with the session which had positive impact on the problem finding session. Also the session was not included in the retest of the IMCH (version 4) carried out with year 5 children (aged between 9 and 10) reported in chapter 9 and it showed children were having slight issues recording the problems correctly. However they carried out the process in the correct order.

Play and Evaluate Game for Usability Problems: Children are provided with the second game, which is believed to have usability problems that children should find. They are also given the pack of post it note or well cut out sheets (on which problems should be written), a pen or pencil for children to write problems down, the coloured markers (red, yellow/orange and green) to indicate the severity of the problem which could be done with just a stroke or dot of the intended severity colour. At this time, the severity scale should be placed at the front such that it is visible and legible for all the child evaluators to see. To aid easy merging process for the next sub session, children should be reminded to write one problem on one sheet as the size of a sheet (e.g. large sheet) could make a child write more than one problem on a sheet. This could make the merging phase more difficult especially if the two problems are different as separating those problems for merging with other evaluators could cause delay in the process.

Merging Problems Found Amongst Evaluators: While the severity scale is still kept visible, children should be facilitated to merge their problems. This merging could be done on a board or on a plain sheet on the table before the children. To do this process, any child should be picked to start first, if there are introvert children, they could be left to follow on later i.e. in this case the extrovert and bolder children could be made to go first. This is suggested by the independent teacher as reported in section 7.4.1 and was seen in both studies that bolder children were always more comfortable to start the merging phase especially if it involves the children walking up to the board. They could be asked on who will like to go first or the facilitator could just make a pick. When a child is to do the process, they should read their problem and then post it on the board or sheet while other goes through their sheets of problems to see if they have the same or similar problems. If similar problems are found, this or these should be posted together with the first problem.

Giving Final Severity Rate: After the problems have been merged, the evaluators should be asked on what final severity should be given to the problem. When they agree on a severity, the facilitator should indicate the agreed severity on the post it sheets with a board marker or any marker and keep the problems together for future use.

10.5 Conclusion

This chapter has stated what the IMCH V5 entails, what it should, explained how the method works and has finally highlighted the guidelines for running the method. The next chapter will conclude this thesis, outline limitations of the research thus far state the contributions and the future work of the research.

10.5.1 Contribution of Chapter to Thesis

This chapter is what could be seen as the manual of the method as the IMCH method was described with its guidelines to guide prospective researchers on how to use this method.

Table 10.1 Comparison of the Five Versions of the IMCH

STAGES OF THE METHOD	**DIFFERENT VERSIONS OF THE INSPECTION METHOD FOR CHILDREN (IMCH)**					
	IMCH V1	**IMCH V2**	**IMCH V3**	**IMCH V4**	**IMCH V5**	**OBSERVATION**
Stage 1	Explain the Evaluation Process / Method	Explain the Evaluation Process / Method	Explain the Evaluation Process / Method	Explain the Evaluation Process / Method	Explain the Evaluation Process / Method	
Stage 2	**Narrate Story** so Children can come up with **Game Criteria**	**Narrate Story** so Children can come up with **Game Criteria**	Short Game Play	**Narrate Story** so Children can **produce Game Criteria**	**Short Game Play (Optional)**	
Stage 3	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **PLAIN** TLS severity scale	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale	**Narrate Story** so Children can come up with **Game Criteria**	**Discuss Game Criteria Produced**	**Narrate Story** so Children can **produce Game Criteria** and **discuss Game Criteria**	
Stage 4	Explain Second (Merging) Phase	Explain Second (Merging) Phase	Show and Tell Actual Evaluation	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale	Show and Tell Actual Evaluation (Optional)	

Stage 5	Children to **merge problems** using post it note on a board and **provide final severity** for each merged problems	Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY FACE TLS severity scale again	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale	Explain Second (Merging) Phase	**Play and evaluate game** based on produced game criteria and **giving a severity rating** using the **SMILEY FACE** TLS severity scale	
Stage 6			Explain Second (Merging Phase)	Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY FACE TLS severity scale again	Explain Second (Merging Phase)	
Stage 7			Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY		Children to **merge problems** using post it note on a board and **Provide final severity** for each merged problems with SMILEY	

			FACE TLS severity scale again		FACE TLS severity scale again	

11 CHAPTER ELEVEN: CONCLUDING THE THESIS

11.1 Introduction

This work set out with an initial aim *to investigate whether older children can effectively carry out a heuristic evaluation on technologies designed for younger children if given appropriate tool.* However outcomes from initial studies (HE with children) and reports from similar studies from literature redirected the aim of the research *to investigate whether children can perform an effective usability inspection method (IMCH) evaluation on technologies designed for them based on their values*. With this aim, the work in this thesis followed an exploratory and empirical approach to achieve its aim which was carried out sequentially in four stages.

This chapter will summarise the work in this thesis, answer the research questions of this thesis and highlight the contributions of the work, placing the relevance of the work within the current state of CCI research. States the limitations and future work of the research and finally the work was concluded with a closing comments by the author.

The chapter is therefore divided into six sections: section 11.2 summarises this thesis work, 11.3 states each research question and how it was answered within the research, 11.4 states the contribution of the research, 11.5 highlights the limitations, 11.6 states the future work and 11.7 concludes the chapter with the author's closing comments.

11.2 Summary of Research

The focus of this work was to design a value based inspection method where children will act as the UIM evaluators. The work also set out to produce an evaluation guideline suited to children to perform the evaluation. Though adults (researchers and practitioners) will facilitate the evaluation using the guideline. During this exploratory work, a coloured face severity scale was developed with which children should rate problems reported during the evaluation. New processes to make adaptations to existing evaluation methods and produce an evaluation method suited to children was identified.

In order to do this, a user centred approach and a mixed method research approach was followed. Studies were carried out sequentially, data triangulation was used to arrive at the method process. The design of the method also went through iterative process and peer debrief sessions.

11.3 Answering the Research Questions

RQ1. Can children perform a heuristic evaluation?

RQ2. How can children's performance in the heuristic evaluation be assessed?

In order to answer RQ1 and RQ2, an in-depth literature was first carried out in HE to have an understanding of the method, how it is run and how participants' performances are assessed. In addition, research objective one (RO1) was carried out where an HE study was carried out with adult experts (people who have experience in performing a heuristic evaluation) and then with children. The first HE was a pilot study with the adults to have practical knowledge on how the HE works in practise. Then the second HE study was with children where observation was used to gather issues children might encounter with the method. With this, the suitability of the method for children was determined then children's data (usability problems) found was used to determine the effectiveness of the method (This is reported in chapter 5 of this thesis). It was evident from the study that the HE method was not yet suitable for children as children encountered problems while performing the evaluation and found very few usability problems as opposed to the number of problems found by adult in the study with adults. This proofs children's performance was poor in performing the heuristic evaluation.

RQ3. In the event of poor performance, what measures could be taken to produce a suitable UIM for children?

To answer RQ3, RO2 was carried out, where two focus groups were carried out with school teachers to gather input towards the modification of the HE method to make it suitable for children (This is reported in chapter 6). Ideas gathered from this study was explored and the decision to create a new method was made and then the first version of the method was created. Part of this first version was explored with children in a design session from which the second version of the method was created. Finally the second version was brought before an independent teacher (a teacher who was not part of the focus group) to scrutinise and make input towards the viability of the method for children. This produced the third version of the new method (IMCH) (This is reported in chapter 7) which was tested iteratively to determine the suitability of the method for the children and the effectiveness of the method as an evaluation method (This is reported in chapters 8 and 9). With this new method children were able to perform an inspection method evaluation and predict / found real problems.

RQ4. Can children's values be incorporated into the new UIM?

The decision to incorporate value into the new method was reached after reviewing teachers' input from chapter 6. In order to understand this concept and implement it in the IMCH method, then a literature review was carried out in social sciences to understand value as a concept and also in CCI to see how this has been included in CCI research (this is reported in chapter 7). After the review of these literatures, it was decided that value could be reflected from the start to the end of the method process by asking children to think about what they consider important when performing the evaluation: from producing the game criteria to the final severity rating.

11.4 Contributions of this Research

This research was able to produce a major and a minor contribution which is discussed below.

11.4.1 Major

The major contribution (MAC) from this research is the IMCH method and an accompanying guideline on how to use the method.

- Currently there are no validated inspection method published in CCI literature. As discussed in chapter three all current evaluation methods with children are user based with none in inspection method. Therefore, this research has produced a novel inspection method for children (IMCH) and an accompanying guidelines with which children can find and predict real usability problems based on their values and game criteria they produced.
- With this IMCH method, there is the avenue for children to produce game criteria based on their values which can inform and direct technology designers when designing technologies for children, that is give them insight to develop appropriate and needed technology for the children.
- As reported in chapter 9, the method does not only allow children predict usability problem but also become critical of the technology under evaluation to identify other issues which could help make the technology more suitable and fun for children.
- The method also provides an avenue for children to have thoughtful discussion during game criteria session (during this session children are able to produce not just general game criteria but value based criteria such that it is not currently captured by existing game heuristics e.g. that it is important that the game is for boys and girls) and when merging problems (this is a useful process for inspection method evaluation) which could help the game designer or researcher gain access into children's thought process and make quick decision of what is really important for them.

- It also allows the children to rate usability problems, so designers can have idea of problem priority when effecting changes to issues identified.
- Industries with tight budget could use the method with small number of children to determine the usability of children's technology and bigger industries could use it with small number of evaluators as a lead way to bigger evaluation decisions.
- This method paves way for CCI researchers (seasoned and new) to develop other inspection methods for children.

11.4.2 Minor

The minor contribution (MIC) is an insight of a process that could be used for the adaptation of evaluation method suited to children. There are different ways recorded in literature through which evaluation methods have been adapted to work with children which involves adults (researchers) designing the method and have children test its viability and suitability for them until the method is fully developed. However, the process undertaken within this research to design the IMCH is novel as there is no record of the involvement of multiple stakeholders (11 teachers, 97 children and 9 researchers who have experience in working with children) and the application of mixed methods in the adaptation of method to suit children especially within inspection method literature. This process could be adopted in future adaptations of method for children.

There are other contributions from the literature review that relates to assessing evaluation method for children and an outcome of the IMCH (e.g. its ability to facilitate discussion amongst the child evaluators and if girls and boys respond differently in this type of evaluation), which has not been followed up in this research. There is also an insight to how children's severity drawing was analysed and coded which could inform the analysis of children's drawing in the future.

11.5 Limitation of this Research

The major limitation of this research work was obtaining appropriate study time, as there was a huge challenge on getting teachers to partake in the focus group studies. Schools which participated gave initial dates for the study but due to business and tight in school schedule these dates were reviewed several times before a convenient date and time was given for the study to be held. Liaising with other schools was difficult, even after liaising with the head teachers (for example: schools could give an appointment date then just before then write to cancel appoints), it was also difficult to get a convenient time for 3 or more teachers to gather for the studies. This led to rescheduling until schools finally withdrew from participating. This caused a big setback for the research as subsequent studies couldn't be carried out until after the focus group studies were concluded and result collected.

Another constraint of this research was access to study groups, given research within ChiCI group is often carried out with school children at the time and convenience of the school, it was sometimes difficult to have access to the children such that their school activities won't be disrupted. Also there was the need to meet with more teachers for the focus group session but as stated earlier this was problematic as teachers had busy schedules to want to fit in other extra activities. This limitation in reaching more target group makes it difficult to make generalisation with findings gathered.

Given the location of my research focus, it was an obvious decision to work with school children in UK as instruments and curriculum used for studies are designed following the UK system. This therefore creates regional barriers as findings could not be extended beyond the system of UK.

Age of the CCI ranges between 5 and 12years (as reported in section 2.2) as opposed to the definition of children given in chapter 2. This however limits this work which focused on children 7 to 11years given the predictions of age appropriateness when carrying out usability testing with children as proposed by Hanna et al., (1997). In view of this, generalisation cannot be made from the findings of this work to cover for all children.

11.6 Future Work

Future work will look to explore the new method more in-depth to determine its ability to make children more critical when performing an evaluation. The method will also go through iterative refinement process to produce a final validated method within the CCI.

There were interesting issues that arose during the first IMCH study e.g. gender difference to respond to the method process this will be looked into to determine the credibility of such claim.

The method will also be tested with wider range and more children to allow for generalisation of the method ease of use and effectiveness of the method with children.

A comparative study will be carried out with a standard user testing method to determine the effectiveness of the IMCH over a user based method and it will be measured using multiple methods of assessing usability evaluation method.

11.7 Researcher's Comment

It is intended that the work and contributions in this thesis benefits the CCI community in the adaptation and development of rigorous evaluation methods to suit children especially in inspection method. Also processes undertaken in this research would be useful to not just seasoned researchers but also new and upcoming researchers in their work with children. Finally it is expected that the outcome of this research will be useful but to evaluators and designers of children's technologies.

It is indeed a pleasant experience to have worked with this class of users (children) as they are honest and enthusiastic group of participants whose contribution made the work in this thesis possible and therefore is endeared in this research and for future research.

It was also an informative and useful exercise and input gathering process for this research to have worked with teachers. As some of their ideas were really help and useful as they suggested though some other suggestions were not so useful for the new method creation but the extent of usefulness is worth the effort. Although recruiting teachers was also a time consuming measure.

12 BIBLIOGRAPHY

Alben, L. (1996). Quality of experience: defining the criteria for effective interaction design. *Interactions*, *3*(3), 11–15. doi:10.1145/235008.235010

Als, B. S., Jensen, J. J., & Skov, M. B. (2005). Comparison of think-aloud and constructive interaction in usability testing with children. In *Proceeding of the 2005 conference on Interaction design and children - IDC '05* (pp. 9–16). New York, New York, USA: ACM Press. doi:10.1145/1109540.1109542

Alsumait, A., & Al-Osaimi, A. (2009). Usability heuristics evaluation for child e-learning applications. In *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services - iiWAS '09* (p. 425). New York, New York, USA: ACM Press. doi:10.1145/1806338.1806417

Azevedo, S., & Ferreira, F. I. (2013). Participation and Learning in a Jenaplan School in the Netherlands: An Ethnographic Research with Children. *Procedia - Social and Behavioral Sciences*, *82*, 599–603. doi:10.1016/j.sbspro.2013.06.316

B. Brederode, P. Markopoulos, M. Gielen, A. Vermeeren, H. de Ridder. (2005). pOwerball: The design of a novel mixed-reality game for children with mixed abilities. In *Paper Presented at the IDC 2005, Boulder, Colorado*. Boulder, Colorado.

Baauw, E., Bekker, M. M., & Barendregt, W. (2005). A Structured Expert Evaluation Method for the Evaluation of Children's Computer Games. In M. F. Costabile & F. Paternò (Eds.), *IFIP TC13 International Conference on Human-Computer Interaction, INTERACT 2005* (Vol. 3585, pp. 457–469). Rome, Italy: Springer Berlin Heidelberg. doi:10.1007/11555261

Baauw, E., Bekker, M. M., & Markopoulos, P. (2006). Assessing the applicability of the structured expert evaluation method (SEEM) for a wider age group. In *Proceeding of the 2006 conference on Interaction design and children - IDC '06* (p. 73). New York, New York, USA: ACM Press. doi:10.1145/1139073.1139095

Baauw, E., & Markopoulous, P. (2004). A comparison of think-aloud and post-task interview for usability testing with children. In *Proceeding of the 2004 conference on Interaction design and children building a community - IDC '04* (pp. 115–116). New York, New York, USA: ACM Press. doi:10.1145/1017833.1017848

Bandura, A. (1971). Social learning theory. *Social Learning Theory*. New York: General Learning Press. doi:10.1111/j.1460-2466.1978.tb01621.x

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*(2), 248–287. doi:10.1016/0749-5978(91)90022-L

Barendregt, W. (2006). *Evaluating Fun and Usability in Computer Games with Children*. Eindhoven University of Technology.

Barendregt, W., Bekker, M. M., & Baauw, E. (2007). Development and evaluation of the problem identification picture cards method. *Cognition, Technology & Work*, *10*(2), 95–105. doi:10.1007/s10111-007-0066-z

Barendregt, W., Lindström, B., Rietz-Leppänen, E., Holgersson, I., & Ottosson, T. (2012). Development and evaluation of Fingu. In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12* (p. 204). New York, New York, USA: ACM Press. doi:10.1145/2307096.2307126

Barnard, Y. F., & Sandberg, J. A. C. (1994). *THE THINK ALOUD METHOD A practical guide to modelling cognitive processes* (First.). London: Academic Press, London.

Bekker, M. M., Baauw, E., & Barendregt, W. (2007). A comparison of two analytical evaluation methods for educational computer games for young children. *Cognition, Technology & Work*, *10*(2), 129–140. doi:10.1007/s10111-007-0068-x

Bernhaupt, R. (2010). *Evaluating User Experience in Games - Concepts and Methods*. (J. Karat & J. Vanderdonckt, Eds.) (Human-Comp.). London: Springer-Verlag London Limited. Retrieved from http://link.springer.com/content/pdf/10.1007/978-1-4471-2726-0.pdf

Blanchet-Cohen, N., & Reilly, R. C. (2013). Teachers' perspectives on environmental education in multicultural contexts: Towards culturally-responsive environmental education. *Teaching and Teacher Education*, *36*, 12–22. Retrieved from http://www.sciencedirect.com/science/article/pii/S0742051X13001054

Boom, J. (2004). Commentary on: Piaget's stages: the unfinished symphony of cognitive development. *New Ideas in Psychology*, *22*(3), 239–247. doi:10.1016/j.newideapsych.2004.11.002

Boyce, M. E. (1996). Organizational story and storytelling: a critical review. *Journal of Organizational Change Management*, *9*(5), 5–26. doi:10.1108/09534819610128760

Brodersen, C., & Iversen, O. S. (2007). Dressing up for school work: Supporting a collaborative environment with heterogeneous technologies. In *ECSCW 2007* (pp. 251–270). London: Springer London. Retrieved from http://link.springer.com/chapter/10.1007/978-1-84800-031-5_14

Bruckman, A., Bandlow, A., & Forte, A. (2003). HCI for kids. In J. A. Jacko & A. Sears (Eds.), *The human–computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (pp. 793–809). Lawrence Erlbaum and Associates. Retrieved from https://eva.fing.edu.uy/pluginfile.php/52761/mod_resource/content/0/biblioteca/Bruckman_Bandlow_-_HCI_for_Kids_-_IN_The_Human-Computer_Interaction_Handbook.pdf

Burns, R. B. (2000). *Introduction to Research Methods* (4th ed.). London, Thousand Oaks, New Delhi: Sage Publications Ltd.

Capra, M. G. /Advise.-S.-J. (2006). Usability problem description and the evaluator effect in usability testing. Retrieved from http://dl.acm.org/citation.cfm?id=1168269

Caracelli, V., & Greene, J. (1993). Data Analysis Strategies for Mixed Method Evaluation Designs. *Education Evaluation and Policy Analysis*, *15*(2), 195–207.

Chaitin, J. (2003). Narratives and Story-Telling. *Beyond Intractability*. Retrieved January 21, 2015, from www.beyondintractability.org/essay/narrative

Chan, H. K., He, H., Chung Wang, W. Y., Saarijärvi, H., Kuusela, H., & Spence, M. T. (2012). Using the pairwise comparison method to assess competitive priorities within a supply chain. *Industrial Marketing Management*, *41*(4), 631–638. Retrieved from http://www.sciencedirect.com/science/article/pii/S0019850111000903

Chattratichart, J., & Lindgaard, G. (2008). A comparative evaluation of heuristic-based usability inspection methods. In *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08* (p. 2213). New York, New York, USA: ACM Press. doi:10.1145/1358628.1358654

Clough, P., & Nutbrown, C. (2012). *A Student's Guide to Methodology* (3rd ed.). London, Thousand Oaks, New Delhi, Singapore: Sage Publications Ltd.

Cockton, G., Lavery, D., & Woolrych, A. (2002). Inspection-based evaluations, 1118–1138. Retrieved from http://dl.acm.org/citation.cfm?id=772072.772142

Cockton, G., & Woolrych, A. (2001). Understanding inspection methods: lessons from an assessment of heuristic evaluation. In A. Blandford & J. Vanderdonckt (Eds.), *in People and Computers* (XV., pp. 171–192). Springer-Verlag. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4471-0353-0_11

Creswell, J. W. (2009). *Research Design - Qualitative, Quantitative and Mixed Methods Approaches*. (V. Knight, S. Connelly, L. Habib, S. K. Quesenberry, & M. P. Scott, Eds.) (3rd ed.). Thousand Oaks, London, New Delhi, Singapore: Sage Publications, Inc.

Creswell, J. W., & Plano Clark, V. (2007). *Designing and Conducting Mixed Methods Research*. (L. C. Shaw, K. Greene, D. Santoyo, & C. Chilton, Eds.). Thousand Oaks, London and New delhi: Sage Publications, Inc.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research*. (V. Knight, L. Habib, A. Dodd, C. A. Hoffman, B. Bauhaus, & M. Markanich, Eds.) (2nd ed.). Thousand Oaks, London, New Delhi, Singapore: Sage Publications, Inc.

Davies, M., & Hughes, N. (2014). *Doning a Successful Research Project - Using Qualitative or Quantitative Methods*. (Aardvark Editorial Limited, Ed.) (2nd ed.). England, New York: Palgrave Macmillan.

Dawson, C. (2002). *Practical Research Methods : A User-friendly Guide to Mastering Research Techniques and Projects*.

Dawson-Tunik, T. L., Fischer, K. W., & Stein, Z. (2004). Do stages belong at the center of developmental theory? A commentary on Piaget's stages. *New Ideas in Psychology*, *22*(3), 255–263. doi:10.1016/j.newideapsych.2004.11.004

de Kock, E., van Biljon, J., & Pretorius, M. (2009). Usability evaluation methods. In *Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on - SAICSIT '09* (pp. 122–131). New York, New York, USA: ACM Press. doi:10.1145/1632149.1632166

Desurvire, H., Caplan, M., & Toth, J. A. (2004). Using heuristics to evaluate the playability of games. In *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04* (p. 1509). New York, New York, USA: ACM Press. doi:10.1145/985921.986102

DeVries, R. (2000). Vygotsky, Piaget, and Education: a reciprocal assimilation of theories and educational practices. *New Ideas in Psychology*, *18*(2-3), 187–213. doi:10.1016/S0732-118X(00)00008-8

Dindler, C., Eriksson, E., Iversen, O. S., Lykke-Olesen, A., & Ludvigsen, M. (2005). Mission from Mars. In *Proceeding of the 2005 conference on Interaction design and children - IDC '05* (pp. 40–47). New York, New York, USA: ACM Press. doi:10.1145/1109540.1109546

Dindler, C., Iversen, O. S., Smith, R., & Veerasawmy, R. (2010). Participatory design at the museum. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction - OZCHI '10* (p. 72). New York, New York, USA: ACM Press. doi:10.1145/1952222.1952239

Dodonov, Y. S., & Dodonova, Y. A. (2011). Basic processes of cognitive development: missing component in Piaget's Theory. *Procedia - Social and Behavioral Sciences*, *30*, 1345–1349. doi:10.1016/j.sbspro.2011.10.260

Donker, A., & Markopoulos, P. (2002). A comparison of think-aloud, questionnaires and interviews for testing usability with children. In *HCI* (pp. 305–316). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4471-0105-5_18

Donker, A., & Reitsma, P. (2004). Usability testing with young children. In *Proceeding of the 2004 conference on Interaction design and children building a community - IDC '04* (pp. 43–48). New York, New York, USA: ACM Press. doi:10.1145/1017833.1017839

Druin, A. (1996a). A place called childhood. *Interactions*, *3*(1), 17–22. doi:10.1145/223500.223506

Druin, A. (1996b). CHIkids. In *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96* (pp. 506–507). New York, New York, USA: ACM Press. doi:10.1145/238386.248153

Druin, A. (1999a). Cooperative inquiry. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99* (pp. 592–599). New York, New York, USA: ACM Press. Retrieved from http://dl.acm.org/citation.cfm?id=302979.303166

Druin, A. (1999b). The Role of Children in the Development of New Technology. Retrieved May 8, 2013, from http://hcil2.cs.umd.edu/trs/99-23/99-23.html

Druin, A. (2002). The role of children in the design of new technology. *Behaviour and Information Technology*, *21*(1), 1–25. doi:10.1080/0144929011010865

Druin, A., Bederson, B., Boltman, A., Miura, A., Knotts-Callahan, D., & Platt, M. (1998, October 15). Chapter 3: Children as Our Technology Design Partners+. Retrieved from http://drum.lib.umd.edu//handle/1903/947

Druin, A., & Solomon, C. (1995). Designing educational computer environment for children (abstract). In *Conference companion on Human factors in computing systems - CHI '95* (pp. 379–380). New York, New York, USA: ACM Press. doi:10.1145/223355.223731

Druin, A., & Solomon, C. (1996). Designing multimedia environments for children: computers, creativity, and kids. Retrieved from http://dl.acm.org/citation.cfm?id=235173

Druin, A., Stewart, J., Proft, D., Bederson, B., & Hollan, J. (1997). KidPad. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97* (pp. 463–470). New York, New York, USA: ACM Press. doi:10.1145/258549.258866

Duh, H. B.-L., & Chen, V. H.-H. (2010). Fantasies in narration. In *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology - ACE '10* (p. 3). New York, New York, USA: ACM Press. doi:10.1145/1971630.1971632

Dumas, J. S., & Redish, J. (1999). *A Practical Guide to Usability Testing*. Intellect Books. Retrieved from http://books.google.com/books?id=4lge5k_F9EwC&pgis=1

Duncker, K. (1945). On problem solving. In *The American Psychological Association*. Washington DC.

Ericsson, K. A., & Simon, H. A. (1993). Protocol Analysis:Verbal Reports as Data (revised edition). *MIT Press, Cambridge, Mass*.

Febretti, A., & Garzotto, F. (2009). Usability, playability, and long-term engagement in computer games. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09* (p. 4063). New York, New York, USA: ACM Press. doi:10.1145/1520340.1520618

Federoff, M. A. (2002). *Heuristics and Usability Guidelines for the Creation and Evaluation* *of Fun in Video Games*. Indiana University. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.89.8294

Feldman, D. H. (2004). Piaget's stages: the unfinished symphony of cognitive development. *New Ideas in Psychology*, *22*(3), 175–231. doi:10.1016/j.newideapsych.2004.11.005

Forlizzi, J., & Ford, S. (2000). The building blocks of experience. In *Proceedings of the conference on Designing interactive systems processes, practices, methods, and techniques - DIS '00* (pp. 419–423). New York, New York, USA: ACM Press. doi:10.1145/347642.347800

Frauenberger, C., Good, J., Keay-Bright, W., & Pain, H. (2012). Interpreting input from children. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12* (p. 2377). New York, New York, USA: ACM Press. doi:10.1145/2207676.2208399

Garrett, R. (2006). Critical storytelling as a teaching strategy in physical education teacher education. *European Physical Education Review*, *12*(3), 339–360. doi:10.1177/1356336X06069277

Gelderblom, H., & Kotzé, P. (2008). Designing technology for young children: What we can learn from theories of cognitive development. In *Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries riding the wave of technology - SAICSIT '08* (pp. 66–75). Wilderness, South Africa: ACM Press. doi:10.1145/1456659.1456668

Gilbert Cockton, Lavery, D., & Woolrych, A. (2003). Inspection based evaluations. In J.Julie & S. Andrew (Ed.), *The human-computer interaction handbook* (J.Julie & ., pp. 1118 – 1138). Erlbaum Associates Inc.

Gilutz, S. & Nielsen, J. (2002). Usability of websites for children: 70 design guidelines. *Nielsen Norman Group*. Retrieved March 28, 2015, from http://www.nngroup.com/reports/kids/

Gobo, G. (2011). Ethnography. In David Silverman (Ed.), *Qualitative Research* (3rd ed., pp. 15–34). London, Thousand Oaks, New Delhi, Singapore: Sage Publications Ltd.

Grimaldi, S., Fokkinga, S., & Ocnarescu, I. (2013). Narratives in design. In *Proceedings of the 6th International Conference on Designing Pleasurable Products and Interfaces - DPPI '13* (p. 201). New York, New York, USA: ACM Press. doi:10.1145/2513506.2513528

Guha, M. L., Druin, A., Chipman, G., Fails, J. A., Simms, S., & Farber, A. (2004). Mixing ideas. In *Proceeding of the 2004 conference on Interaction design and children building a community - IDC '04* (pp. 35–42). New York, New York, USA: ACM Press. doi:10.1145/1017833.1017838

Guha, M. L., Druin, A., & Fails, J. A. (2013). Cooperative Inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction*, *1*(1), 14–23. doi:10.1016/j.ijcci.2012.08.003

Gutman, J. (1982). A means-end chain model based on consumer categorization processes. *Journal of Marketing*, *46*(2), 60–72.

Gutman, J. (1991). Exploring the nature of linkages between consequences and values. *Journal of Business Research*, *22*(2), 143–148. doi:10.1016/0148-2963(91)90048-3

Haigh, C., & Hardy, P. (2011). Tell me a story--a conceptual exploration of storytelling in healthcare education. *Nurse Education Today*, *31*(4), 408–11. doi:10.1016/j.nedt.2010.08.001

Hanna, L., Neapolitan, D., & Risden, K. (2004). Evaluating computer game concepts with children. In *Proceeding of the 2004 conference on Interaction design and children building a community - IDC '04* (pp. 49–56). New York, New York, USA: ACM Press. doi:10.1145/1017833.1017840

Hanna, L., Risden, K., & Alexander, K. (1997). Guidelines for usability testing with children. *Interactions*, *4*(5), 9–14. doi:10.1145/264044.264045

Harpur, A. P., Boyce, K. E., & McConnell, N. C. (2013). An investigation into the circumstances surrounding fatal dwelling fires involving very young children. *Fire Safety Journal*, *61*, 72–82. doi:10.1016/j.firesaf.2013.08.008

Hartson, H., Andre, T., & Williges, R. (2001). Criteria for evaluating usability evaluation methods. *International Journal of Human ….* Retrieved from http://www.idemployee.id.tue.nl/g.w.m.rauterberg/lecturenotes/0H420/hartson.pdf

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behaviour & Information Technology*, *25*(2). doi:10.1080/01449290500330331

Hertzum, M., & Jacobsen, N. (2003). The evaluator effect: A chilling fact about usability evaluation methods. *International Journal of Human- …*, *15*(1), 183–204. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S15327590IJHC1304_05

Hertzum, M., & Jacobsen, N. E. (2001). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-Computer Interaction*, *13*(4), 421–443. doi:10.1207/S15327590IJHC1304_05

Hocko, J. (2002). *Categorizing the "Badness" of Usability Problems*. Waltham, MA.

Holdford, D. (2008). Content analysis methods for conducting research in social and administrative pharmacy. *Research in Social and Administrative Pharmacy*, *4*(2), 173–181. Retrieved from http://www.sciencedirect.com/science/article/pii/S1551741107000241

Hollingsed, T., & Novick, D. G. (2007). *Usability inspection methods after 15 years of research and practice*. *Proceedings of the 25th annual ACM international conference on Design of communication - SIGDOC '07*. New York, New York, USA: ACM Press. doi:10.1145/1297144.1297200

Hornbæk, K., & Frøkjær, E. (2008). Comparison of techniques for matching of usability problem descriptions. *Interacting with Computers*, *20*(6), 505–514. doi:10.1016/j.intcom.2008.08.005

Horton, M. (2012). *Improving Validity and Reliability in Children's Self Reports of Technology Use*. University of Central Lancashire.

Hourcade, J. P., Perry, K. B., & Sharma, A. (2008). PointAssist: helping four year olds point with ease. In *Proceedings of the 7th international conference on Interaction design and children - IDC '08* (p. 202). New York, New York, USA: ACM Press. doi:10.1145/1463689.1463757

Inkpen, K. M. (2001). Drag-and-drop versus point-and click mouse interaction styles for children. *ACM Transactions on Computer-Human Interaction*, *8*(1), 1–33. doi:10.1145/371127.371146

ISO 9241-11. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) -- Part 11: Guidance on usability*. Switzerland. Retrieved from http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883

ISO 9241-210. (2010). *Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems*. Retrieved from https://bsol.bsigroup.com/en/Bsol-Item-Detail-Page/?pid=000000000030263992

Iversen, O. S., Buur, J., & N, D.-A. (2003). User Centred Design through the Keyhole : Video Design Cases. In *Proceedings of IFIP TC 13 International Conference on Human-Computer Interaction* (pp. 431–438). Zurich, Switzerland: IOS Press.

Iversen, O. S., Halskov, K., & Leong, T. W. (2010). Rekindling values in participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference on - PDC '10* (p. 91). New York, New York, USA: ACM Press. doi:10.1145/1900441.1900455

Ivory, M. Y., & Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, *33*(4), 470–516. doi:10.1145/503112.503114

Jacobsen, N. E., Hertzum, M., & John, B. E. (1998). The evaluator effect in usability tests. In *CHI 98 conference summary on Human factors in computing systems - CHI '98* (pp. 255–256). New York, New York, USA: ACM Press. doi:10.1145/286498.286737

Jacobsen, N., Hertzum, M., & John, B. (1998). The evaluator effect in usability studies: Problem detection and severity judgments. In *… of the Human Factors and …* (pp. 1336–1340). Chicago. Retrieved from http://pro.sagepub.com/content/42/19/1336.short

Jaime Montemayor, Allison Druin, Jim Hendler, A. S. (1999). PETS: A Personal Electronic Teller of Stories. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.4627

Jeffries, R., Miller, J., Wharton, C., & Uyeda, K. (1991). User interface evaluation in the real world: a comparison of four techniques. *Proceedings of the SIGCHI …*, *91*(c). Retrieved from http://dl.acm.org/citation.cfm?id=108862

Jensen, J. J., & Skov, M. B. (2005). A review of research methods in children's technology design. In *Proceeding of the 2005 conference on Interaction design and children - IDC '05* (pp. 80–87). New York, New York, USA: ACM Press. doi:10.1145/1109540.1109551

Johnstone, C. J., Bottsford-Miller, N. A., & Thompson, S. J. (2006). Using the Think Aloud Method (Cognitive Labs) To Evaluate Test Design for Students with Disabilities and English Language Learners. *Technical Report 44 University of Minnesota, National Center on Educational Outcomes*. Retrieved June 3, 2013, from http://www.cehd.umn.edu/nceo/OnlinePubs/Tech44/

Jones, T. (1991). An Empirical Study of Children's Use of Computer Pointing Devices. *Journal of Educational Computing Research*, *7*(1), 61–76.

Kafka, T., Economos, C., Folta, S., & Sacheck, J. (2011). Children as subjects in nutrition research: a retrospective look at their perceptions. *Journal of Nutrition Education and Behavior*, *43*(2), 103–9. doi:10.1016/j.jneb.2010.03.002

Kakiashvili, T., Koczkodaj, W. W., & Woodbury-Smith, M. (2012). Improving the medical scale predictability by the pairwise comparisons method: Evidence from a clinical data study. *Computer Methods and Programs in Biomedicine*, *105*(3), 210–216. Retrieved from http://www.sciencedirect.com/science/article/pii/S0169260711002586

Kano, A., Horton, M., & Read, J. C. (2010). Thumbs-up scale and frequency of use scale for use in self reporting of children's computer experience. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI '10* (p. 699). New York, New York, USA: ACM Press. doi:10.1145/1868914.1869008

Kearney, R. (2002). *On Stories* (1st ed.). London; New York: Routledge. Retrieved from http://eds.a.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzcyMzQwX19BTg2?sid=418ed1b4-bd15-4061-9556-515ddb2479c6@sessionmgr4004&vid=0&format=EB&rid=1

Kelly, S. R., Mazzone, E., Horton, M., & Read, J. C. (2006). Bluebells: a design method for child-centred product development. In *Proceedings of the 4th Nordic conference on Human-computer interaction changing roles - NordiCHI '06* (pp. 361–368). New York, New York, USA: ACM Press. doi:10.1145/1182475.1182513

Kesselring, T., & Müller, U. (2011). The concept of egocentrism in the context of Piaget's theory. *New Ideas in Psychology*, *29*(3), 327–345. doi:10.1016/j.newideapsych.2010.03.008

Knox, C. A., & Burkhart, P. V. (2007). Issues related to children participating in clinical research. *Journal of Pediatric Nursing*, *22*(4), 310–8. doi:10.1016/j.pedn.2007.02.004

Köffel, C., & Haller, M. (2008). Heuristics for the evaluation of tabletop games. Evaluating User Experiences in Games,. In *Workshop at the 2008 Conference on Human Factors in Computing Systems. CHI '08.* Florence, Italy. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.99.3219

Korhonen, H. (2011). The explanatory power of playability heuristics. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology - ACE '11* (p. 1). New York, New York, USA: ACM Press. doi:10.1145/2071423.2071473

Korhonen, H., & Koivisto, E. M. I. (2006). Playability heuristics for mobile games. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services - MobileHCI '06* (p. 9). New York, New York, USA: ACM Press. doi:10.1145/1152215.1152218

Korhonen, H., & Koivisto, E. M. I. (2007). Playability heuristics for mobile multi-player games. In *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts - DIMEA '07* (p. 28). New York, New York, USA: ACM Press. doi:10.1145/1306813.1306828

Krippendorff, K. (1980). Content Analysis: An Introduction to Its Methodology. London: Sage Publications Ltd.

Krueger, R. A. (1994). *Focus Group Interview: A Practical Guide for Applied Research* (Second Edi.). Thousand Oaks, CA: Sage Publications, Inc.

Krueger, R. A., & Casey, M. A. (2000). *Focus Groups: A practical guide for applied research*. (D. C. Laughton, E. Carr, D. E. Axelsen, & V. Cheng, Eds.)*Focus Groups: A practical guide for applied research* (3rd ed.). California; London; New Delhi: Sage Publications Ltd.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX Curve: A method for evaluating long-term user experience. *Interacting with Computers*, *23*(5), 473–483. doi:10.1016/j.intcom.2011.06.005

Lazar, D. J., Feng, D. J. H., & Hochheiser, D. H. (2010). *Research Methods in Human-Computer Interaction*. John Wiley & Sons. Retrieved from http://books.google.com/books?id=H_r6prUFpc4C&pgis=1

Lewis J.R. (2006). Usability Testing. In N. Hoboken (Ed.), *In G. salvendy (Ed.) Handbook of Human Factor and Ergonomics* (3rd ed., pp. 1275 – 1316). John Wiley.

Liben, L. S. (Ed.). (1983). *Piaget and the Foundations of Knowledge*. Hillsdale, New Jersy, London: Lawrence Erlbaum Associates.

Ling, C., & Salvendy, G. (2009). Effect of evaluators' cognitive style on heuristic evaluation: Field dependent and field independent evaluators. *International Journal of Human-Computer Studies*, *67*(4), 382–393. doi:10.1016/j.ijhcs.2008.11.002

Lourenço, O. (2012). Piaget and Vygotsky: Many resemblances, and a crucial difference. *New Ideas in Psychology*, *30*(3), 281–295. doi:10.1016/j.newideapsych.2011.12.006

MacFarlane, S., & Pasiali, A. (2005). Adapting the Heuristic Evaluation Method for Use with Children. In *Interact*. Rome. Retrieved from http://chici.uclan.ac.uk/references/adapting_the_heuristic_evaluation.pdf

MacFarlane, S., Sim, G., & Horton, M. (2005). Assessing usability and fun in educational software. In *Proceeding of the 2005 conference on Interaction design and children - IDC '05* (pp. 103–109). New York, New York, USA: ACM Press. doi:10.1145/1109540.1109554

Mack, R., & Montaniz, F. (1994). Observing, predicting and analysing usability problems. In *Usability inspection methods* (pp. 295 – 339). John Wiley & Sons.

Maier, H. W. (1978). The Cognitive Theory of Jean Piaget. In W. W. Boehm (Ed.), *Three Theories of Child Development* (3rd ed., pp. 12–70). New York, Hagerstown, San Francisco, London: Harper & Row, Publishers.

Malone, T., & Lepper, M. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning, *3*, 223 – 253.

Malone, T. W. (1982). Heuristics for designing enjoyable user interfaces. In *Proceedings of the 1982 conference on Human factors in computing systems - CHI '82* (pp. 63–68). New York, New York, USA: ACM Press. doi:10.1145/800049.801756

Marco, J., Cerezo, E., Baldasarri, S., Mazzone, E., & Read, J. C. (2009). Bringing tabletop technologies to kindergarten children. In *BCS-HCI, 2009*. Cambridge, UK: BCS.

Markopoulos, P., & Bekker, M. (2003a). Interaction design and children. *Interacting with Computers*, *15*(2), 141–149. doi:10.1016/S0953-5438(03)00004-3

Markopoulos, P., & Bekker, M. (2003b). On the assessment of usability testing methods for children. *Interacting with Computers*, *15*(2), 227–243. doi:10.1016/S0953-5438(03)00009-2

Markopoulos, P., & Bekker, M. M. (2002). How to compare usability tseting methods with children participants. Workshop Interaction Design for Children. In *Proceedings of the IFIP 9th International Conference on Human-Computer Interaction - INTERACT' 05* (pp. 153–158). Rome: Shaker Publishing.

Markopoulos, P., Read, J. C., MacFarlane, S., & Hoysniemi, J. (2008). Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers. Retrieved from http://dl.acm.org/citation.cfm?id=1386295

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396.

Mason, J., & Hood, S. (2011). Exploring issues of children as actors in social research. *Children and Youth Services Review*, *33*(4), 490–495. doi:10.1016/j.childyouth.2010.05.011

Matusov, E., & Hayes, R. (2000). Sociocultural critique of Piaget and Vygotsky. *New Ideas in Psychology*, *18*(2-3), 215–239. doi:10.1016/S0732-118X(00)00009-X

Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(2). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385

Mazzone, E. (2008). Determining value in informant design with children, 251–252. Retrieved from http://dl.acm.org/citation.cfm?id=1531826.1531909

Mazzone, E. (2012). *Designing with children: reflections on effective involvement of children in the interaction design process*. University of Central Lancashire.

McLeod, S. (2007). Lev Vygotsky - Developmental Psychology. *Simply Psychology*. Retrieved February 26, 2015, from http://www.simplypsychology.org/vygotsky.html

McLeod, S. (2009). Jean Piaget - Cognitive Theory. *Simply Psychology*. Retrieved February 23, 2015, from http://www.simplypsychology.org/piaget.html

McLeod, S. (2011). Bandura - Social Learning Theory. *Simply Psychology*. Retrieved February 26, 2015, from http://www.simplypsychology.org/bandura.html

Mokhtar, N. H., & Kamarulzaman, M. F. A. H. & S. (2011). The Effectiveness of Storytelling in Enhancing Communicative Skills. *Procedia - Social and Behavioral Sciences*, *18*, 163–169. doi:10.1016/j.sbspro.2011.05.024

Molich, R., McGinn, J. (Jen), & Bevan, N. (2013). You say "disaster", i say "no problem." In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13* (p. 301). New York, New York, USA: ACM Press. doi:10.1145/2468356.2468410

Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of the ACM*, *33*(3), 338–348. doi:10.1145/77481.77486

Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Morse, J. M. (2003). Principles of Mixed Methods and Multimethod Research Design. In *Handbook of Mixed Methods in Social and Behavioural Research* (pp. 189–208). Thousand Oaks, CA: Sage Publications, Inc.

National Story telling. (n.d.). Storytelling by the National Storytelling Network. *What is strorytelling*. Retrieved January 21, 2015, from www.storynet.org/resources/whatisstorytelling.html

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92* (pp. 373–380). New York, New York, USA: ACM Press. doi:10.1145/142750.142834

Nielsen, J. (1992). The usability engineering life cycle. *Computer, 25*(3), 12–22. doi:10.1109/2.121503

Nielsen, J. (1993a). *Usability Engineering*. San Francisco, CA: Morgan Kaufmann Publishers Inc. Retrieved from http://dl.acm.org/citation.cfm?id=529793

Nielsen, J. (1993b). *Usability Engineering*. USA: Morgan Kaufmann Publishers Inc.

Nielsen, J. (1994). Usability inspection methods. In *Conference companion on Human factors in computing systems - CHI '94* (pp. 413–414). New York, New York, USA: ACM Press. doi:10.1145/259963.260531

Nielsen, J. (1995). Severity Ratings for Usability Problems. *Jakob Nielsen's Alertbox*. Retrieved from http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

Nielsen, J., & Mack, R. L. (1994). Usability inspection methods. Retrieved from http://dl.acm.org/citation.cfm?id=189200

Norman, D. A., & Draper, S. W. (1986). *User centered system design: new perspectives on human-computer interaction*. Lawrence Erlbaum Associates. Retrieved from http://books.google.co.uk/books/about/User_centered_system_design.html?id=wt1QAAAAMAAJ&pgis=1

NSPCC. (2013). Legal definition of a child | NSPCC. *NSPCC Factsheet*. Retrieved September 22, 2014, from http://www.nspcc.org.uk/Inform/research/briefings/definition_of_a_child_wda59396.html

Obrist, M., Moser, C., Fuchsberger, V., Tscheligi, M., Markopoulos, P., & Hofstätter, J. (2011). Opportunities and challenges when designing and developing with kids @ school. In *Proceedings of the 10th International Conference on Interaction Design and Children - IDC '11* (pp. 264–267). New York, New York, USA: ACM Press. doi:10.1145/1999030.1999078

Oosterholt, R., Kusano, M., & de Vries, G. (1996). Interaction design and human factors support in the development of a personal communicator for children. In *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96* (pp. 450–457). New York, New York, USA: ACM Press. doi:10.1145/238386.238603

Oosterholt, R., Kusano, M., & Vries, G. De. (1996). Interaction design and human factors support in the development of a personal communicator for children. In *CHI* (pp. 450–457). Vancouver BC,: ACM. doi:0-89791-777-4/96/04

Oxford-University-Press. (2015a). Critical. *Language Matters*. Retrieved March 14, 2015, from http://www.oxforddictionaries.com/definition/english/critical

Oxford-University-Press. (2015b). Storytelling. *Oxford English Dictionary*.

Oxford-University-Press. (2015c). Values. *Oxford Dictionaries*. Retrieved January 23, 2015, from www.oxforddictionaries.com/definition/english/value

Paavilainen, J. (2010). Critical review on video game evaluation heuristics. In *Proceedings of the International Academic Conference on the Future of Game Design and Technology - Futureplay '10* (p. 56). New York, New York, USA: ACM Press. doi:10.1145/1920778.1920787

Pardo, S., Vetere, F., & Howard, S. (2006). Teachers' involvement in usability testing with children. In *Proceeding of the 2006 conference on Interaction design and children - IDC '06* (p. 89). New York, New York, USA: ACM Press. doi:10.1145/1139073.1139097

Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods*. (D. C. Laughton, V. Novak, D. E. Axelsen, K. Journey, & K. Peterson, Eds.) (3rd ed.). California; London; New Delhi: Sage Publications, Inc.

Piaget, J. (1971). *Science of Education and the Psychology of the Child* (cased.). London: Longman Group Limited.

Pinelle, D., Wong, N., & Stach, T. (2008). Heuristic evaluation for games: usability principles for video game design. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems - CHI '08* (p. 1453). New York, New York, USA: ACM Press. doi:10.1145/1357054.1357282

Pinelle, D., Wong, N., Stach, T., & Gutwin, C. (2009). Usability heuristics for networked multiplayer games. In *Proceedinfs of the ACM 2009 international conference on Supporting group work - GROUP '09* (p. 169). New York, New York, USA: ACM Press. doi:10.1145/1531674.1531700

Posner, B. Z., & Munson, J. M. (1979). The importance of values in understanding organizational behavior. *Human Resource Management*, *18*(3), 9–14. doi:10.1002/hrm.3930180303

Read, J. C. (2005). The ABC of CCI. *Interfaces*, *62*(Spring 2005), 8–9. Retrieved from http://www.bcs.org/upload/pdf/interfaces62.pdf

Read, J. C. (2007). Validating the Fun Toolkit: an instrument for measuring children's opinions of technology. *Cognition, Technology & Work*, *10*(2), 119–128. doi:10.1007/s10111-007-0069-9

Read, J. C., & Bekker, M. M. (2011). The Nature of Child Computer Interaction. In *BCS-HCI '11 Proceedings of the 25th BCS Conference on Human-Computer Interaction* (pp. 163–170). Swinton.

Read, J. C., Horton, M., Sim, G., Gregory, P., Fitton, D., & Cassidy, B. (2013). CHECk. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13* (p. 187). New York, New York, USA: ACM Press. doi:10.1145/2468356.2468391

Read, J. C., Hourcade, J. P., Markopoulos, P., & Iversen, O. S. (2014). Child computer interaction SIG. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14* (pp. 1135–1138). New York, New York, USA: ACM Press. doi:10.1145/2559206.2559214

Read, J. C., & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction, 81–88. doi:10.1145/1139073.1139096

Read, J. C., & Markopoulos, P. (2013). Child–computer interaction. *International Journal of Child-Computer Interaction*, *1*(1), 2–6. doi:10.1016/j.ijcci.2012.09.001

Read, J. C., Markopoulos, P., & Druin, A. (2011). A community for child computer interaction. In *CHI 2011,*. Vancouver BC: ACM. Retrieved from http://chi2011.org/communities/child-computer-interaction/CCI.pdf

Read, J. C., Sim, G., Gregory, A. J., Xu, D., & Ode, J. B. (2013). Children Designing Serious Games. *EAI Endorsed Transactions on Game-Based Learning*, *13*(1-12), 1–9. doi:10.4108/trans.gbl.01-06.2013.e5

Read, J., & Fine, K. (2005). Using survey methods for design and evaluation in child computer interaction. In *Workshop on Child Computer Interaction: …*. Retrieved from http://www.researchgate.net/publication/228651205_Using_survey_methods_for_design_and_evaluation_in_child_computer_interaction/file/3deec518618d104b10.pdf

Read, J., MacFarlane, S., & Casey, C. (2001). Measuring the usability of text input methods for children. *People and Computers XV—Interaction …*, 559–572. doi:10.1007/978-1-4471-0353-0

Read, J., MacFarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. *… Design and Children*, *2*. Retrieved from http://www.researchgate.net/publication/228870976_Endurability_engagement_and_expectations_Measuring_children's_fun/file/3deec518618d0828ce.pdf

Resnick, M. (1991). Multilogo: a study of children and concurrent programming. *Interactive Learning Environment*, *1*(3), 158–170.

Revelle, G. L., & Strommen, E. F. (1990). The effects of practice and input device used on young children's computer control. *Collegiate Microcomputer*, *8*(4), 33–41. Retrieved from http://dl.acm.org/citation.cfm?id=99021.99026

Röcker, C., & Haar, M. (2006). Exploring the Usability of Video Game Heuristics for Pervasive Game Development in Smart Home Environments. In C. Magerkurth, M. Chalmers, S. Bjork, & L. Schafer (Eds.), *Proceedings of Third International Workshop on Pervasive Gaming Applications (PerGames'06)* (pp. 124–131). Dublin, Ireland.

Rode, J., Blythe, M., & Nardi, B. (2012). Qualitative research in HCI. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12* (p. 2803). New York, New York, USA: ACM Press. doi:10.1145/2212776.2212725

Rubin, J., & Chisnell, D. (2011). *Handbook of Usability Testing: Howto Plan, Design, and Conduct Effective Tests (Google eBook)*. John Wiley & Sons. Retrieved from http://books.google.com/books?id=l_e1MmVzMb0C&pgis=1

S. Papert. (1980). *Mindstorms, Children, Computers and Powerful Ideas*. Great Britain,: Basic Books.

S. Papert. (1988). The conservation of Piaget: The computer as a grist to the constructivist mill. In G. Forman & P. Pufall (Eds.), *Constructivism in the Computer Age* (pp. 3–13). Hillsdale, NJ: Lawrence Erlbaum Associates.

Salian, K. (2012). *Heuristic Evaluation by Older Children: Is the current method feasible for them?* Preston.

Salian, K., Sim, G., & Read, J. (2013). Can Children Perform a Heuristic Evaluation? In *APCHI '13*. Bangalore, India: ACM. Retrieved from http://www.researchgate.net/publication/258006951_Can_Children_Perform_a_Heuristic_Evaluation/file/9c960526948bd5bf00.docx

Scaife, M., & Rogers, Y. (1998). Kids as informants: telling us what we didn't know or confirming what we knew already?, 27–50. Retrieved from http://dl.acm.org/citation.cfm?id=303430.303434

Scaife, M., Rogers, Y., Aldrich, F., & Davies, M. (1997). Designing for or designing with? Informant design for interactive learning environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97* (pp. 343–350). New York, New York, USA: ACM Press. doi:10.1145/258549.258789

Schaffer, N. (2007). *Heuristics for Usability in Games White Paper Noah Schaffer , Certified Usability Assistant (CUA) April 2007*. Retrieved from http://gamesqa.files.wordpress.com/2008/03/heuristics_noahschafferwhitepaper.pdf

Scholtz, J. (2004). Usability Evaluation User-Centered Evaluations. In *Encyclopedia of Human Computer Interaction*. IAD National Institute of Standards and Technology.

Schwartz, S. H. (1994). Are There Universal Aspects in the Structure and Contents of Human Values? *Journal of Social Issues*, *50*(4), 19–45. doi:10.1111/j.1540-4560.1994.tb01196.x

Scott, J. (2000). Children as Respondents: The Challenge for Quantitative Methods. In P. Christensen & A. James (Eds.), *Research with Children – Perspectives and Practices* (pp. 98–119). London: Falmer Press.

Sears, A. (1997). Heuristic Walkthroughs: Finding the Problems Without the Noise. *International Journal of Human-Computer Interaction, 9*(3), 213–234. doi:10.1207/s15327590ijhc0903_2

Silver, D. (2001). Songs and storytelling: bringing health messages to life in Uganda. *Education for Health (Abingdon, England), 14*(1), 51–60. doi:10.1080/13576280010015362

Sim, G. (2009). *Evidence Based Design of Heuristics : Usability and Computer Assisted Assessment by*. University of Central Lancashire.

Sim, G. (2012). Designing the anti-heuristic game. In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12* (p. 308). New York, New York, USA: ACM Press. doi:10.1145/2307096.2307153

Sim, G., Cassidy, B., & Read, J. C. (2013). Understanding the fidelity effect when evaluating games with children. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13* (pp. 193–200). New York, New York, USA: ACM Press. doi:10.1145/2485760.2485769

Sim, G., MacFarlane, S., & Read, J. (2006). All work and no play: Measuring fun, usability, and learning in software for children. *Computers & Education*, *46*(3), 235–248. doi:10.1016/j.compedu.2005.11.021

Skov, M. B., & Stage, J. (2005). Supporting problem identification in usability evaluations. In *Proceedings of the 17th Australia conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*. Canberra, Australia.

Sousa, P., Fonseca, H., Gaspar, P., & Gaspar, F. (2015). Usability of an internet-based platform (Next.Step) for adolescent weight management. *Jornal de Pediatria*, *91*(1), 68–74. doi:10.1016/j.jped.2014.05.010

Strommen, E. F. (1993). Preschool Children at the Interface: A Cognitive Model of Device Difficulty. In *ERIC* (pp. 990–999). Education Resource Information Centre. Retrieved from http://files.eric.ed.gov/fulltext/ED362206.pdf

Subrahmanyam, K., Kraut, R. E., Greenfield, P. M., & Gross, E. F. (2000). The Impact of Home Computer Use on Children's Activities. *The Future of Children - CHILDREN AND COMPUTER TECHNOLOG*, *10*(2), 123–144.

Tashakkori, A., & Creswell, J. W. (2007). The New Era of Mixed Methods. *Journal of Mixed Methods Research*, *1*(1), 3–7. doi:10.1177/2345678906293042

Tashakkori, A., & Teddlie, C. (2003). *Handbook of Mixed Methods in Social and Behavioural Research*. Thousand Oaks, CA: Sage Publications, Inc.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of Mixed Methods Research*. (V. Knight, S. Connelly, L. Habib, S. K. Quesenberry, & C. Duksta, Eds.) (1st ed.). Los Angeles, London, New Delhi, Singapore, Washington DC: Sage Publications, Inc.

Tu, W., & Zhou, X.-H. (2000). Pairwise comparisons of the means of skewed data. *Journal of Statistical Planning and Inference*, *88*(1), 59–74. Retrieved from http://www.sciencedirect.com/science/article/pii/S0378375899002062

UNICEF. United Nations Convention on the Right of a Child (1990). London: UN. Retrieved from http://www.unicef.org.uk/Documents/Publication-pdfs/UNCRC_PRESS200910web.pdf

Väänänen-Vainio-Mattila, K., & Wäljas, M. (2009). Developing an expert evaluation method for user eXperience of cross-platform web services. In *Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era on - MindTrek '09* (p. 162). New York, New York, USA: ACM Press. doi:10.1145/1621841.1621871

van Kesteren, I. E. H., Bekker, M. M., Vermeeren, A. P. O. S., & Lloyd, P. A. (2003). Assessing usability evaluation methods on their effectiveness to elicit verbal comments from children subjects. In *Proceeding of the 2003 conference on Interaction design and children - IDC '03* (p. 41). New York, New York, USA: ACM Press.

doi:10.1145/953536.953544

Vanden Abeele, V., Zaman, B., & De Grooff, D. (2011). User eXperience Laddering with preschoolers: unveiling attributes and benefits of cuddly toy interfaces. *Personal and Ubiquitous Computing*, *16*(4), 451–465. doi:10.1007/s00779-011-0408-y

Vermeeren, A. P. O. S., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction Extending Boundaries - NordiCHI '10* (p. 521). New York, New York, USA: ACM Press. doi:10.1145/1868914.1868973

Villalta, M., Gajardo, I., Nussbaum, M., Andreu, J. J., Echeverría, A., & Plass, J. L. (2011). Design guidelines for Classroom Multiplayer Presential Games (CMPG). *Computers & Education*, *57*(3), 2039–2053. doi:10.1016/j.compedu.2011.05.003

Vissers, J., De Bot, L., & Zaman, B. (2013). MemoLine. In *Proceedings of the 12th International Conference on Interaction Design and Children - IDC '13* (pp. 285–288). New York, New York, USA: ACM Press. doi:10.1145/2485760.2485836

Vygotsky, L. S. (1978). *Mind in Society: The development of Higher Psychological Proceses*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, Massachusetts, London: Harvard University Press.

Vygotsky, L. S. (1997). Interaction between Learning and Development. In M. Gauvain & M. Cole (Eds.), *Readings on the development of Children* (Second., pp. 29–36). New York: W H Freeman and Company. Retrieved from http://www.psy.cmu.edu/~siegler/vygotsky78.pdf

Wang, D., Li, J., Zhang, J., & Dai, G. (2008). A pen and speech-based storytelling system for Chinese children. *Computers in Human Behavior*, *24*(6), 2507–2519. doi:10.1016/j.chb.2008.03.014

Wang, W., & Rubart, J. (2006). A cognitive and social framework for shared understanding in cooperative hypermedia authoring. In *Proceedings of the seventeenth conference on Hypertext and hypermedia - HYPERTEXT '06* (p. 53). New York, New York, USA: ACM Press. doi:10.1145/1149941.1149953

WebFinance. (2015). Values. *Business Dictionary*. Retrieved January 23, 2015, from www.businessdictionary.com/definitions/values.html

Wodike, O. A., Sim, G., & Horton, M. (2014). Empowering Teenagers to Perform a Heuristic Evaluation of a Game. In *Proceedings of the 28th International BCS Human Computer Interaction Conference (HCI 2014)*. Southport, UK: British Computer Society. Retrieved from http://ewic.bcs.org/content/ConWebDoc/54238

Woods, R. L., Satgunam, P., Bronstad, P. M., & Peli, E. (2010). Statistical analysis of subjective preferences for video enhancement. In B. E. Rogowitz & T. N. Pappas (Eds.), *Proc. of SPIE-IS&T Electronic Imaging* (Vol. 7527, pp. 75270E–75270E–1–10). Human Vision and Electronic Imaging XV,. doi:10.1117/12.843858

Woolrych, A., & Cockton, G. (2002). Testing a Conjection based on the DR-AR Model of UIM Effectiveness. *Proceedings of HCI, 2,* 30 – 33.

Xu, D., Mazzone, E., & MacFarlane, S. (2006). In search for evaluation methods for children's tangible technology. In *Proceeding of the 2006 conference on Interaction design and children - IDC '06* (pp. 171–172). New York, New York, USA: ACM Press. doi:10.1145/1139073.1139079

Xu, D., Read, J., Sim, G., Mcmanus, B., & Qualter, P. (2009). Children and'smart'technologies:

can children's experiences be interpreted and coded? In … *People and Technology* (pp. 224–231). Retrieved from http://dl.acm.org/citation.cfm?id=1671038

Yehuda, H., & McGinn, J. (2007). Coming to terms. In *CHI '07 extended abstracts on Human factors in computing systems - CHI '07* (p. 1899). New York, New York, USA: ACM Press. doi:10.1145/1240866.1240918

Zaman, B. (2005). Evaluating games with children. In *Proceedings of Interact 2005 Workshop on Child …* (pp. 9–11). Retrieved from http://soc.kuleuven.be/com/mediac/cuo/admin/upload/Zaman_Evaluating games with children.PDF

Zaman, B. (2007). Introducing contextual laddering to evaluate the likeability of games with children. *Cognition, Technology & Work*, *10*(2), 107–117. doi:10.1007/s10111-007-0067-y

Zaman, B. (2009a). Introducing a pairwise comparison scale for UX evaluations with preschoolers. *Human-Computer Interaction–INTERACT 2009*, 634–637. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-03658-3_68

Zaman, B. (2009b). Introduction and validation of a pairwise comparison scale for UX evaluations and benchmarking with preschoolers. In *User Experience Evaluation Methods in Product …*. Retrieved from https://lirias.kuleuven.be/handle/123456789/268687

Zaman, B., Vanden Abeele, V., Markopoulos, P., & Marshall, P. (2009). Tangibles for children,. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09* (p. 4729). New York, New York, USA: ACM Press. doi:10.1145/1520340.1520727

Zuckerman, O., Arida, S., & Resnick, M. (2005). Extending tangible interfaces for education. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05* (p. 859). New York, New York, USA: ACM Press. doi:10.1145/1054972.1055093

13 APPENDICES

The following are appendices of document used for studies and data that have been collected during studies in this research. These have been represented according to the chapters where the relevant data is presented and document used. Below are list of the appendix outline:

Appendix 1 – Chapter 5 documents, this include:

- **A.** Heuristic Set/Severity Rating Scale
- **B.** Evaluators' Data Collection Form (EDCF)
- **C.** Merging Form
- **D.** Observers' Data Collection Form
- **E.** Evaluators' Data
- **F.** Observers' Data

Appendix 2 – Chapter 6 documents which include:

- **A**. Consent Form
- **B**. Heuristic Method Description Sheet

Appendix 3 – Chapter 7 Documents

Appendix 4 – Chapter 8 Documents

Appendix 5 – Chapter 9 Documents

Appendix 1 Chapter 5 Documents

13.1 Appendix 1 Chapter 5 Documents

A. Playability Game Heuristics and Severity Rating Scales

PLAYABILITY GAME HEURISTICS by Hannu Korhonen and Elina M.I. Koivisto

	Mobility Heuristics
1.	The game and play sessions can be started Quickly
2.	The game accommodates with the surroundings
3.	Interruptions are handled reasonably
	Gameplay Heuristics
4.	The game provides clear goals or supports player created goals
5.	The player sees the progress in the game and can compare the results
6.	The players are rewarded and rewards are meaningful
7.	The player is in control
8.	Challenge, strategy, and pace are in balance
9.	The first-time experience is encouraging
10.	The game story supports the gameplay and is meaningful
11.	There are no repetitive or boring tasks
12.	The players can express themselves
13.	The game supports different playing styles
14.	The game does not stagnate
15.	The game is consistent
16.	The game uses orthogonal unit differentiation4
17.	The player does not lose any hard-won possessions
	Game Usability Heuristics
18.	Audio-visual representation supports the game
19.	Screen layout is efficient and visually pleasing
20.	Device UI and game UI are used for their own purposes
21.	Indicators are visible
22.	The player understands the terminology
23.	Navigation is consistent, logical, and minimalist
24.	Control keys are consistent and follow standard conventions
25.	Game controls are convenient and flexible
26.	The game gives feedback on the player’s actions
27.	The player cannot make irreversible errors
28.	The player does not have to memorize things unnecessarily
29.	The game contains help

SEVERITY RATING SCALE by Jakob Nielsen

0 = I don't agree that this is a usability problem at all
1 = Cosmetic problem only: need not be fixed unless extra time is available on project
2 = Minor usability problem: fixing this should be given low priority
3 = Major usability problem: important to fix, so should be given high priority
4 = Usability catastrophe: imperative to fix this before product can be released

Appendix B: **EVALUATORS' USABILITY PROBLEM COLLECTION FORM** IDENTITY ☐ MALE ☐ FEMALE ☐

Problem found (write a single problem in each space)	**Heuristic(s) violated (Insert heuristic no)**	**Task (Where it was violated)**	**How was it found**	**Severity Rating (0 – 4)**
(e.g. game freezes after first level)	14	☐ Creating a player	☐ Scanning for problems	3
		☐ Navigating within the game UI	☐ Systematically searching for problems	
		☒ Actual game play	☐ Trying to force errors	
		☐ Finishing the game play	☒ Following users task	
		☐ Creating a player	☐ Scanning for problems	
		☐ Navigating within the game UI	☐ Systematically searching for problems	
		☐ Actual game play	☐ Trying to force errors	
		☐ Finishing the game play	☐ Following users task	
		☐ Creating a player	☐ Scanning for problems	
		☐ Navigating within the game UI	☐ Systematically searching for problems	
		☐ Actual game play	☐ Trying to force errors	
		☐ Finishing the game play	☐ Following users task	

Appendix C: **EVALUATORS PROBLEM MERGING FORM**

Group Identity [ ] Number in Group [ ]

Problem Categories	Frequency (How many Evaluators Identified this problem?)	Severity Rating (0 - 4)

Appendix D: **OBSERVERS' DATA CAPTION FORM1** Observer's name__________________________ Evaluator Group ☐

Evaluator	Problem Statement	What area is the problem statement related to? (Please tick as appropriate)					
		HS	Task	DCF	EI	SR	Others
e.g. 1 or A	Unable to allocate a heuristic no to problem found	✓					

FULL MEANINGS (KEY)

- HS – Heuristic Set
- DCF – Data Collection Form,
- EI – Evaluation Instruction: - Instruction for the whole evaluation process – How to fill the data collection form and what the whole process entails.
- SR – Severity Rating,
- Task – Step by step instruction on how to play the game
- Others – Any other area of problem that is not in the list

Appendix E: Adult Experts' Individual Problem Reports from HE pilot

IDENTITY "A" GENDER "MALE"				
PROBLEMS FOUND	HEURISTIC VIOLATED	TASK	HOW WAS IT FOUND	Severity Rating (0 - 4)
No instructions	4, 9 and 10	Actual game play	Following users task	4
Screen did not rotate	2	Navigating within the game UI	"	2
Just kept going, didn't really know what was going on	11 and 6	Actual game play	Following users task	3
Level 2 seemed to be exactly the same	6 and 11	Actual game play	Following users task	3
Feedback is not very useful	26	Actual game play	Following users task	2

IDENTITY "B" GENDER "MALE"				
Hard to tell how to start game	21	Actual game play	Following users task	2
No instruction given on how to play game	4 and 9	Actual game play	Following users task	4
Player has to hold down fingers to interact	4, 1, 26, 21	Actual game play	Following users task	3
Fingus disapear after first time out	15 and 21			3
No feedback on why player is right or wrong with answer i.e to answer given	21 and 26			4

IDENTITY "C" GENDER "MALE"				
No idea what to do	4	Actual game play	Following users task	4
No positive feedback or guidance on what to do first time	9	"	"	2
Difficulty changes disproportunately (gets easier then harder then easier)	8 and 15	"	"	3
Fruit positioning makes touching all fruits at once difficult	25	"	"	2
No help found	29	Navigating within the game UI	Systematically searching for problems	3
Start page makes no sense to me	26	"	"	2

Appendix F: Observers' Categorised Data for HE Child Evaluators

OBSERVERS' PROBLEM LIST FOR SINGLE EVALUATION PHASE

S/NO	Globally emerged Problem after merging	Predefined Themes (categories)	Problem identified within predefined themes	Codes for problems within predefined themes
1.	Form Issues (16)	DCF	Not sure where to write problem on form (1)	H1FDa1
			Not putting Heuristic number on form (3)	H1FDdc3
			Generic problem filling in form (4)	H1FDabcd4
			Not sure of identity purpose on form (1)	H1FDb1
		HS	Having problems attaching heuristic to problem on the form (4)	H1FHSabcd4
		SR	Not recording severity ratings on forms (3)	H1FSRabd3
2.	Children (2)	Other	Influenced by Peers (1)	H1ChOa1
		EI	Children not engaged (1)	H1ChEIa1
3.	Game play (12)	Other	Unsure how to play game (4)	H1GOab4
		Task	Unsure how to play game (8)	H1GTabcd8
4.	Not recording problem	Other	Evaluator encountered problem	H1NOacd3

	(7)		but did not write down (3)	
		Task	Evaluator encountered problem but did not write down (4)	H1NTc4
5.	Understanding (8)	HS	Reading List of Heuristics not playing (1)	H1UHa1
			Problem understanding heuristics (2)	H1UHbd2
		Task	Not knowing what to do (1)	H1UTb1
		EI	Unclear instructions (2)	H1UEIbd2
		SR	Did not understand severity ratings (2)	H1USRab2
6.	Missing information (1)	EI	Missing informing evaluators some instruction (1)	H1MEIa1

OBSERVERS' PROBLEM LIST FOR GROUP HEURISTIC EVALUATION PHASE TWO

S/NO	Globally emerged Problem after merging	Predefined Themes (categories)	Problem identified within predefined themes	Actual individual problem in problem identified within predefined themes
1	Communication (5)	PCatI	Problem with evaluators agreeing same issue/problem (3)	H2CPCbcd3

		SR	Lack of discussion from evaluators on severity ratings (2)	H2CSRcd2
2	Bias (3)	Other	Facilitator bias children merging problem (3)	H2BObc2
3	Understanding (3)	SR	Understanding Severity (3)	H2USRbd3

Appendix 2 Chapter 6 Documents

13.2 Appendix 2 Chapter 6 Documents

Appendix 2A: FG Consent Form

FOCUS GROUP CONSENT/ETHICAL FORM FOR SCHOOL TEACHERS AS PRESENTED BY: OBELEMA AKOBO WODIKE

A PhD student and member of the ChiCI group at the University of Central Lancashire who will be the moderator for the focus group

I am here today to carry out focus group from which data collected will be used to aid my research. The focus group will be lasting for 30mins and I will be moderating. Discussions will be audio recorded, transcription of the recording will be done in whole within the university and transcript will be stored on my computer and used strictly for my research. Voices, names of individual and name of school will be anonymous and names mentioned during the discussion will be changed when used in presentations or in published document. The audio recording will be deleted after being cross checked with the transcript.

Though your contribution is very vital for my research you are not obliged to participate so feel free to leave at any time without a reason. Ask any questions if you are confused by any statement or sentence.

If you have read, understood the above written and have no more questions, could you please check the boxes below, sign and date this form to give your consent.

I am happy to participate in the focus group ☐

I understand I am not obliged to participate and could leave at any time during the discussion without stating any reason ☐

I understand discussions during the focus group will be audio recorded ☐

I know the audio recording will be transcribed in whole and used for research purposes ☐

I agree that names and voices will be anonymous and names mentioned will be changed ☐

I understand all these processes, I know what it involves and I am happy to give my consent ☐

____________________________ ______________________

Signature Date

Appendix 2B: DESCRIPTION OF THE METHOD (Heuristic Evaluation) FOR FOCUS GROUP

HEURISTICS

Heuristic Evaluation is an expert evaluation method in which usability experts carry out inspection of technology with the use of rule of thumbs or guidelines in this case referred to as heuristics. Heuristics was originally created to evaluate the usability of user interface but over time several research work have been carried out that led to the creation of game and playability heuristics. Till now the playability and game heuristics are used to evaluate both adult and children games but all these heuristic evaluations are done by adults; this is the reason for my research (to carryout heuristic evaluation for children's technology with children).

HOW HEURISTIC WORKS

The heuristic evaluation process is done in two phases, in the first phase the evaluators individually evaluate the technology and find problems while in the second phase, evaluators come together and merge all their individual problems into a single categorized list with severity ratings which would have been agreed by the group and frequency (number of evaluators that individually found and recorded the problem). The process is broken down further:

Phase 1

Evaluators are given an application (it can be an interface or a game, for this study it is a game), a problem sheet, a sheet with the set of heuristic and a pen. They are allowed and given time to

1. Play the game.
2. Find usability problem while game is being played.
3. Record usability problem (that is, whenever a problem is found, they write it in the problem sheet)

Phase 2

1. Evaluators group themselves and produce a categorized list of merged usability problem with severity rating and frequency.

For my heuristic evaluation study, two of my colleagues are present to observe the whole process during the heuristic evaluation.

Appendix 3 Chapter 7 Documents

13.3 Appendix 3 Chapter 7 Documents

3A: **<u>Gathering Game Criteria Form</u>**

Group __________ Children's Year___________ Children's Age Range____________

S/No	Game Criteria Identified	Frequency of Identification

Appendix 3B: <u>**DESIGNING A SEVERITY SCALE**</u>

Group________________ **Identity_________________**

INSTRUCTION: Draw Pictures to represent a 'not a problem at all' to a 'very bad problem'. You could draw as many or few pictures in between the points as you want.

Appendix 3C: Younger Children's Data - **Merged Game Criteria**

S/No	CRITERIA	Groups
1	Remind me of when I was small	1
2	Hide and catch people	1
3	I could dance	1
4	It's a challenge, moves are hard, it's not easy, it gets harder	1, 2, 4, 5, 6
5	Copy the tactics	1
6	Game can be on another device (Wii)	1
7	Beat People	1
8	You can learn e.g. Learn to bake and Learn how to work your brain to make your brain healthier	1, 4, 6
9	It moves when you want it to	1
10	Action and shooting	2
11	I could use my imagination	2
12	A game with octopus is really interesting	2
13	That teaches me to be what I want to be	2, 3, 4
14	Fun	3, 4, 6
15	It has to make people like it and children will want to play with it	3
16	It has loads of stuff	3
17	You can do some stuff, you can do what you want to do, create, make and design e.g. you can change the wall paper	1, 3, 4, 5
18	You get medals and hit highest scores	3, 4
19	If you get it wrong, it makes a sound so you know	4
20	You can choose character or player that you like where some are boys and some are girls	5
21	You can see your score	5
22	Activity you like the most and you can guess	6
23	You can play anything	6

3D: Younger Children's Data - **Creating Themes for Criteria**

S/No	CRITERIA (Themes)	Groups
1	Remind me of when I was small **(memorable)**	1
2	Hide and catch people **(fascinating)**	1
3	I could dance **(fascinating)**	1
4	It's a challenge, moves are hard, it's not easy, it gets harder **(challenging)**	1, 2, 4, 5, 6
5	Copy the tactics **(learning)**	1
6	Game can be on another device (Wii) **(flexible)**	1
7	Beat People **(Progress)**	1
8	You can learn e.g. Learn how to work your brain to make your brain healthier, you can learn to bake **(learning)**	1, 4, 6
9	It moves when you want it to **(progress)**	1
10	Action and shooting **(fascinating)**	2
11	I could use my imagination **(learning)**	2
12	A game with octopus is really interesting **(fascinating)**	2
13	That teaches me to be what I want to be **(learning)**	2, 3, 4
14	It should be Fun/funny **(fascinating)**	3, 4, 6
15	It has to make people like it and children will want to play with it **(fascinating)**	3
16	It has loads of stuff **(flexible)**	3
17	You can do some stuff, you can do what you want to do, create, make and design e.g. you can change the wall paper **(flexible)**	1, 3, 4, 5
18	You get medals and hit highest scores **(Progress)**	3, 4
19	If you get it wrong, it makes a sound so you know **(Notification)**	4
20	You can choose character or player that you like where some are boys and some are girls **(flexible)**	5
21	You can see your score **(Notification)**	5
22	Activity you like the most and you can guess **(flexible)**	6
23	You can play anything **(flexible)**	6

Appendix 3E: Classification of Game Criteria into Themes and Matching Existing Heuristics/Guidelines

S/No	Themes	Younger Children's Game Criteria	Matching Existing Game Heuristics/Guidelines
1	Memorable	Remind me of when I was small	
2	Fascinating	Hide and catch people	
		I could dance	
		Action and shooting	
		A game with octopus is really interesting	The child is interested in the eLearning program characters because … (2) they are interesting to him, …(Alsumait & Al-Osaimi, 2009)
		It should be fun and funny	
		It has to make people like it and children will want to play with it	The game is enjoyable to replay (Desurvire et al., 2004)
3	Challenging	It's a challenge, moves are hard, it's not easy or it gets harder	A good game should be easy to learn and hard to master (Alsumait & Al-Osaimi, 2009; Desurvire et al., 2004; Federoff, 2002)
4	Learning	Copy the tactics	
		You can learn e.g. Learn how to work your brain to make your brain healthier, you can learn to bake	One reward of playing should be the acquisition of skill (Federoff, 2002) The program supports the child's cognitive curiosity through surprises, paradoxes, humour, and dealing with topics that already interest the child. (Alsumait & Al-Osaimi, 2009)

		I could use my imagination	The eLearning program allows the child to use his imagination… (Alsumait & Al-Osaimi, 2009)
		That teaches me to be what I want to be	
5	Flexible	Game can be on another device (Wii)	…program may be used on a variety of equipment and platforms such as laptops, PDA. (Alsumait & Al-Osaimi, 2009)
		It has loads of stuff	…program is enjoyable and interesting. It uses e-stories, games, simulations, role playing, and activities to gain the attention and maintain the motivation of learners (Alsumait & Al-Osaimi, 2009) Include a lot of interactive props for the player to interact with (Federoff, 2002)
		You can do some stuff, you can do what you want to do, create, make and design e.g. you can change the wall paper	Allow players to build content. (Federoff, 2002)
		You can choose character or player that you like where some are boys and some are girls	
		Activity you like the most and you can guess	
		You can play anything	
6	Progress	Beat People	

		It moves when you want it to	…The program, on the other hand, needs to respond immediately to the child (Alsumait & Al-Osaimi, 2009)
		You get medals and hit highest scores	The game should give rewards (Federoff, 2002)
7	Notification	If you get it wrong, it makes a sound so you know	Use sound to provide meaningful feedback (Federoff, 2002)
		You can see your score	A player should always be able to identify their score/status in the game. (Desurvire et al., 2004; Federoff, 2002)

3F: USING TRAFFIC LIGHT SYSTEM TO JUDGE PROBLEM SEVERITY (INPUT FROM YEAR 2 CHILDREN)

S/No	Games	Problem	Traffic Light Colours				Groups (No. in group)
			Green	**Yellow**	**Red**	**Undecided**	
1	JamMo	If you do something wrong it does not tell you			4		1 (4)
					3	1	2 (4)
					4		3 (4)
					4		4 (4)
					2		5 (2)
					1		6 (1)
			-	**-**	**18**	**1**	**Total**
		Dragging items is a bit of a problem		4			1
				3		1	2
				4			3
				4			4
				2			5
					1		6
			-	**17**	**1**	**1**	**Total**
2	LeesCircus	Volume on the laptop was too low		1	3		1
			1		2	1	2
				2	2		3
					4		4
				1	1		5 (3)
					1		6
			1	**4**	**13**	**1**	**Total**
		It is difficult to drag items with the mouse		3	1		1
				3	1		2
				2	2		3
				1	3		4
					2		5
					1		6

			-	**9**	**10**	**-**	**Total**
		Could not click on pictures			4		1
			2		1	1	2
					4		3
					4		4
					2		5
					1		6
			2	**-**	**16**	**1**	**Total**
3	Milo	Instructions on how to play each sub game were too long			4		1
			2		2		2
			2	2			3
					4		4
					2		5
				1			6
			4	**3**	**12**	**-**	**Total**
		Explanation was not enough		4			1
				4			2
					4		3
					4		4
				2			5
					1		6
			-	**10**	**9**	**-**	**Total**
		Did not understand how to catch the flies		1	3		1
			2		2		2
			2	2			3
					4		4
					2		5
					1		6
			4	**3**	**12**	**-**	**Total**
		Did not know how to give the fly to the toad	3	1			1
			1			3	2
				4			3

					4		4
					2		5
				1			6
			4	**6**	**6**	**3**	**Total**
		It is impossible to follow any tactic when trying to click on the crabs	1	1	2		1
			1	2	1		2
					4		3
					4		4
					2		5
					1		6
			2	**3**	**15**	**-**	**Total**
4	Anti-heuristics space-invader	The process of restarting the game is long			4		1
			1		3		2
				2	2		3
					4		4
				1	1		5
					1		6
			1	**3**	**15**	**-**	**Total**
		Screen turns black with inappropriate feedback	1	2	1		1
			1	3			2
					4		3
					4		4
					2		5
					1		6
			2	**5**	**13**	**-**	**Total**

<u>COLOUR CODE AND KEY</u>

GREEN – Not a problem

YELLOW – A bit of a problem

RED – A bad problem

UNDECIDED – used more than one colour

3G: Description of younger children's severity scale drawing

Groups	Identity	Step 1	Step 2	Step 3	Step 4	Step 5
1	A	A sun flower (flower - random)				
	B	A smiley face girl (Faces)	A sad face girl			
	C	A smiley face girl (Faces)	A sad face girl	An angry face girl		
	D	A sad face being (faces)	A Straight face being	A smiley face being		
2	A	A house with two windows (House - random)	Another house with 4 windows and smoke coming out of the chimney			
	C	Two persons' conversation (random)				
	D	A house on the grass with smoke coming out of the chimney (House - random)				

3	A	Someone robbing (Persons - random)	The person jailed	The person is freed and asked not to steal again		
	B	A person behind bars (persons)	Someone driving freely in a car			
	C	A person stuck half way to climb the top of the house (persons)	A ladder was provided with which the person climbed			
	D	A sad face lady (faces)	A straight face lady	A smiley face	A grinning smiley lady	A group of smiling ladies
4	A	A stick person tumbling over (Not a problem) (A stick person)	A stick person half way stable (A bit of problem)	A stick person at an angle position (A bit more of a problem)	A stick person standing (A big problem)	
	B	A person in a house and looking out of the window	A person trying to go out through	A person almost falling over from	A person falling over with his head (very bad)	

		(Happy) (persons)	the window (Worried)	the window (not good)		
	C	A smiley faced person on a roller skate (Not a problem) (Faces)	A sad face person on a roller skate with one troubled wheel (A bit of a problem)	A sad face person on a roller skate with two troubled wheel (A big bad problem)		
	D	A smiley face girl (happy problem) (faces)	A sad face girl (bad problem)			
5	A	A smiley butterfly (happy) (Faces)	A sad butterfly (sad)			
	C	A sad face butterfly (faces)	A smiley face butterfly			
6	A	Random picture of a pink cat (animal - random)	A blue dog	A dog wearing a purple ribbon		

Faces 8, flower 1, House 2, animal 1, random 1, persons 5

3H: Children's Severity Text Label

identity	**Text**
4A	Not a problem, A bit of problem, A bit more of a problem, A big problem
4B	Happy, Worried, Not good, Very bad
4C	Not a problem, A bit of a problem, A big bad problem
4D	Happy problem, Sad problem
5A	Happy, sad

3I: Older Children's severity drawing score sheet

Table 7.6: Scores of the pictures based on the codes

S/No	PICTURE IDENTITY	LEVEL OF PICTURE	TEXT	FIT TO RATE PROBLEM	CONSISTENT OR REPRESENTATION OF 'UP' CONCEPT	TOTAL SCORE
1	1A	1	1	1	1	4
2	1B	1	1	1	1	4
3	1C	1	1	1	1	4
4	1D	1	1	1	1	4
5	2A	1	1	1	1	4
6	2B	1	1	1	1	4
7	2C	1	1	1	1	4
8	2D	1	1	0	0	2
9	3A	1	0	0	0	1
10	3B	1	0	0	0	1
11	3C	1	1	1	1	4
12	3D	1	1	1	1	4
13	4A	1	1	0	1	3
14	4B	1	1	0	1	3
15	4C	1	1	1	1	4
16	4D	1	1	1	1	4
17	5A	1	1	1	1	4
18	5B	1	1	1	0	3
19	5C	1	1	1	1	4
20	5D	1	1	1	1	4
21	6A	1	0	0	1	2
22	6B	1	1	1	1	4
23	6C	1	1	0	0	2
24	6D	1	1	1	1	4
TOTAL		24	21	17	19	81

To calculate the mean score = $\frac{Total\ number\ of\ coded\ picture\ scores}{Total\ number\ of\ Child\ Participants} = \frac{\sum X}{N} = \frac{81}{24} = 3.375 = 3.4$

3J: Older Children's Game Criteria in Groups

S/No	GAME CRITERIA OF YEARS 5 AND 6 CHILDREN BY THEIR GROUPS					
	GROUP 1 (FREQUENCY)	**GROUP 2 (FREQUENCY)**	**GROUP 3 (FREQUENCY)**	**GROUP 4 (FREQUENCY)**	**GROUP 5 (FREQUENCY)**	**GROUP 6 (FREQUENCY)**
1	It should be fun learning not just learning – for a learning game (1) F1 & L/M/C1	It should be adventurous (3) CH2	Some questions in it like Maths puzzles and stuff (1) CH3	Fun like Fun Maths (4) F4 & L/M/C4	It should be flexible to be on every game console (3) F/A/I5	It has to be an action game (2) CH6
2	Game should have different sections like a quiz (buttons at the bottom for the different sections) You should be able to click any section by answering the quiz (3) CH1 & F/A/I1	It should have a good storyline and a plot to the storyline (4) S2	Eye Catching – good graphics (1) G/A3	Cute animals should be used (4) G/A4	It should be challenging (2) CH5	It has to be hard (1) CH6
3	Challenging –	It should have instructions (2) H/IN2	It should be something interactive (1) F/A/I3	It should be for all child age group (4) AA4	There should be different levels of difficulty	It has to be adventurous (2) CH6

	Starts off easy then gets harder and harder (3) CH1				(2) CH5	
4	Have hints and help (3) H/IN1	It should be flexible that you can make your own characters (3) F/A/I2	You can make your characters better like upgrading it (1) F/I/A3	It should use cartoons but should relate to real people (3) G/A4	There should be continuity and connection between games (2) L/M/C5	A bit of fighting (2) CH6 & F6
5	When you get it right, you should get reward or coins, you should be able to use the coins or reward to go to the next phase and buy stuff for your character (3) R1	It could be played on different platforms (3) F/A/I2	It should have different levels (3) CH3	It should have questions about what you have done about a level (2) AS4	At the end of every 10 levels you should fight a boss (3) CH5	Fun but you should learn (3) L/M/C6
6	It should have different levels (2) CH1	It should have different levels (3) CH2	You should get rewards for completing a level (1) R3	You should remember what you have learnt in the game after playing- You should be able to	There should be health package for power upgrade (1) R5	It should have good graphics - good pictures, people (4) G/A6

				draw what you saw (4) L/M/C4		
7	It should be adventurous (2) F1/CH1	You should have rewards (coins) for doing things (4) R2	It should be challenging (2) CH3	It should be easier for younger ones (3) AA4	It should have more characters to unlock (3) F5/ F/A/I5	Content of the game should be related to real world e.g. people (4) MtRW6
8	For little ones, It should be something that can help them learn the game (1) L/M/C1	You should be able to use your rewards to buy stuff to continue the game (4) R2	It should have different world (2) CH3	There should be different levels and different levels for different age group (4) CH4/AA4	It should be a multiplayer game (2) F5	You get a reward for every correct puzzle you solve in a puzzle game (3) R6
9	It should be something interesting for them (1) AA1	You should be able to test your skills and you can get better (3) AS3	For the little ones, they should be able to learn something for when they are a bit older in an educational game (1) L/M/C3	There should be an explanation (4) H/IN4	There should be animations (2) G/A5	It should be a battle game (1) F6
10	It should be colourful (1) G/A1	It should be challenging (4) F2	The game should have a storyline - The last level should be able to	There should be a speaker to read out	It should be 3D (3) G/A5	It should be a puzzle game (2) F6/CH6

			end the story where the player battles someone to end (1) S3	stuff for people that cannot read (4) F/A/I4		
11	Let the game be able to read out instructions and hints for people that cannot read (2) F/A/I1	It should be educational such that you could learn from the game but should be fun (4) L/M/C2	You should be able to have a house and decorate it (1) F3	There should be an how to play i.e. An instruction button (4) H/I4	There should be 3 world (1) CH5	You can build stuff and make friends (3) F/A/I6
12	Let each section or sub game be age appropriate, so each player can choose the section that is their age grade (2) AA1	You should be able to remember what you learnt and what is in the game (4) L/M/C2	You should be able to compare what is in the game to what is in real world (4) MtRW3		It should have a name that will distinct it from other games (2) MtRW5	It should allow you to easily join the game (3) F/A/I6
13	It should be fun and addictive (1) F1	For young children, it should be what they can understand (4) L/M/C2	You should be able to find things that will give you power up (3) AS3		You should be able to save someone (2) CH5/F5	

14	It should not be only one level and boring (1) CH1	It should have button to press that will read out to them in case they don't know how to read (2) F/A/I2	You should be able to collect coins or get stars when you are doing good (4) R3		It should be for boys and girls not boys alone (3) F/A/I5	
15		It should not be complicated and should make them laugh (3) F/A/I2	You should have information to help you with the questions (4) H/I3		You should be able to overcome obstacles (3) AS/P5	
16		A good name for the game (something catchy) (4) G/A/D2			It should not be copyright. In other words, it should not be a copied game (3) G/A/D5	
17		It has to be colourful and has good graphics (4) G/A/D2				

3K: **Merging Criteria, Identifying Exact Frequencies and Group**

S/No	CATEGORISED CRITERIA	GROUPS
1	**REWARD** When you do things or get it right or for correct puzzles or for completing a level, you should get reward or coins or stars, you should be able to use the coins or reward to go to the next phase and buy stuff for your character to continue game	**1, 2, 3, 6**
	There should be health package for power upgrade	**3, 5**
2	**CHALLENGE** It should be adventurous	**1, 2, 6**
	There should be some questions in it like Maths puzzles and stuff	**3**
	It has to be an action game, a bit of fighting / should be a battle game or at the end of every 10 levels you should fight a boss	**5, 6**
	It should be challenging, different levels of difficulty like starts of easy then gets harder - it has to be hard and should have different world	**1, 2, 3, 5, 6**
	It should be a puzzle game	**6**
3	**AGE APPROPRIATE** It should be for all child age group - let each section or sub game be age appropriate, so each player can choose the section that is their age grade	**1, 4**
	For little ones, It should be something they can understand and that can help them learn the game	**1, 2**
	There should be different levels for different age group (4) AA4	**4**
	It should not be complicated (i.e. easier) but interesting for them and makes them laugh	**1, 2, 4**

4	**LEARNING, MEMORABILITY & CONSISTENCY** It should be fun learning not just learning – for a learning game e.g. fun like fun Maths	**1, 2, 4, 6**
	There should be continuity and connection between games	**5**
	You should remember what you have learnt in the game after playing- You should be able to draw what you saw	**2, 4**
	For the little ones, they should be able to learn something for when they are a bit older in an educational game	**3**
5	**STORY** It should have a good storyline and a plot to the storyline. The last level should be able to end the story where the player battles someone to end	**2, 3**
6	**MATCH TO REAL WORLD** Content of the game should be related to real world e.g. people	**3, 6**
7	**ASSESSMENT & PROGRESS** You should be able to overcome obstacles	**5**
	It should have questions about what you have done about a level	**4**
	You should be able to test your skills and you can get better	**3**
8	**GRAPHICS, AESTHETICS & DESIGN** It has to be colourful, eye catching and has good graphics and pictures	**1, 2, 3, 6**
	It should not be copyright. In other words, it should not be a copied game and should have a name (something catchy) that will distinct it from other games	**2, 5**

	It should use cartoons but should relate to real people. Cute animals should also be used	**4, 5**
9	**FLEXIBILITY, ACCESSIBILITY & INTERACTIVITY** It should be flexible to be on every game console or at least on different platforms (3) F/A/I5	**3, 5**
	Game should have different sections like a quiz (buttons at the bottom for the different sections) You should be able to click any section by answering the quiz (3) CH1 & F/A/I1	**1**
	It should be something interactive (1) F/A/I3	**3**
	It should be flexible that you can unlock more characters or can make your own characters and even make your characters better like upgrading it (3) F/A/I2	**2, 3, 5**
	Let the game be able to read out instructions, hints and stuff for people that cannot read (2) F/A/I1	**1, 2, 4**
	It should allow you to easily join the game (3) F/A/I6	**6**
	The game should be for boys and girls not boys alone (3) F/A/I5	**5**
10	**HELP, INSTRUCTION** You should have information to help you with the questions (4) H/I3	**1, 3**
	There should be an explanation on how to play i.e. An instruction button (4) H/I4	**2, 4**
11	**GAME PLAY** You should be able to have a house and decorate it	**3**
	You can build stuff and make friends	**6**
	It should be fun and addictive	**1**
	It should be a multiplayer game and you should be able to save someone	**5**

3L: Children's Criteria Matched to Existing Heuristics / Guidelines

S/No	CATEGORISED CRITERIA	ASSOCIATED EXISTING HEURISTIC / GUIDELINE	TYPE OF HEURISTIC / GUIDELINE	REFERENCE
1	**REWARD** When you do things or get it right or for correct puzzles or for completing a level, you should get reward or coins or stars, you should be able to use the coins or reward to go to the next phase and buy stuff for your character to continue game	**GP3 / GAME PLAY** Players are rewarded **GP3 / EUH 3** Rewards are meaningful **GAME PLAY** The game should give rewards that immerse the player more deeply in the game by increasing their capabilities and expanding their abilities to customize	ELearning; Game Play; and Playability Heuristics	(Alsumait & Al-Osaimi, 2009; Federoff, 2002; Korhonen & Koivisto, 2006); (Desurvire et al., 2004)
	You should be able to have a house and decorate it	**GAME PLAY** The game should give rewards that immerse the player more deeply in the game by increasing their capabilities (power-up), and expanding their ability to customize.	Playability Heuristics	(Desurvire et al., 2004)
	There should be health package for power upgrade	**GAME PLAY** … by increasing their capabilities (power-up),	Playability Heuristics	(Desurvire et al., 2004)
2	**CHALLENGE** It should be adventurous	**CLEAR NARRATIVE**	CCG design guidelines	(Villalta et al., 2011)

		The narrative should be composed of quests and challenges that define collaborative activities in a sequential and precise pattern		
	There should be some questions in it like Maths puzzles and stuff			
	It has to be an action game, a bit of fighting / should be a battle game or at the end of every 10 levels you should fight a boss			
	It should be challenging, different levels of difficulty like starts of easy then gets harder - it has to be hard and should have different world	**CHALLENGE** The game should provide different challenge levels for different players **GAMEPLAY –** There should be variable difficult levels **CUH 4 CHALLENGING THE CHILD** The e-learning program is easy to learn, but hard to master. The application is paced to apply pressure but not frustrate the child. The difficulty level varies so that the child has greater challenges as he develops mastery. **GAME PLAY** Pace the game to apply pressure but not frustrate the player. Vary the difficulty level so that the player has greater	Game Heuristics; Child Usability Heuristics; and Playability and Usability Heuristics	(Alsumait & Al-Osaimi, 2009; Desurvire et al., 2004; Febretti & Garzotto, 2009; Federoff, 2002; Pinelle et al., 2008)

		challenge as they develop mastery. Easy to learn, hard to master.		
	It should be a puzzle game			
3	**AGE APPROPRIATE** It should be for all child age group - let each section or sub game or level be age appropriate, so each player can choose the section that is their age grade	**GLOBAL QUESTION 1a** Is the challenge right for the target group?	SEEM Question	(Baauw et al., 2005)
	For little ones, It should be something they can understand and that can help them learn the game	**NUH_1** The child understands all terminology used in the program **EUH_1** The vocabulary and terminology used are appropriate for the learners. **GAMEPLAY, General -** Players should understand and be able to identify goals **GU5** The player understands the terminology **INTERACTIVITY & GUIDANCE** The user's interaction with the game must be simple and intuitive and not add unnecessary complexity to the game	ELearning Usability; and Game Usability Heuristics; and CCG design guidelines	(Alsumait & Al-Osaimi, 2009; Korhonen & Koivisto, 2006; Schaffer, 2007; Villalta et al., 2011)
	It should not be complicated (i.e. easier) but interesting for them and makes them laugh	**CUH 1 MULTIMEDIA REPRESENTATION** The e-learning program includes surprises, humour and	ELearning Heuristics	(Alsumait & Al-Osaimi, 2009)

		interesting representations for the child and avoids unnecessary multimedia representations as they can confuse children that have just started to work with the program **CUH_6** The program supports the child's cognitive curiosity through surprises, paradoxes, humour, and dealing with topics that already interest the child.		
4	**LEARNING, MEMORABILITY & CONSISTENCY** It should be fun learning not just learning – for a learning game e.g. fun like fun Maths	**EUH 3 MOTIVATION TO LEARN** The e-learning program is enjoyable and interesting. It uses e-stories, games, simulations, role playing, and activities to gain the attention and maintain the motivation of learners. The application provides the learner with frequent and varied learning activities that increase learning success	ELearning heuristics	(Alsumait & Al-Osaimi, 2009)
	There should be continuity and connection between games	**GP12 / MECHANICS** Game is consistent / Game should react in a consistent, challenging and exciting way to the player's actions	Playability Heuristics;	(Alsumait & Al-Osaimi, 2009; Desurvire et al., 2004; Korhonen

		CUH Multimedia representations assist the e-learning process. Application should react in a consistent, challenging, and exciting way to the child's actions (e.g., appropriate video clips with the music **GAME PLAY** The game does not stagnate **SG3 CONTINUITY** Provide an asynchronous persistence game world and mechanics that allow the player to feel progress.	Child Usability Heuristics; Game Play Heuristics; Social Game Heuristics	& Koivisto, 2006; Paavilainen, 2010)
	You should remember what you have learnt in the game after playing- You should be able to draw what you saw	**GAME INTERFACE** Follow the trends set by the gaming community to shorten the learning curve	Game and Pervasive Game Heuristics	(Desurvire et al., 2004; Federoff, 2002; Köffel & Haller, 2008; Pinelle et al., 2008; Röcker & Haar, 2006; Schaffer, 2007)
	For the little ones, they should be able to learn something for when they are a bit older in an educational game	**ACTION GUIDE** The game must have a systematic design that includes the educational and ludic aspects,	CCG design guidelines	(Villalta et al., 2011)

		through a script that specifies action sequences, possibilities for action, and events that might take place both in the virtual world and in the real world.		
5	**STORY** It should have a good storyline and a plot to the storyline. The last level should be able to end the story where the player battles someone to end	**GAMEPLAY/GP7** Create a great storyline / Game story supports game play and is meaningful **Global Question 2b** Does the flow of the game meet the expectations? Is the story line logical **CLEAR NARRATIVE** The game must have a base story that allows the participants' immersion	Gameplay and Playability Heuristics, SEEM Question, and CCG design guidelines	(Baauw et al., 2005; Desurvire et al., 2004; Federoff, 2002; Korhonen & Koivisto, 2006; Villalta et al., 2011)
6	**MATCH TO REAL WORLD** Content of the game should be related to real world e.g. people	**CUH_5** The child is interested in the e-learning program characters because (1) they are like the child (2) they are interesting to him (3) they are drawn from the child's own culture. **Match between system and the real world** The system should speak the users' language, with words, phrases and concepts familiar to	ELearning Usability; Usability; and Social Game Heuristics	(Alsumait & Al-Osaimi, 2009; Molich & Nielsen, 1990; Jakob Nielsen, 1993b; Paavilainen, 2010)

		the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order. **SG1** Use common and familiar themes from popular culture which can be understood easily.		
7	**ASSESSMENT & PROGRESS** You should be able to overcome obstacles	Players should feel in control, so they need the time and information to respond to threats and opportunities. That is, players should see enemies, obstacles, and power-ups coming.	Usability Heuristics	(Schaffer, 2007)
	It should have questions about what you have done about a level such that you can test your skills and can get better	**EUH 2 ASSESSMENT** The e-learning program includes self-assessments that advance child achievement and provides sufficient feedback (audio, video) to the child to provide corrective directions. **GP2** The player sees the progress in the game and can compare the results **INTERACTIVITY & GUIDANCE** The game must offer guidance, both for individual and collective action, through	ELearning Usability and Gameplay Heuristics; and CCG design guidelines	(Alsumait & Al-Osaimi, 2009; Korhonen & Koivisto, 2006; Villalta et al., 2011)

		precise, timely and constant information regarding success and failure in performance.		
8	**GRAPHICS, AESTHETICS & DESIGN** It has to be colourful, eye catching and has good graphics and pictures	**CUH 2 DESIGN ATTRACTIVE SCREEN LAYOUT** The screen layout is efficient and visually pleasing. It should appear simple, i.e., uncluttered, readable, and memorable − The font choice, colours and sizes are consistent with good child screen design.	ELearning Usability Heuristics	(Alsumait & Al-Osaimi, 2009)
	It should not be copyright. In other words, it should not be a copied game and should have a name that will distinct it from other games	**EUH_3** The ELearning program incorporates novel characteristics	ELearning Usability Heuristics	(Alsumait & Al-Osaimi, 2009)
	It should use cartoons but should relate to real people. Cute animals should also be used	**GAME STORY** Player is interested in the characters because (1) They are like me (2) they are interesting to me (3) The characters develop as actions occur	Playability Heuristics	(Desurvire et al., 2004)
9	**FLEXIBILITY, ACCESSIBILITY & INTERACTIVITY** It should be flexible to be on every game console or at least on different platforms	**EUH_5** The e-learning program may be used on a variety of equipment and platforms such as laptops, PDA	ELearning Usability Heuristics	(Alsumait & Al-Osaimi, 2009)

	Game should have different sections like a quiz (buttons at the bottom for the different sections) You should be able to click any section by answering the quiz			
	It should be something interactive	**GAME PLAY** Include a lot of interactive props for the player to interact with	Game Play Heuristics	(Federoff, 2002)
	It should be flexible that you can unlock more characters or can make your own characters and even make your characters better like upgrading it	**GAME PLAY** The game should give rewards that immerse the player more deeply in the game by increasing their capabilities and expanding their abilities to customize	Playability Heuristics	(Desurvire et al., 2004)
	Let the game be able to read out instructions, hints and stuff for people that cannot read	**USABILITY** Players should be given context sensitive help while playing so that they do not get stuck or have to rely on manual **M6** Intuitive controls mapped in a natural way ('speech input and output to achieve intuitive and easy interaction)	Playability and Pervasive Game Heuristics	(Desurvire et al., 2004; Röcker & Haar, 2006)
	It should allow you to easily join the game	**MP4.** The game helps the player to find others players and game instances – If the game design includes game instances, the player should be able to easily find and join them.	Multi-player Heuristics Social Game Heuristics, and Network	(Korhonen & Koivisto, 2007; Paavilainen, 2010; Pinelle, et al., 2009)

		SG1 - Provide easy and quick access to the game as the threshold for play should be as minimal as possible **SIMPLE SESSION MANAGEMENT** Provide session management that allows players to start new games, and that allows them to find and join appropriate games	multi-player game heuristics	
	The game should be for boys and girls not boys alone			
10	**HELP, INSTRUCTION** You should have information to help you with the questions	**NUH_10** The child should be given help while using the program. Help should be easy to search. Any help provided is focused on the child's task, and lists simple concrete steps to be carried out. **USABILITY** Players should be given context sensitive help while playing so that they do not get stuck or have to rely on manual **GAME PLAY** The game should give hints, but not too many **GU12** The game contains help **INTERACTIVITY & GUIDANCE**	ELearning Usability; Playability; Game; Game Usability Heuristics; and CCG design guidelines	(Alsumait & Al-Osaimi, 2009; Desurvire et al., 2004; Federoff, 2002; Korhonen & Koivisto, 2006; Villalta et al., 2011)

		The game must offer guidance, both for individual and collective action, through precise, timely and constant information regarding success and failure in performance.		
	There should be an explanation on how to play i.e. An instruction button	**CUH 4 CHALLENGE THE CHILD** The child should have enough information to start to use the program when he turns it on. **NUH_6** Instructions for the use of the program are visible or easily retrievable, so that the child does not have to memorize unnecessary things. **U5,6,8** Upon initially turning the game on, the player has enough information to get started to play; context sensitive help; no manual needed to play (easy, quick and intuitive interaction)	ELearning Usability; and Pervasive Game Heuristics	(Alsumait & Al-Osaimi, 2009; Röcker & Haar, 2006)
11	**GAME PLAY** You can build stuff and make friends	**GAME PLAY** Allow players to build stuff **FLEXIBLE MATCHMAKING** Provide matchmaking features to help people find players with similar interests	Game Heuristics Networked Game Heuristics	(Federoff, 2002; Pinelle et al., 2009)

	It should be fun and addictive	**EUH_3** The e-learning program is enjoyable and interesting. It uses e-stories, games, simulations, role playing, and activities to gain the attention and maintain the motivation of learners. **Game Play** The game is enjoyable to replay/ Make the game replayable	ELearning; Playability; and Game Heuristics	(Alsumait & Al-Osaimi, 2009; Desurvire et al., 2004; Federoff, 2002)
	It should be a multiplayer game and you should be able to save someone	**SUPPORT SOCIAL INTERACTION** Provide support for planned and opportunistic social interactions. Games should also provide features that encourage conversation and cooperation between players.	Networked Game Heuristics	(Pinelle et al., 2009)

3M: Older Children's Severity Drawing Described in Groups

CHILDREN'S PICTURE TYPES IN LEVELS

IDENTITY	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	LEVEL 5	LEVEL 6	LEVEL 7	LEVEL 8	LEVEL 9	LEVEL 10
1A	A big smile with a tooth out on a square face	An average smile on a square face	A small smile on a square face	A smaller smile on a square face	A straight square face	A square face with little sad expression	A square face with more sad expression	A square face with big sad expression	A square face with wriggled lip sad expression	A square face with very big sad expression and a dribbling nose
1B	Four pencils but only one will be circled	Four pencils but only two will be circled	Four Pencils and only three will be circled	Four pencils, All four will be circled						

1C	A face with a side smile	A little squeezed but straight face	A sad face								
1D	Smiley face with a tooth out	A straight face	A squeezed mouth sad face	A big sad face							
2A	A smiling face	A straight face	A sad face								
2B	Face with opened teeth	Face with closed squeezed lips	Angry face with twisted eyes								
2C	A girlish smiley faced with tongue out	A banana person with straight face	A person with standing hair and sad face	A person with glasses and a big sad face							
2D	A globe for earth	An Egyptian Symbol	Another globe for the moon	A star representing the galaxy	A face of a bear						

3A	A face linked to a text box that reads power up!!! Power up!!! Power up!!!	A person and a dog in front of a door	A person trying to press a button at one end of a stairs and another person at the other end of the stairs, a spiral curve and a star crest tin with coins around it, above the stairs with coin on each level of the stairs								
3B	A stick person, a drawing titled pontle home, another drawing titled pontle to resort and some coins										

3C	A face with little smile	A straight face	A sad face								
3D	Two star faces that are smiling	A dog standing and doing nothing	A dog trying to climb on a slide	Two dogs facing themselves head on as in a challenge							
4A	A stick person sitting at a table with a big screen in front and another stick person standing beside the screen	A stick person sitting at a table with another stick person holding something and standing in front									
4B	A stick girl jumping on a trampoline	A stick girl sitting on a couch									
4C	Likert type scale point 1 **OR** A stick person at a lounge on a couch watching TV	Likert type scale point 2 **OR** A stick person jumping on a trampoline	Likert type scale point 3 **OR** A stick person sitting at a table	Likert type scale point 4 **OR** A stick person lying down	Likert type scale point 5 **OR** A stick person sitting idle on a couch						
4D	A girl standing in front of a table filled	A girl sitting at a table for tea time	A stick person								

	with cakes for her birthday	with a plate with food on the table	sitting on the couch							
5A	A picture of an unclear animal	A picture of a pumpkin with hands and legs	A picture of an anglerfish							
5B	An undulated shaped picture	A fish shaped picture with a straight face	A sad face astronaut who has fainted							
5C	A smiling face	A straight face	A sad face							
5D	A smallest sized pumpkin with smallest eyes and teeth	A smaller sized pumpkin with smaller eyes and teeth	A small sized pumpkin with small eyes and teeth	A big sized with big teeth and big eyes and longer arrow						
6A	A person with a smiling face	Another person with a straight face	Another person with a sad face							
6B	A person with hands up and teeth wide open	A person with straight face	A person with face down and putting up a sad face							

6C	A picture of a person trying to be killed by a monster, an alien and some stick persons	Two persons one represent computer and one representing a human being									
6D	A stick person with a smiling face	A stick person with a straight face	A stick person with a sad face because he is being frightened by a monster								

Appendix 4 Chapter 8 Documents

13.4 Appendix 4 Chapter 8 Documents - Children's Data for the first CVBIM EXPERIMENT

Appendix 4A: Detailed Criteria for each group (1 to 3)

GROUP 1	GROUP 2	GROUP 3
GC1A – It should be fun (1)	**GC2A** – It should be fun doing it	**GC3A** – Have a maze for which you will need to solve Maths problems to get out (10)
GC1B – There should be other maths problems like subtraction and multiplication (7)	**GC2B** – There shouldn't be lots to do before you start playing	**GC3B** – Inside the Maze, do mini maths questions in seconds to go pass monsters (9)
GC1C – It should be played in different world (5)	**GC2C** – It should be creative	**GC3C** – You could use brightness to make game option stand out
GC1D – You should get stars when you do it right (2)	**GC2D** – It should have good graphics	**GC3D** – The player should provide the problem and solution
GC1E – You should be able to use your star to upgrade your character or change it (2)	**GC2E** – It should be challenging such that it makes you think not too easy but not too hard that you can't do it	**GC3E** – You should get rewards for doing stuff
GC1F – There should be different bikes and you should be able to change		**GC3F** – You should be able to use your reward to customise your outfit
GC1G – It should start off easy and become harder – challenging (4)		**GC3G** – There should be an alien you have to defeat when you complete all levels
GC1H – You should be able to play with someone (8)		**GC3H** – It should have a good theme song

Appendix 4B: Detailed Criteria for each group (4 to 6)

GROUP 4	GROUP 5	GROUP 6
GC4A – Though it's a maths game, it shouldn't be all about Maths - do something else first for fun (1)	**GC5A –** It got to have a character	**GC6A –** It should be comparing numbers
GC4B – It shouldn't be too easy (4)	**GC5B –** There should be a football shooting at the right sum	**GC6B –** It should be adventurous
GC4C – There should be limited amount of time to answer the questions (9)	**GC5C –** In addition to the football, there should be other sports to choose from	**GC6C –** You should play 10 out of 10 before unlocking the island
GC4D – There should be lots of levels to complete the game (5)	**GC5D –** It should be challenging	**GC6D –** You should be able to collect coins to unlock characters and weapons
GC4E – It should be for boys and girls	**GC5E –** There should be different difficulty levels	**GC6E –** Should be able to play against other characters
GC4F – You should be able to unlock different backgrounds	**GC5F –** It could be like a guessing game	**GC6F –** If you get one wrong answer you get another go on the answer before you go back if you get it wrong again
GC4G – Do not put too many glitches in	**GC5G –** You should get a price whenever you pass a level	**GC6G –** You should be able to do additions and subtractions

GC4H – You should be able to collect coins to upgrade your vehicle (2)	**GC5H –** You should be able to build stuff **GC5I –** There should be bad guys to shoot (6)	**GC6H –** You should be able to fly to other places where there are right answers **GC6I –** There should be a story about a person when you get the story you progress (10) **GC6J –** Whenever you win, you go on to a new level (5) **GC6K –** Add sums to a box fight by answering questions to progress **GC6L –** There should be questions and the answer options should come up for you to answer the questions

Appendix 4C: Similarities of criteria across groups

S/No	Merged Game Criteria between Groups	Groups						Total Group
		1	2	3	4	5	6	
	It should not only be about maths but it should be fun too	*	*		*			3
2	You should get rewards for doing tasks right and you should be able to use your reward for acquiring items during game play	*		*	*	*	*	5
3	There should be different bikes and you should be able to change	*						1
4	There should be other maths problems e.g. Addition, subtraction and multiplication	*					*	2
5	There should be different difficulty level or world to complete the game	*			*	*	*	4
6	It should be challenging such that it makes you think not too easy but not too hard that you can't do it	*	*		*	*		4
7	You should be able to play against other characters or persons e.g. an alien or bad guys	*		*		*	*	4
8	There shouldn't be lots to do before you start playing		*					1
9	It should be creative		*					1
10	There should be limited amount of time to answer the questions			*	*			2
11	You will need to solve Maths problems to get out of a place or to move on			*			*	2
12	You could use brightness to make game option stand out			*				1

13	The player should provide the problem and solution			*				1
14	It should be for boys and girls				*			1
15	You should be able to unlock different backgrounds				*			1
16	Do not put too many glitches in				*			1
17	There should be a football shooting at the right sum					*		1
18	It could be like a guessing game					*		1
19	You should be able to build stuff					*		1
20	It should be comparing numbers						*	1
21	It should be adventurous						*	1
22	If you get one wrong answer you get another go on the answer before you go back if you get it wrong again						*	1
23	You should be able to fly to other places where there are right answers						*	1
24	Add sums to a box fight by answering questions to progress						*	1
25	There should be questions and the answer options should come up for you to answer the questions						*	1
26	It should have good graphics		*					1
27	It should have good theme song			*				1
28	In addition to the football, there should be other sports to choose from					*		1
29	There should be a story about a person when you get the story you progress (12)						*	1

Appendix 4D: Evaluators Individual Problem Report and Merged problems

GROUP ONE					
S/No	Problems identified	individual severity	Merged Problem	Final Severity	Frequency
1	Too Confusing	R	Too Confusing	R	1
2	Difficult	Y	Difficult	Y	1
3	Long Intro	G	Long Intro	G	2
4	Long Intro	G			
5	At the start there's a video, nothing wrong at the start when you choose easy, hard it doesn't show you how to start only says choose game		At the start there's a video, nothing wrong at the start when you choose easy, hard it doesn't show you how to start only says choose game		1
6	It will not let me get to the home page	R	Game keeps freezing	R	4
7	Won't let me back to menu	R			
8	The game froze (The main menu)	R			
9	After the game, the main menu freezes	R			
10	It keeps freezing	R			
11	None interduction	R	None interdiction	R	1

GROUP TWO					
S/No	Problems identified	individual severity	Merged Problem	Final Severity	Frequency
1	Nothing Sliced because they were too close	Y	Numbers are too close together	R	2
2	O's are too close to numbers	G			
3	Numbers are too close together so you may slice the wrong one or not slice one at all	Y			
4	I don't exactly understand as it is too fast	Y	I don't exactly understand as it is too fast	R	1
5	I put the right answer but it said it was wrong	Y	I put the right answer but it said it was wrong	Y	1
6	It won't let you press retry	R	It won't let you press retry	Y	1
7	I don't understand what to do	R	Don't understand what to do at the start as instructions are not very clear	Y	4
8	Not very clear what to do	R			
9	Hard to understand what to do	Y			
10	The instructions aren't clear	G			
11	Don't know how to use it at the start	R			
12	It freezes	R	Game freezes especially when one finishes	R	4
13	It froze every single time so it was annoying	Y			
14	It has frozen twice	R			
15	It froze	G			
16	Froze everytime a game ended	R			
17	Freezes (won't work)	Y			
18	Every time you finish, it freezes	R			

GROUP THREE					
S/No	Problems identified	individual severity	Merged Problem	Final Severity	Frequency

<table>
<tr><td>1</td><td>It doesn't stay on the game and it glitches</td><td>R</td><td>It doesn't stay on the game and it glitches</td><td>R</td><td>1</td></tr>
<tr><td>2</td><td>There isn't a language button if you are from a different country</td><td>R</td><td>There isn't a language button if you are from a different country</td><td>R</td><td>1</td></tr>
<tr><td>3</td><td>When it doesn't click or go onto what I want I have to clear it and go back on the game</td><td>R</td><td>When it doesn't click or go onto what I want I have to clear it and go back on the game</td><td>R</td><td>1</td></tr>
<tr><td>4</td><td>It hasn't told me clearly what to do</td><td>R</td><td rowspan="2">It wasn't clear on what to do</td><td rowspan="2">Y</td><td rowspan="2">2</td></tr>
<tr><td>5</td><td>I did not get what to do on the 3 one</td><td>Y</td></tr>
<tr><td>6</td><td>Not loading game</td><td>R</td><td rowspan="5">Game freezes every time</td><td rowspan="5">R</td><td rowspan="5">4</td></tr>
<tr><td>7</td><td>It keeps on freezing every time</td><td>R</td></tr>
<tr><td>8</td><td>It won't let me retry for two times now</td><td>R</td></tr>
<tr><td>9</td><td>The game froze</td><td>R</td></tr>
<tr><td>10</td><td>It freezes every time I go to finish the game</td><td>R</td></tr>
<tr><td>11</td><td>It doesn't add right some times</td><td>G</td><td>It doesn't add right some times</td><td>G</td><td>1</td></tr>
<tr><td>12</td><td>Often it doesn't let you move it</td><td>Y</td><td>Often it doesn't let you move it</td><td>Y</td><td>1</td></tr>
<tr><td>13</td><td>It should have told you the question before it actually coming up and should have been in bolder writing</td><td>Y</td><td>It should have told you the question before it actually coming up and should have been in bolder writing</td><td>Y</td><td>1</td></tr>
<tr><td>14</td><td>It takes a while to load then it goes back to the start</td><td>Y</td><td rowspan="2">It takes a while to load</td><td rowspan="2">Y</td><td rowspan="2">2</td></tr>
<tr><td>15</td><td>It is taking a long time to load</td><td>G</td></tr>
</table>

GROUP FOUR

S/No	Problems identified	individual severity	Merged Problem	Final Severity	Frequency
1	Bit Childish	G	Bit Childish	G	1
2	At the start it goes on the internet	Y	At the start it goes on the internet	Y	1
3	When you lose, you can't restart or go to the home screen	R	Game won't restart when you lose neither will it go to home	Y	4
4	Whenever the game ends you cannot go back to home or retry	Y			
5	Not restarting	Y			
6	Won't restart game once you have failed it	Y			

GROUP five

S/No	Problems identified	individual severity	Merged Problem	Final Severity	Frequency
1	Can't tell which is 9 and which is 6	R	Can't tell which is 9 and which is 6	R	1
2	Starting again is annoying	Y	Starting again is annoying	Y	1
3	When I get a score, it doesn't show up	Y	When I get a score, it doesn't show	Y	1
4	It keeps going on to facebook	R	It keeps going on to facebook	R	1
5	Game froze three times	R	Game freezes when trying to restart after failing	R	4
6	Every time I die, I have to close the game and restart	R			
7	Won't let me restart the game when I fail it	R			
8	It won't let me replay (frozen)	R			
9	How to start a new game	G/Y	How to start a new game	G/Y	1

GROUP six - years 4 in St Augustine school					
S/No	**Problems identified**	**individual severity**	**Merged Problem**	**Final Severity**	**Frequency**
1	It sometime freezes	R	Game froze	Y/R	4
2	The problem is when it freezes	R			
3	When the level finishes, then it make it freeze	Y			
4	It went frozen	Y			
5	You can't catch up with the falling ice	R	You can't catch up with the falling ice	R	1

Appendix 4E: Full Details of Merged Problems for Each Group

GROUP 1 (Severity Rating)	GROUP 2 (Severity Rating)	GROUP 3 (Severity Rating)
MP1A – Too Confusing (R)	**MP2A** – Numbers are too close together (R)	**MP3A** – It doesn't stay on the game and it glitches (R)
MP1B – Difficult (Y)	**MP2B** – I don't exactly understand as it is too fast (R)	**MP3B** – There isn't a language button if you are from a different country (R)
MP1C – Long Intro (G)	**MP2C** – I put the right answer but it said it was wrong (possibly unreal) (Y)	**MP3C** – When it doesn't click or go onto what I want I have to clear it and go back on the game (merge to game freezes) (R)
MP1D – At the start there's a video, nothing wrong at the start when you choose easy, hard it doesn't show you how to start only says choose game (-)	**MP2D** – It won't let you press retry (merged to game freezes) (Y)	**MP3D** – It wasn't clear on what to do (Y)
MP1E – Game keeps freezing (R)	**MP2E** – Don’t understand what to do at the start as instructions are not very clear (Y)	**MP3E** – Game freezes every time (R)
MP1F – None Interduction (R) (Incomprehensible)		**MP3F** – It doesn’t add right sometimes (G) (possibly unreal)
	MP2F – Game freezes especially when one finishes (R)	**MP3G** – Often it doesn't let you move it (merge to game freezes) (Y)
		MP3H – It takes a while to load (Y)

		MP3I – It should have told you the question before it actually coming up and should have been in bolder writing (Unreal) (Y)

Appendix 4F: Full Details of Merged Problems for Each Group

GROUP 4 (Severity Rating)	GROUP 5 (Severity Rating)	GROUP 6 (Severity Rating)
MP4A – Bit Childish (G) **MP4B** – At the start it goes on the internet (Y) **MP4C** – Game won't restart when you lose neither will it go to home (Y)	**MP5A** – Can't tell which is 9 and which is 6 (R) **MP5B** – Starting again is annoying (Y) **MP5C** – It keeps going on to Facebook (R) **MP5D** – Game freezes when trying to restart after failing, so don't know how to start a new game (R) **MP5E** – When I get a score, it doesn't show (Y) (Unreal) **MP5F** – How to start a new game (G/Y)	**MP6A** – Game froze (Y/R) **MP6B** – You can't catch up with the falling ice (R)

Table 8.8 Identifying game criteria themes and matching them to existing heuristics/guidelines

S/No	Themes and Criteria	Groups	Existing Game guideline/Heuristics
MGC1	**Aesthetics/Design**		
A	It should have good theme song (3)	3	Application should react in a consistent, challenging, and exciting way to the child's actions (e.g., appropriate video clips with the music). (Alsumait & Al-Osaimi, 2009) Should use visual and audio effects to arouse interest (Federoff, 2002)
B	It should have good graphics (2)	2	The font choice, colours and sizes are consistent with good child screen design (Alsumait & Al-Osaimi, 2009)
C	You could use brightness to make game option stand out (3)	3	Use noticeable and distinct avatars that have intuitive information mappings (Pinelle et al., 2009) Make effects of the Artificial Intelligence (AI) clearly visible to the player by ensuring they are consistent with the player's reasonable expectations of the AI actor. (Desurvire et al., 2004)
MGC2	**Game Content Preference**		
A	There should be questions and the answer options should come up for you to answer the questions (6)	6	

B	There should be a football shooting at the right sum (5)	5	**Value based**
C	It could be like a guessing game (5)	5	
D	It should be comparing numbers (6)	6	
E	It should be adventurous (6)	6	
F	You should be able to build stuff (5)	5	Allow players to build content (Federoff, 2002)
MGC3	**Game Progress**		
A	If you get one wrong answer you get another go on the answer before you go back if you get it wrong again (6)	6	The e-learning program provides sufficient feedback (audio, video) to the child to provide corrective directions. (Alsumait & Al-Osaimi, 2009)
B	Add sums to a box fight by answering questions to progress (6)	6	The e-learning program includes self-assessments that advance child achievement (Alsumait & Al-Osaimi, 2009)
MGC4	**Multiplayer**		
A	You should be able to play against other characters or persons (1, 6)	1, 6	The game supports communication. The game helps the player to find other players and game instances (Korhonen & Koivisto, 2007)
MGC5	**Minimal Frustration**		

A	There shouldn't be lots to do before you start playing (2)	2	Get the player involved quickly and easily. (Federoff, 2002)
B	Do not put too many glitches in (4)	4	The interface should be as non-intrusive as possible (Federoff, 2002)
MGC6	**Fun**		
A	It should not only be about maths but it should be fun too (1, 2 & 4)	1, 2 & 4	"The e-learning program is enjoyable and interesting." (Alsumait & Al-Osaimi, 2009) "The game is fun for the Player first, the designer second and the computer third." (Desurvire et al., 2004)
MGC7	**Useful Reward**		
A	You should get rewards for doing stuff and you should be able to use your reward (1-6)	1-6	"The game should give rewards" (Federoff, 2002) "The game should give rewards that immerse the player more deeply in the game by increasing their capabilities (power-up), and expanding their ability to customize." (Desurvire et al., 2004)
MGC8	**Flexibility and Accessibility**		

A	There should be other maths problems e.g. Addition, subtraction and multiplication (1, 6)	1, 6	**Value based**
B	There should be different bikes and you should be able to change (1)	1	
C	In addition to the football, there should be other sports to choose from (5)	5	**Value base**
D	You should be able to unlock different backgrounds (4)	4	"Include a lot of interactive props for the player to interact with" (Federoff, 2002)
E	It should be for boys and girls (4)	4	**Value based**
F	The player should provide the problem and solution (3)	3	"Allow players to build content" (Federoff, 2002)
G	You should be able to fly to other places where there are right answers (6)	6	Game play should be balanced with multiple ways to win. (Desurvire et al., 2004)
MGC9	**Challenge**		
A	There should be limited amount of time to answer the questions (3, 4)	3, 4	
B	You will need to solve Maths problems to get out of a place or to move on (3, 6)	3, 6	
C	There should be different difficulty level or world to complete the game (1,4, 5, 6)	1,4, 5, 6	"Learning information is provided in layers or on different levels, in contrast to the linear approach more common to e-learning." (Alsumait & Al-Osaimi, 2009)

			“There should be variable difficulty level” (Federoff, 2002)
D	It should be challenging such that it makes you think not too easy but not too hard that you can't do it (1, 2, 4, 5)	1, 2, 4, 5	The e-learning program is easy to learn, but hard to master. The application is paced to apply pressure but not frustrate the child. The difficulty level varies so that the child has greater challenges as he develops mastery. (Alsumait & Al-Osaimi, 2009) A good game should be easy to learn and hard to master (Federoff, 2002)
MGC11	**Inspiring/Imaginative**		
A	It should be creative (2)	2	The e-learning program incorporates novel characteristics (Alsumait & Al-Osaimi, 2009)
MGC12	**Game Story**		
A	There should be a story about a person when you get the story you progress (Storyline)	6	Create a great storyline (Federoff, 2002) The game must have a base story that allows the participants’ immersion. (Villalta et al., 2011)

APPENDIX 4H: OBSERVERS DATA FOR 1^{ST} CVBIM STUDY

What problem did children encounter and how has this affected them while performing the evaluation?

Researcher's General Reflective Note

1. Girls were boy attentive than boys
2. Older Children were bored during the show and tell session as children who were not given the show and tell still understood the method process. Though the show and tell was useful for the younger children.
3. The show and tell session slowed down the evaluation process.
4. Children who had a break in their study due to insufficient time had issues recollecting problems they previously encountered.

Appendix 4H: Observers' Data

Table 7.19 Observers' Problems identified for evaluators and facilitator

S/No	Problem Area	Problem Themes	Groups	Effects of Problem Theme
1	Game Play (12)	Use of one/crossed hand (4)	1, 4, 5	Slowed down
		Confused (1)	1	
		Problem finding level (1)	2	
		Not following (7) / Disregarded (2) instruction	1	Not finding problems
				Plays another game
				Wrote two problems on one sheet
			2	Plays on different level
			3	Chose same level/game twice
			1, 4	Wrote problems at the end
			1	Not know how to play
		Game froze (10)	1,	Game play obstruction
				Helping out one another
			1, 4	Confused

			2, 4	Slowed children down
			4	Annoyance
				Lack of enjoyment (2)
			6	Kept seeking assistance
		Unclear Game Instruction (5)	3	Children did not understand instruction
		Couldn't find sub game (1)	5	Lags behind
		Study break Effect (1)	6	Unsure of problems previously encountered
		Fast paced game (2)	1, 2	Difficult to find sum
		Stating out problems found (1)	1	Potential peer bias
		Lack of Game play experience (2)	1	Slowed down
			1	Confused
			6	Unsure of action
2	Game Criteria (7)	Distracted by game sound (1)	4	Not listening to narration
		Shy Start (1)	5	-
		Taking pictures with iPad (1)	1	Not concentrating
		First time tablet user (1)	1	Not attentive to explanation
		No interaction (1)	1	Bored
		Not playing game (1)	3	
		Lack of Game play experience (3)	6	
				Confused
				Not contributing
3	Merging Phase (1)	Not following instruction (1)	1	Delayed the merging phase
4	Severity Rating (3)	Did not rate problems (1)	1	-
		No group interaction (1)	2	Difficulty agreeing on severity
		Evaluators' Effect (1)	6	

Appendix 4I: All Observed Problems for Each Group

GROUP 1

FACILITATOR

COMMENTS

S/No	Comments	Comment Association	CA Theme
1	They might not know what a mat game is	Game Criteria	Observer's inference

EVALUATORS

PROBLEMS

S/No	Problems	Identity	Problem Association	CA Theme
1	Not used to playing therefore, moving slowly and confused	C	Gameplay	Slowed down – lack of game play experience
2	Playing with one hand, therefore moving slowly	B	Gameplay	Slow movement – use of one hand
3	Using one hand to push button and it is taking time because the surface is too large- Moving slowly and confused	A	Game play	Using one hand and confused – Moving slowly
4	He is not attentive, he is playing with the tablet as this is the first time he is seeing one	C	Game criteria	Not attentive to instruction – first time tablet user
5	Getting a bit bored as there is no interaction between them	All	Game criteria	Bored – No interaction
6	Taking pictures with the iPad and not concentrating	A	Game Criteria	Playing with iPad not concentrating
7	Not looking for a problem but enjoys playing. Just wants to play and not finding problems	C	Game play	Not following instruction – own decision
8	Does not know how to play yet he does not think he needs all the instructions.	C	Gameplay	Exercising own decision
9	The game is stuck so evaluator does not know what to do	C	Game play	Game froze – confused
10	Takes time to find the sum because the numbers are running too quick on the screen	A	Game play	Fast pace game
11	Game froze so evaluators can't play anymore	All	Game play	Game froze – play obstruction
12	Plays another game not following rules	C	Game play	Not following rules

13	D's iPad is frozen so B is helping out	B, D	Game play	Helping out
14	Not following rules	C	Merging Problem	Not following rules
15	Did not rate problems	C	Severity Rating	No severity rating
16	Enjoys saying out problems found while playing, this could influence other evaluators	C	Game play	Potential peer bias
17	Chooses to write down problems at the end of the evaluation instead of doing so at the point problem was found as stated in the instruction – He could forget problems found	C	Game play	Did not follow instruction
18	Wrote down problems found at the end, this delayed other from moving on to the next phase (merging phase)	C	Others	Procedure delay
19	Not following instruction as he wrote two problems on one post it sheet instead of one problem per sheet	B	Evaluation Instruction	Not following instruction

COMMENTS

S/No	Comments	Identity	Problem Association	
1	Used to play games therefore, knows what to do	D	Game play	Game familiarity aids understanding
2	Game froze for C and A is helping out on C's iPad	A	Game play	Game froze – Helping out
3	Sitting next to each other as D is helping C when game froze	C, D	Game play	Game froze – Helping out

GROUP 2

EVALUATORS

PROBLEMS

S/No	Problems	Identity	Problem Association	CA Theme
1	Chose hard instead of medium even though she was told her twice	D	Gameplay	Did not follow game instruction
2	Game Froze, slowed her down	B	Gameplay	Slowed down – game froze

3	The numbers flying too fast, she needs time to find the correct sum	D	Game play	Fast game pace – unable to find sum
4	Used easy instead of medium so the game is so easy that she is not noting down problems	A	Game play	Did not follow game instruction
5	Problems finding the right level and this slows her down	A	Game play	Problem finding level – slowed down
6	Game froze, slows her down	B	Game play	Slowed down – Game froze
7	She became bored when game froze	B	Game play	Game froze – became bored
8	Having issues agreeing on the severity rate of the problem. There is no group interaction	All	Severity Rating	Evaluators' Effect – Severity Rate

COMMENTS

S/No	Comments	Identity	Problem Association	
1	More attentive than boys (Not a problem) and did not have problems with the game	All	Game play	Girls more attentive than boys
2	Read all the instructions	All	Game play	Reads all instructions
3	They follow the instructions	All	Game play	Follows instructions
4	Reading all the instructions but there is a problem with the iPad that she loses patience yet she wrote down notes	C	Game play	Reads instruction – write down problems
5	Reads the instructions	D	Game play	Reads instructions

GROUP 3

FACILITATOR

PROBLEMS

S/No	Problems	Problem Association	
1	Using too long sentences, children got bored	Game Criteria	Too long sentences causes boredom
2	Facilitator should explain and show children the 4 sub games as children are	Evaluation Instruction (Game Play)	Lack of clear game instruction

	confused about the sub games		

EVALUATORS

PROBLEMS

S/No	Problems	Identity	Problem Association	CA Theme
1	Moves the iPad instead of using the arrows for direction	B	Gameplay	Did not understand game instruction
2	Not playing games, she is very bored	B	Game Criteria	Bored
3	Didn’t know which sub game	D	Game play	Did not understand game instruction
4	Didn’t understand which sub game to play	B	Game play	Did not understand game instruction
5	Chose wrong level (medium instead of hard). This prevents child from finding problems because the level is easy	C	Game play	Chose wrong level -
6	Chose same sub game, therefore did not find any more problems	C	Game play	Misunderstood instruction

COMMENTS

S/No	Comments	Identity	Problem Association	CA Themes
1	Takes time to read instructions	C	Game play	Reads instruction
2	They agree well to severity colours – No problems	A, C	Severity Rating	Good severity agreement

GROUP 4

FACILITATOR

PROBLEM

S/No	Problems	Problem Association	
1	Spending too much time writing full sentence on the board. Children tend to get bored	Game Criteria	Lengthy sentences – lack of interaction - Boredom

2	Too much time explaining the faces and colours	Evaluation Instruction	Long explanations
3	Confused explanation – Facilitator made an explanation first on the iPad then went back to menu and restart	Evaluation Instruction	Confused explanation

EVALUATORS

PROBLEMS

S/No	Problems	Identity	Problem Association	CA Theme
1	Uses right hand to press buttons on the left of the iPad and this slows him down	C	Gameplay	Slow movement – use of hand on crossed side
2	He is paying attention to the sound of the iPad rather than paying attention to the narration	D	Narration	Sound distraction
3	Game froze causes children to restart and putting them behind and not knowing what to play	C, B	Game play	Game froze
4	Play using 1 hand – not that quick	All	Game play	Slow movement – uses one hand
5	Game Froze, So he is annoyed and not enjoying	B	Game play	No enjoyment – Game froze
6	Not enjoying that much	C	Game play	No enjoyment
7	Does not understand the instruction – Child says he prefers to use mark rather than pen for writing down problems	B	Evaluation Instruction	Own decision
8	Not following instruction, wrote problems after playing all games	A	Evaluation Instruction	Own decision

COMMENTS

S/No	Comments	Identity	Problem Association	
1	Uses 2 hands, so he is quick	D	Game play	Use of two hands – quick movement
2	Both use two hands – They are quick	A, B		Use of two hands – quick movement

GROUP 5

FACILITATOR

PROBLEMS

S/No	Problems	Problem Association	
1	Children are getting bored while facilitator writes on the bored	Game Criteria	No interaction - boredom
2	Facilitator should say every sub game has different levels	Evaluation Instruction (Game Play)	Need for clearer instruction

COMMENTS

S/No	Comments	Comment Association	
1	The game instruction are more clear now – children can go further with less confusion	Game play	Clear instruction – less confusion

EVALUATORS

PROBLEMS

S/No	Problems	Identity	Problem Association	CA Theme
1	They are shy at the beginning	All	Severity	Shy start
2	He is a bit behind because he couldn't find prime freeze	C	Game play	Couldn't find sub game – Lags behind
3	They are using one hand, slows down/ are in a hurry	All	Game play	Use of one hand – slows movement

COMMENTS

S/No	Comments	Identity	Problem Association	
1	Looks at each other's iPad but not to copy	C, D	Game play	N/A
2	Uses one hand	A, D	Game play	N/A
3	Uses two hands	B	Game play	N/A
4	Uses one and two hands depends if she has to press buttons for "drive"	C	Game play	N/A

5	More relax because the instructions are clear	A, C, D	Game play	Clearer instructions aids relaxation

GROUP 6

EVALUATORS

PROBLEMS

S/No	Problems	Identity	Problem Association	CA Theme
1	Lack of game play experience, could not contribute, bored and looking confused	B	Game Criteria	Lack of Game play experience – Boredom and confusion
2	Not sure what to do, kept watching others play	B	Game play2	Unsure of Action
3	Game freezes, kept seeking assistance	A, B, E	Game play2	Game froze
4	Do not play games before now, so made just one contribution based on the Maths blaster (intro game) game	B	Game Criteria	Lack of Game play experience - minor contribution
5	Break in study made children unsure of procedure and not sure of problems encountered previously.	A,B,D,E	Other	Effect of Study break
6	Could not agree on final severity	A,B,D, E	Severity Rating	Evaluator’s Effect in severity rating

Appendix 5 Chapter 9 Documents

13.5 Appendix 5 – Chapter 9 Document: Last CVBIM study with children

5A: GAME CRITERIA According to Group

S/No	GROUP 1	GROUP 2	GROUP 3
1	There should be different games for different maths problems i.e. you can either do addition or subtraction	There should be different challenge levels	There should theme
2	There should be different difficulty levels	There should be clear instructions	It should have instructions
3	You should be able to choose game type either maths only or maths and fun	Different Maths Problem game (e.g. add, subtraction)	It have a background
4	You need a catchy name	It should be fun maths not boring maths	It should have music
5	It should be the same as what is done in school	It should be for different years of children i.e. there should be different levels targeted at different age of children	There should be different levels
6	You should be able to log in, save and continue later	It should be for both boys and girls	You should receive award at the end of the game
7	Be able to challenge others (e.g. Friends)	You should be able to play online with different children	You should be able to pause the game
8	You should be able to play it on different devices	It should not just be questions to answer but it should be a quest i.e. there should be a storyline	You should be able to print your reward
9	You should be able to win prices (rewards)		It should be in a safe environment so people cn't hack into it.

10	Your scores should be able to turn to coins and you should be able to get stuff and change backgrounds		
11	You should be able to take your reward away (e.g. in the form of a certificate)		
12	You should be able to power up when you get it all right		
13	There should be video tutorial or help button to teach someone who doesn't know how to play		
14	You should be able to do it all over when you a level		
15	There should be a bonus level after you got it right that will last for a limited time		
16	There should be different modes and backgrounds		
17	You should be able to change language		
18	It should be for boys and girls		
19	There should be quite a lot of levels so you don't finish it ASAP		
20	You should be able to reset or make your own levels		

21	You should be able to send what you have done to yourself and friends		

5B: Children's Usability Problems Found

GROUP 1

S/No	**Individual problem**	**Individual severity**	**Merged Problems**	**Final Severity**
1.	It is not clear what to do	Y	It is not clear what to do	Y
2.	Sometimes the numbers get stuck behind the zeros and you can't get the numbers	G	Sometimes the numbers get stuck behind the zeros and you can't get the numbers	G
3.	On sub-zero mines, if you get it, more zeros should appear	G	On sub-zero mines, if you get it, more zeros should appear	G
4.	The music is good but it puts me off	G	The music is good but it puts me off	G
5.	When you fail, You can't go back to home	R	When you lose, it freezes such that you can't go back to home	R
6.	When you lose you freeze	R		
7.	It's hard to move the ice cubes	G/Y	It is hard to drag the ice cubes	G/Y
8.	It's hard to move the ice cubers	G/Y		
9.	Hard to drag	Y/R		

GROUP 2

S/No	Individual problem	Individual severity	Merged Problems	Final Severity
1.	The ice cubes fall too fast and reaches the top of the screen too quickly	Y/G	Try not to let numbers go too fast as the ice cube numbers fall too fast	G
2.	Faster	G		
3.	Try not to let numbers go too fast	G		
4.	I don’t know how to do it	R	I don’t understand the game	Y/R
5.	I do by tscyo	Y/R		
6.	I don’t understand the game	Y/R		
7.	I don’t understand why it is called cool 21	G	I don’t understand why it is called cool 21	G
8.	A bit too easy for older children	G	A bit too easy for older children	G
9.	After a while it stops and won’t let you click anything	R	It freezes after a while that you can’t click retry	R
10.	On your second go it freezes again and won’t let you play	R		
11.	You can’t click the retry button	R		
12.	Pcos it won’t work	R		
13.	Won’t let you retry game	R		

GROUP 3

S/No	Individual problem	Individual severity	Merged Problems	Final Severity
1.	Make sure you can pause it	Y	Make sure you can pause it	Y
2.	When you make to sometimes it doesn’t work	R	When you make it sometimes it doesn’t work	R
3.	It won’t let you press play again	Y	It won’t let you retry	Y/R
4.	Make the buttons easier to use	Y		
5.	Make sure you can try again	R		
6.	It won’t let you retry	R		
7.	It keeps going on something else	Y	It keeps going on something else	Y

5C: OBSERVED PROBLEMS

S/No	GROUP 1		GROUP 2		GROUP 3	
	Observed Problems (Identity)	**Problem Area**	**Observed Problems**	**Problem Area**	**Observed Problems**	**Problem Area**
1	Take a long time to contribute to the conversation (C, D)	Game Criteria	'A' tried to help 'D' think about what they would like in a game (D)	GC	Went into Facebook from the main menu (B)	GP2
2	D made point that has already been made by A (D)	GC	'C' stated that it is difficult to think of what would make a good maths game	GC	Went into the wrong sub game (A)	GP2
3	Couldn't get his sound working (D)	GP	'B' ended up loading Facebook from within the menu	GP	Had to be told by A not to touch the Q'S on sub-zero game (D)	GP2
4	Unsure how to replay the sub-zero mine game (All)	GP	'C' had to read and explain the instructions to 'D'	GP	Could not restart the game once on the game over screen (ALL)	GP2
5	Game Froze and IPad screen kept rotating the wrong way (C)	GP	Could not start new level once the game was complete (All)	GP	Struggled to choose game options without being shown (C)	GP2
6	Unsure of object that was not a number in cool 21 game (D)	GP	'C' has to explain how to record the problems to (D)	GP	Frostris "I don't get this" (D)	GP2
7	Wrote more than one problem on a post it note (B)	GP	Didn't know how to get back to the main menu from sub game start screen (C, B)	GP	Stated they had a problem but did not write it down (D,B)	MP
8	Struggled to play ice cube game. Didn't seem sure of the controls (C)	GP	'C' stated they didn't quite understand the game	GP	Struggled talking about SR's but did seem to understand them (ALL)	SR
9	Talked about ratings as how good /easy/ hard the game was	SR	Started writing second problem on the same post it note (A, B, C)	GP		

	rather than about the problems (All)					
			‘C’ seemed to ‘control’ what ‘D’ did throughout the study	Other		
			D received issue trying to sign into Apple Game Centre	GP		
			‘C’ struggled to play the frostris sub game – didn’t know what to do	GP		
			C and D Wrote the same problem multiple times for different sub games	MP		

5D: Children's Game Criteria matched to Existing Heuristics

S/No	Theme and Criteria	Group	Existing Heuristics
1	**Flexibility and Accessibility**		
A	There should be different games for different maths problems i.e. you can either do addition or subtraction (flexibility)	1,2	
B	You should be able to play it on different devices (flexibility)	1	The e-learning program may be used on a variety of equipment and platforms such as laptops, PDA. (Alsumait & Al-Osaimi, 2009)
C	You should be able to reset or make your own levels when you finish a level so you can do it all over. (flexibility)	1	Make the game Replayable (Federoff, 2002)
D	It should be for different years of children i.e. there should be different levels targeted at different age of children (Flexible)	2	
E	You should be able to change language (Flexibility and Accessibility)	1	
F	It should be for boys and girls (Flexibility)	1,2	
2	**Challenge**		
A	There should be lots of level so you don't finish it ASAP but let the difficulty levels be different, also you should be able to create your own level (Challenge)	1-3	There should be variable difficulty level (Federoff, 2002) Vary the difficulty level so that the player has greater challenge as they develop mastery. Easy to learn, hard to master. (Desurvire et al., 2004)s
3	**Fun**		
A	It should be fun Maths not boring maths so you should be able to choose game type either maths only or maths and fun (Fun)	1,2	
4	**Aesthetics and Design**		
A	You need a catchy name/theme (Aesthetics /Design)		
B	It should have music (Aesthetics/Design)	3	
C	There should be different modes and backgrounds (Aesthetics)	1,3	

5	**Match to Real World**		
A	It should be the same as what is done in school (Match to Real World)	1	The e-learning program interface employs simple words, phrases and concepts familiar to the child and makes information appear in a natural and logical order. (Alsumait & Al-Osaimi, 2009)
6	**User Control**		
A	You should be able to pause the game (User control)	3	
B	You should be able to log in, so you can save and continue later (User Control)	1	Players should be able to save games in different states. (Federoff, 2002) The child can easily turn the application on and off, and can save his user profile in different states. (Alsumait & Al-Osaimi, 2009)
7	**Multiplayer**		
A	Be able to challenge others (e.g. Friends) (Multiplayer)	1,2	
8	**Useful Reward**		
A	You should be able to win prices (rewards) when you get right, your scores should also turn to prices that you can use to get stuff, change backgrounds and get power ups. And be able to take your reward away (useful reward)	1,3	Rewards are meaningful (Alsumait & Al-Osaimi, 2009) The game should give rewards that immerse the player more deeply in the game by increasing their capabilities (power-up), and expanding their ability to customize. (Desurvire et al., 2004)
B	There should be a bonus level after you got it right that will last for a limited time (Reward)	1	
9	**Security**		
A	It should be in a safe environment so people can't hack into it. (Security)	3	
10	**Game Story**		
A	It should not just be questions to answer but it should be a quest i.e. it should have a storyline (Game Story)	2	

11	**Help and Instruction**		
A	There should be video tutorial or help button to teach someone who doesn't know how to play and let the instructions be clear (Help and Instruction)	1-3	The game should give hint but not too many (Federoff, 2002) The e-learning program includes interesting tutorials or flashes that mimic lessons in the program. (Alsumait & Al-Osaimi, 2009)
12	**Sharing**		
A	You should be able to send what you have done to yourself and friends (Sharing)	1	

14 PUBLISHED WORK

On the following page is a work inspired and published by this thesis.