INTERNET CROWDSOURCING FOR GENERATIVE DESIGN

HAO WU

This Thesis is submitted to the Department of Design, Manufacture and Engineering Management, University of Strathclyde, for the degree of Doctor of Philosophy.

COPYRIGHT STATEMENT

AUTHOR’S DECLARATION

This thesis is the result of the author’s original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Singed

Date: 28/11/2016

Thank you very much, my mum and dad.

Thank you for your wholeheartedly dedication to me.

I love you.

ACKNOWLEDGEMENT

In the first place, I specially and strongly thank my supervisor, Professor Jonathan Corney. Since I was an undergraduate student in the department of Design, Manufacture and Engineering Management, Professor Corney helped me a lot whatever on my study or about my life till now. Because of his acceptance of my PhD application, I would have chance to start and continue my research life. Professor Corney gives me not only his academic support to my PhD research, but also so many wonderful chances, for example, encouraging and supporting me to attend international academic conferences in my research area, and visiting Sichuan University in China (where I graduated from) for international collaborations for so many times.

I would like to thank Professor Yi Qin. From the first day when I made up my mind to study in Strathclyde University, Professor Qin warmly helped me how to start a completely new life in Glasgow, and how to study in a different education system. Especially at the beginning of my PhD study, Professor Qin kindly gave me many suggestions on how to be a PhD student, and how to start a PhD study.

Specially thank Dr Michael Grant. During my PhD research work, Dr Grant always shared his enthusiastic ideas and suggestions, and supported me for carrying out my experiments.

I also hugely appreciate to Mr Colin Andrews, Dr Gokula Vijayumar Annamalai Vasantha and Dr Andrew Lynn. Especially thank them for sharing their expertise in Internet technology and statistical analysis.

Special mention goes to Senyong Chen, Yankang Tian, Jie Zhao, Quanren Zeng, Song Yang, Xiaoyu Liu, Kunlan Huang, Mutian Li, Xinyu Yang, and all my colleagues and friends in the department of Design, Manufacture and Engineering Management. Thank you all for sharing happiness and sorrows with each other during these years. Because all of you, my life in Glasgow was rich and happy.

Finally, but by no means least, thanks to all members in Group 6 of the “SWB Brazilian Summer School Group Project”: Bruna de Alvarenga Massarioli, David Souza Pinto, Felipe Camara de Alvarenga, Gabriel de Oliveira Cordeiro and Guilherme Augusto Eduardo Aguiar. Chapter 5 of the thesis incorporates some results generated by them.

PUBLICATIONS

An Evaluation Methodology for Crowdsourced Design

Hao Wu, Jonathan Corney, Michael Grant

Published in journal: Advanced Engineering Informatics
Volume 29, Issue 4, Pages 757-1126 (October 2015)

Relationship Between Quality and Payment in Crowdsourced Design

Hao Wu, Jonathan Corney, Michael Grant

Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD),
Published in**:** 2014 Computer Supported Cooperative Work in Design (CSCWD)
Electronic ISBN: 978-1-4799-3776-9; USB ISBN: 978-1-4799-3775-2
Print on Demand (PoD) ISBN: 978-1-4799-3777-6

Crowdsourcing Measures of Design Quality

Hao Wu, Jonathan Corney, Michael Grant

Proceedings of 34th Computers and Information in Engineering Conference, ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference
Published in: 34th Computers and Information in Engineering Conference
ISBN: 978-0-7918-4629-2

The Application of Crowdsourcing for 3D Interior Layout Design

Hao Wu, Jonathan Corney, Michael Grant

Proceedings of the 20th International Conference on Engineering Design (ICED 15)
Published in: the 20th International Conference on Engineering Design (ICED 15)
ISBN: 978-1-904670-67-4; ISSN: 2220-4334

ABSTRACT

In recent years, the ‘‘power of the crowd” has been repeatedly demonstrated and various Internet platforms have been used to support applications of collaborative intelligence in tasks ranging from open innovation to image analysis. However, crowdsourcing applications in the fields of design research and creative innovation have been much slower to emerge. So, although there have been reports of systems and researchers using Internet crowdsourcing to carry out generative design, there are still many gaps in knowledge about the capability and limitations of the technology. Indeed the process models developed to support traditional commercial design (e.g. Pugh’s Total Design, Agile, Double-Diamond etc.) have yet to be established for Crowdsourced Design (cDesign). As a contribution to the development of such a general model this thesis proposes the cDesign framework to support the effective use crowdsourcing for generative design. Within the cDesign framework the effective evaluation of design quality is identified as a key component that not only enables the leveraging of a large, virtual workforce’s creative activities but is also fundamental to almost all iterative optimisation processes.

This thesis first describes a brief history of the internet crowdsourcing and how crowdsourcing has been used in the design domain. Then the aims and objectives of this research work are listed which are followed by the methodologies chosen for the research. After that three crowdsourced design case studies are presented and discussed. The results of these case studies are integrated to establish the cDesign framework. A fourth crowdsourced design case study on the validation of this cDesign framework is described before a final discussion on the significance of this research work presented. The thesis concludes, by identifying the limitations of the work and opportunities for future research.

Acronyms

CDEC: crowdsourced design evaluation criteria

cDesign Framework: crowdsourced design framework

HBGA: Human-based Genetic Algorithms

IDT: Iterative Design Task

MTurk: Amazon's Mechanical Turk

NIDT: Non-iterative Design Task

CONTENTS

CHAPTER 1 1

INTRODUCTION 1

I. BACKGROUND OF CROWDSOURCING & CROWDSOURCED DESIGN 2

1. Crowdsourcing 2
2. Origins of Crowdsourced Design 3

II. MOTIVATION & PURPOSES 4

III. AIMS & OBJECTIVES 5

1. Aims 5
2. Objectives 5

IV. STRUCTURES OF THE THESIS 6

V. CHAPTER SUMMARY 7

CHAPTER 2 11

LITERATURE REVIEW 11

I. INTRODUCTION 12

1. Crowdsourcing Definition 12

II. CROWDSOURCING AS AN INNOVATION METHODOLOGY 13

1. Crowdsourcing as the Scientific Method 13
2. Open Innovation and Crowdsourcing 15
3. Linguistic Study 16

III. CROWDSOURCING PROCESS 17

1. Crowdsourcing Platforms Used for Creative Design 17
2. Assessment Methods & Reward Models 18

IV. HUMAN-BASED GENETIC ALGORITHMS 25

V. DESIGN METHODOLOGY 27

1. Total Design (Stuart Pugh) 27
2. Ulrich & Eppinger 29
3. Wideblue: design process 30

VI. KNOWLEDGE GAPS 31

VII. CHAPTER SUMMARY 32

CHAPTER 3 39

RESEARCH DESIGN & 39

METHODOLOGY 39

I. INTRODUCTION 40

II. RESEARCH PARADIGM 40

III. RESEARCH QUESTIONS & APPROPRIATE METHODOLOGIES 43

1. What Question to Study? 43

2. What Data are Relevant and What Data to Collect? 44

3. How to Analyse the Results? 45

IV. RESEARCH VALIDATION 46

1. Construct Validity 46

2. Internal Validity 46

3. External Validity 46

V. CHAPTER SUMMARY 48

CHAPTER 4 49

CASE STUDY A: A DESK LAMP DESIGN TASK & 49

CASE STUDY B: A LIVING ROOM LAYOUT DESIGN TASK 49

I. INTRODUCTION 50

II. Case Study B: Relationship between Design Quality and Payment to Workers 56

2.1 Living Room Layout Design Task (for Creating Designs) 57

2.2 Qualifying Design Quality: Living Room Layout Evaluation Task (for Design Evaluation) 60

III. RESULTS OF CASE STUDY B 61

3.1 Demographic Information 62

3.2 Design Evaluation Results 63

IV. DISCUSSION 64

V. CHAPTER SUMMARY 70

CHAPTER 5 73
EXPERIMENTAL CASE STUDY C 73
3D KITCHEN LAYOUT DESIGN.. 73
I. INTRODUCTION.. 74
II. EXPERIMENT PROCESS.. 74
1. Stage 1.. 75
2. Stage 2.. 76
3. Stage 3 & 4 76
III. EXPERIMENTAL DESIGN - 3D INTERIOR DESIGN EXPERIMENT 77
1. Stage 1: Specification 77
2. Stage 2: Validation (Prototype)... 79
3. Stage 3: Execution.. 83
4. Stage 4: Evaluation... 83
IV. DESIGN COMBINATION AND RESULTS... 85
1. 1st Generation Layout Design and Evaluation Results 85
2. Design Combination and Evaluation Results.. 85
V. DISCUSSION 92
VI. CHAPTER SUMMARY.. 93
CHAPTER 6 95
CDESIGN FRAMEWORK... 95
I. INTRODUCTION.. 96
II. CROSS CASE ANALYSIS 98
III. COMPARISONS OF EXPERIMENTAL CASES FINDINGS..................................... 100
3.1 Experimental Case One: Desk Lamp Design Task.. 100
3.2 Experimental case two: living room layout design task 100
3.3 Experimental case three: 3D kitchen room layout design task.......................... 100
IV. GENERAL CDESIGN FRAMEWORK AND DETAILED PROCESSES 103
1. Stage 1... 104
2. Stage 2... 105

3. Stage 3 & 4 105
4. Stage 3&4, Non-iterative Design Task (NIDT) 106
5. Stage 3&4, Iterative Design Task (IDT) 107
V. CHAPTER SUMMARY 109
CHAPTER 7 111
VALIDATION OF THE CDESIGN FRAMEWORK 111
CASE STUDY D: 111
CAR KEY FOB DESIGN EXPERIMENT 111
I. INTRODUCTION 112
II. METHODOLOGY AND TOOLS 113
III. CAR KEY FOB DESIGN EXPERIMENT 114
IV. EXPERIMENT RESULTS 119
1. 1st Generation Creation Drawings 119
2. 1st Generation Evaluation Task 120
3. 2nd Generation Creation Task 124
4. 2nd Generation Evaluation Task 124
5. 3rd Generation Creation Task 127
6. 3rd Generation Evaluation Task 129
7. 4th Generation Creation Task 129
8. 4th Generation Evaluation Task 131
V. DISCUSSION & CHAPTER SUMMARY 132
CHAPTER 8 134
DISCUSSION 134
I. INTRODUCTION 135
II. DISTRIBUTION TRENDLINES OF EACH EXPERIMENT 135
1. Distribution Trendlines in Living Room Layout Design Task & Car Key Fob Design Task 139
2. Evaluation ranking distribution in 3D kitchen room layout design task 147
III. CHAPTER SUMMARY 155

CHAPTER 9 156

CONCLUSION, LIMITATION & 156

FUTURE WORK 156

I. INTRODUCTION 157

II. RESEARCH SUMMARY 158

III. CONTRIBUTION TO KNOWLEDGE 161

IV. LIMITATIONS 164

V. FUTURE WORK & CONCLUSION 166

APPENDIX 1: All living room layout designs generated by MTurk workers in case study A 169

APPENDIX 2: All 1st generation car key fob designs generated by MTurk workers in case study D 175

APPENDIX 3: A report of demographic information of crowd workers in this research 181

1. Gender 182

2. Education Level 184

3. Nationality 185

4. Design Experience 187

5. Submission Time 193

6. The Acceptance for Submissions 195

FIGURES

Figure 2 - 1 Example tasks of invitation of tenders on Taskcn.. 19

Figure 2 - 2 Screenshot of GE jet engine bracket challenge 20

Figure 2 - 3 Examples of submitted bracket 22

Figure 2 - 4 Yu's HBGA process 26

Figure 2 - 5 Total Design model and process 28

Figure 2 - 6 Product design specification (PDS) 29

Figure 2 - 7 Ulrich & Eppinger design process 29

Figure 2 - 8 Wideblue design process 30

Figure 3 - 1 Beech's research design map (2005) 41

Figure 3 - 2 The choice of research methods related to ontology, Beech (2005) 42

Figure 3 - 3 Research methodology illustration 47

Figure 4 - 1 Desk lamp design task workflow 50

Figure 4 - 2 Case study A: examples of crowdsourced designs from the first generation 54

Figure 4 - 3 Case study A: best five desk lamp designs from the first generation 54

Figure 4 - 4 Case study A: second generation desk lamp designs 55

Figure 4 - 5 Case study A: the best 3 desk lamps 55

Figure 4 - 6 Case Study B: Main stages of the experiment: 1. Requesters upload problems to MTurk; 2. Turkers find tasks via MTurk; 3. Google Drawing is the tool for participants to draw the living room layout; 4. Workers upload their results to Requesters as well as share the drawings via Google Drawing; 5. Once solutions are approved, Payment will be given to Turkers by MTurk. 59

Figure 4 - 7 Case Study B: examples of the approved living room layout designs 63

Figure 4 - 8 Relationship between the number of designs v.s. time 65

Figure 4 - 9 Spread of average design marks v.s. payment level 65

Figure 4 - 10 Spread of ranking v.s. payments where best designs have low ranks (i.e., 1st or 2nd).. 68

Figure 4 - 11 Best Fit line: average ranking v.s. payment 68

Figure 4 - 12 Best Fit Line: most highly ranked designs v.s. payment 69

Figure 4 - 13 Examples of the best ranking layouts 69

Figure 5 - 1 The workflow of the 3D kitchen layout design experiment.............................. 80

Figure 5 - 2 Examples of the kitchen designs by Homestyler 80

Figure 5 - 3 Ten evaluation criteria 84

Figure 5 - 4 Ten 1st generation layout designs, link for high quality image 90

Figure 5 - 5 Layout Combination Progress 91

Figure 5 - 6 Scores of the best 3 designs from generation 1 and the generation 2 (combined designs) 92

Figure 6 - 1 Integration and triangulation of research findings in case studies.................... 99

Figure 6 - 2 The cDesign methodology.. 103

Figure 6 - 3 The workflow and detailed process for Non-iterative Design Task (NIDT)....... 106

Figure 6 - 4 The workflow and detailed process for Iterative Design Task (IDT) by using HBGA (to create designs) and cDEC (to evaluate designs)... 107

Figure 7 - 1 The workflow of the car key fob design experiment 114

Figure 7 - 2 Examples of the 1st generation drawings.. 120

Figure 7 - 3 Example of design ranking form 1 123

Figure 7 - 4 Example of design tanking form 2 123

Figure 7 - 5 2nd generation creation task (the combination process using the best three designs from the 1st generation).. 124

Figure 7 - 6 Evaluation rankings results of the 2nd generation designs 125

Figure 7 - 7 10 drawings of group 1 designs – 2nd generation, combination 1 125

Figure 7 - 8 10 drawings of group 2 designs – 2nd generation, combination 1 126

Figure 7 - 9 10 drawings of group 3 designs – 2nd generation, combination 1 126

Figure 7 - 10 The best three designs from all 2nd generation car key fob designs.............. 126

Figure 7 - 11 3rd generation creation task (the combination process using the best three designs from the 2nd generation)... 127

Figure 7 - 12 10 drawings of group 1 designs – 3rd generation, combination 2 128

Figure 7 - 13 10 drawings of group 2 designs – 3rd generation, combination 2 128

Figure 7 - 14 10 drawings of group 3 designs – 3rd generation, combination 2 128

Figure 7 - 15 The best three drawings in the 3rd generation designs 129

Figure 7 - 16 4th generation creation task (the combination process using the best three designs from the 3rd generation) 130

Figure 7 - 17 10 drawings of group 1 designs – 4th generation, combination 3 130
Figure 7 - 18 10 drawings of group 2 designs – 4th generation, combination 3 131
Figure 7 - 19 10 drawings of group 3 designs – 4th generation, combination 3 131
Figure 7 - 20 The best three 4th generation drawings... 132
Figure 7 - 21 Evaluation ranking results of the best designs from each generation 133

Figure 8 - 1 A normal distribution showing the design quality and the frequency 137
Figure 8 - 2 Evaluation does not distinguish a standard performance 137
Figure 8 - 3 Trendline for the average ranking results... 138
Figure 8 - 4 The normal probability plots, trendlines and r^2 information of group 1 average ranking results in living room layout design task... 141
Figure 8 - 5 The normal probability plots, trendlines and r^2 information of group 1 average ranking results in living room layout design task... 142
Figure 8 - 6 The normal probability plots, trendlines and r^2 information of the average ranking results in car key fob design task (1st generation) 145
Figure 8 - 7 The normal probability plots, trendlines and r^2 information of the average ranking results in car key fob design task (2nd generation)... 146
Figure 8 - 8 The normal probability plot of the 1st generation designs in 3D kitchen room layout design task 151
Figure 8 - 9 The normal probability plots for the group 1 of the 2nd generation evaluation rankings.. 152
Figure 8 - 10 The normal probability plots for the group 2 of the 2nd generation evaluation rankings.. 152
Figure 8 - 11 The normal probability plots for the group 3f the 2nd generation evaluation rankings.. 153
Figure 8 - 12 The normal probability plots for the top 6 rankings..................................... 154
Figure 8 - 13 The normal probability plots for the final ranking.. 154

Figure 9 - 1 The cDesign framework 162
Figure 9 - 2 Walker's Design Family Tree 166

Appendix 3- 1 Gender Proportion in all tasks 182
Appendix 3- 2 Gender proportion of design tasks... 183
Appendix 3- 3 Gender proportion of design evaluation tasks.. 184

Appendix 3- 4 Education level in all experiments. 185
Appendix 3- 6 An example of the design experience options of collection form............... 188
Appendix 3- 7 An example of the collection results form 189
Appendix 3- 8 Exact nationality of crowd worker in all experiments................................. 190
Appendix 3- 9 Continent of crowd workers 191
Appendix 3- 10 Number of participants in different continents 192
Appendix 3- 12 All design experience proportion 193
Appendix 3- 14 Submission time information and the distribution curve 195
Appendix 3- 15 The acceptance information... 196

TABLES

Table 2 - 1 Reported components of crowdsourced design tasks - 1 31

Table 2 - 2 Reported components of crowdsourced design tasks - 2 32

Table 3 - 1 Research methodology and technique 47

Table 4 - 1 correlations between payment and the average marks 66

Table 4 - 2 Ranking Results 67

Table 5 - 1 Scores for the 1st generation designs 90

Table 5 - 2 Scores for the combination designs 91

Table 6 - 1 Key elements preparation 101

Table 6 - 2 Specification Tasks 104

Table 6 - 3 Validation Tasks 105

Table 8 - 1 Average ranking data in each group (i.e., assessment of 10 MTurk workers) for the living room layout design task 140

Table 8 - 2 Average ranking data for each generation in car key fob design task 143

Table 8 - 3 R^2 information of the 1st generation evaluation ranks 144

Table 8 - 4 R^2 information of the 2nd, 3rd and 4th generation evaluation ranks 144

Table 8 - 5 Evaluation scores of the 1st generation 3D kitchen room layout designs 147

Table 8 - 6 Ranking results of the 1st generation 3D kitchen room layout designs 149

Table 8 - 7 The average of the 1st generation 150

Table 8 - 8 The average rankings of the 2nd generation designs in 3D kitchen room layout design task 151

Table 8 - 9 Average rankings in the top 6 design evaluation stage and the final design evaluation stage 153

Appendix 3- 5 All collected nationality information from crowd workers 185

Appendix 3- 11 Design experience information of participants 192

Appendix 3- 13 Submission time information 194

CHAPTER 1

INTRODUCTION

I. BACKGROUND OF CROWDSOURCING & CROWDSOURCED DESIGN

1. Crowdsourcing

Many commercial design tasks have not been done by lone individuals activity for many years, but rather carried out by groups of various people (i.e., marketing, designers, manufacturers, engineers, purchasing managers, salesmen and after-sale service workers, etc.) (Hsiao & Chou 2004)(Cooper & Kleinschmidt 1986)(Dougherty 1992). So different kinds of people play their respective roles in a complex design processes. Perhaps this number of people could be regarded as a crowd, that is to say, design and creativity has been seen attributable to the crowd for many years. However, over time, the nature of the "crowd" has been changing.

In 2006, "crowdsourcing" was defined by Jeff Howe as "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call"(Howe 2006). However, these people do not have the same composition as the internal company "crowd" discussed earlier. This new type of "crowd" is made up by anonymous groups (Yochai Benkler 2006). In this crowd, members do not know each other; they usually undertake tasks individually, and then if their results are accepted, they will be rewarded by typically small amounts of money. (Kittur et al. 2008). Crowdsourcing groups include online product communities (Brabham 2009)(Jeppesen & Frederiksen 2006)(Kozinets et al. 2008), virtual communities of special interests (Hogue 2011), the general public (Chilton 2009)(Haklay & Weber 2008), and employees who typically would not participant in the tasks to be completed (Stewart et al. 2009).

Since 2006, crowdsourcing has been applied to many tasks, but there has been surprisingly little work reported on the distribution of design tasks.

2. Origins of Crowdsourced Design

One of the key methodologies, known as Human-based Genetic Algorithms (HBGA), that has been widely applied in Crowdsourced design was first reported almost five years. Since then HBGAs have been employed in some of the most impressive examples of collaborative, innovation tasks (Yu & Nickerson 2011)(Yu & Nickerson n.d.)(Yu & Sakamoto 2011). Interestingly the origins of the approach pre-date the internet and are based on research which showed that controlled combinations could develop creativity (Osborn 1957)(Amabile n.d.). From this an online combination process based on the HBGA was established (Yu & Nickerson n.d.)(Yu & Nickerson 2011)(Bao et al. 2011).

The basic concept of an HGBA is not hard to grasp; essentially new ideas are basically separated into different generations. In the first generation, participants from the crowdsourcing platform create the first group of designs. Then a second crowd evaluates the first generation and chooses several pairs for the combination process to construct the second generation (i.e., generation 1 evaluation). In generation 2, some of the ideas were selected directly from the top ranked generation 1 designs, and others were collected by combining pairs chosen from the first generation (i.e., generation 2 combination). Then, the third generation applies the evaluation process to the second generation combination process again to create generation 3 (Yu & Nickerson n.d.)(Yu & Nickerson 2011). So, iteratively, generation after generation, new ideas could be sequentially created.

II. MOTIVATION & PURPOSES

This research is motivated by the almost universal need of original equipment manufacturers (OEMs) to innovate new products and services. Although the procedures for enabling employee led design and innovation are well established it has become increasingly clear that this approach is often limited by a certain amount of "group think" and, often, a lack of resource or time to allow truly innovative solutions. One solution to these problems is to look outside the company, beyond the employees and facilitate external workers to provide design concepts. This approach has been termed "open innovation" and although it can take many forms (such a social media discussion or customer focus groups) the use of crowd sourcing is an obvious method.

However exactly how commercial crowdsourcing sites can be employed in product development is far from clear. So while the academic literature has demonstrated that crowds can design (Yu & Nickerson 2011)(Yu & Sakamoto 2011)(Nickerson et al. 2011) and also evaluate designs (Bao et al. 2011)(Herr et al. 2011). It is not clear how these technologies can be combined in to an overall process.

This context has given rise to the opportunity to extend traditional design methodologies and use new internet technology for creative activities. Because both the generation and evaluation of designs can be done by the crowd there is potential to explore large numbers of concepts (much larger than would typically be considered by an internal design team). So the motivations of this research work are:

1) The commercial need for tools and methods to improve the volume and variety of design concepts generated.
2) The engineering need for tools and methods to enable effective evaluation of large numbers of candidate designs.

Give this motivation the research hypothesis is that:

1) Internet crowdsourcing technology can be used to create effective open innovation process which are all executed by crowd workers (including design process and design evaluation process).
2) A crowdsourced design framework can be identified that is sufficiently general it can support the creation of a wide range of design tasks

III. AIMS & OBJECTIVES

1. Aims

The reported research literature suggests less is known about crowdsourcing applications in design area, the aims of this investigation are:

1) Conduct a number of case studies to identify the choice and parameters inherent in using crowdsourcing for design
2) Identify and generalise the process of creating crowdsourced design tasks
3) Apply the proposed framework to a design task and assess its completeness and effectiveness.

2. Objectives

To reach the aims of this research, the following objectives need to be investigated:

1) **Understand the mechanism** of crowdsourcing, the crowdsourcing platform and the basic processes of crowdsourced design.
2) **Investigate the relationships between parameters** such as payment, quality, quantity and evaluation in Crowdsourced design.
3) **Establish if there are limits** to the design representations that can be supported by commercial crowdsourcing platforms. Specifically investigating if 3D, as well as 2D, designs can be crowdsourced.
4) **Establish the general crowdsourced design model** (both design process and design evaluation process are executed by crowd workers, including collecting the crowdsourced design evaluation criteria), and detail each stage.
5) **Validate the crowdsourced design model** using a large scale design case study.

IV. STRUCTURES OF THE THESIS

After determining the aims and objectives of this research, the structure of the thesis is shown as follows:

1) Introduction:

   In the Introduction chapter, an overall view of crowdsourcing and crowdsourced are described. Then the motivations, aims and objectives of this research are discussed. Finally the structure of the thesis is shown.

2) Literature review:

   In the Literature Review chapter, the details of the development of crowdsourcing and crowdsourced design are described. And the research gap is discussed to guide the following research work.

3) Methodologies:

   In the Methodologies chapter, the research methodology and research design are discussed, as well as the exact methods and techniques applied in this research.

4) Experimental case study A & B:

   In the case study A & B chapter, two case studies (case study A: desk lamp design task, case study B: living room layout design task) on crowdsourced design are described and discussed. These case studies were designed to investigate the objectives 1 and 2.

5) Experimental case study C:

   In the case study C chapter, one crowdsourced design experiment (3D kitchen room layout design task) is described and discussed to investigate the research objective 3.

6) Crowdsourced Design (cDesign) framework

   In the establishment of cDesign framework chapter, a general crowdsourced design model is integrated from the results of case studies A, B & C. The establishment of cDesign framework solve the objective 4.

7) Validation of the framework: experimental case study D

   In the cDesign framework validation chapter, one crowdsourced design task (car key fob design task) is described and discussed. Case study D investigate the research objective 5.

8) Discussions

   In the Discussion chapter, the demographic information collected from crowdsourcing platforms are discussed. Then a statistical analysis of the design evaluation used in each case studies is discussed.

9) Conclusion, limitation & future work

   In the last chapter, this research work is concluded. The limitations and the future work are described.

V. CHAPTER SUMMARY

In this chapter, an overall view on crowdsourcing and crowdsourced design is briefly described. Then the motivation of the research is discussed, followed by the determination of the aims and objectives of the research. Finally the structure of this thesis is illustrated. To well understand the development of crowdsourcing in design and the related areas, the literature review of this research is shown in the next chapter.

Reference

Amabile, T., *Creativity in Context: Update to the Social Psychology of Creativity*, Westview Press, June 7, 1996.

Bao, J., Sakamoto, Y. & Nickerson, J., 2011. Evaluating Design Solutions Using Crowds. In *Seventeenth Americas Conference on Information Systems, August 4th-7th, Detroit, Michigan*. Available at:

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2201651 [Accessed June 25, 2013].

Brabham, D., 2009. Crowdsourced advertising: how we outperform Madison Avenue. *Flow: A Critical Forum on Television and Media Culture*. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Crowdsourced+Advertising:+How+We+Outperform+Madison+Avenue#0 [Accessed July 17, 2013].

Chilton, S., 2009. Crowdsourcing is radically changing the geodata landscape: Case study of OpenStreetMap. *24th International Cartographic Conference*. Available at: http://raptor1.bizlab.mtsu.edu/S-Drive/DMORRELL/Mgmt 4990/Crowdsourcing/22_6.pdf [Accessed July 17, 2013].

Cooper, R.G. & Kleinschmidt, E.J., 1986. An Investigation into the New Product Process: Steps, Deficiencies, and Impact. *Journal of Product Innovation Management*, 3(2), pp.71–85. Available at: http://doi.wiley.com/10.1111/1540-5885.320071.

Dougherty, D., 1992. Interpretive Barriers To Successful Product Innovation In Large Firms. *Organization Science*, 3.

Haklay, M. & Weber, P., 2008. Openstreetmap: User-generated street maps. *Pervasive Computing, IEEE*, pp.12–18. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4653466 [Accessed July 17, 2013].

Herr, C. et al., 2011. THE ROLE OF COLLECTIVE INTELLIGENCE IN DESIGN. , pp.1–10. Available at: http://web.arch.usyd.edu.au/~mpau4467/papers/11_PauliniMurtyMaher.pdf [Accessed June 25, 2013].

Hogue, C., 2011. Crowdsourcing for science. *Chemical & Engineering News*. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:CROWDSOURCING+FOR+SCIENCE#7 [Accessed July 17, 2013].

Howe, J., 2006. The rise of crowdsourcing. *Wired magazine*, (14), pp.1–5. Available at: http://sistemas-humano-computacionais.wikidot.com/local--files/capitulo:redes-sociais/Howe_The_Rise_of_Crowdsourcing.pdf [Accessed June 22, 2013].

Hsiao, S.-W. & Chou, J.-R., 2004. A creativity-based design process for innovative product design. *International Journal of Industrial Ergonomics*, 34(5), pp.421–443. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0169814104000873 [Accessed May 29, 2013].

Jeppesen, L.B. & Frederiksen, L., 2006. Why Do Users Contribute to Firm-Hosted User Communities? The Case of Computer-Controlled Music Instruments. *Organization Science*, 17(1), pp.45–63. Available at:

http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1050.0156 [Accessed May 22, 2013].

Kittur, A., Chi, E.H. & Suh, B., 2008. Crowdsourcing user studies with Mechanical Turk. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, p.453. Available at: http://portal.acm.org/citation.cfm?doid=1357054.1357127.

Kozinets, R. V., Hemetsberger, a. & Schau, H.J., 2008. The Wisdom of Consumer Crowds: Collective Innovation in the Age of Networked Marketing. *Journal of Macromarketing*, 28(4), pp.339–354. Available at: http://jmk.sagepub.com/cgi/doi/10.1177/0276146708325382 [Accessed May 28, 2013].

Nickerson, J., Sakamoto, Y. & Yu, L., 2011. Structures for creativity: The crowdsourcing of design. *… Workshop on Crowdsourcing …*, pp.1–4. Available at: http://crowdresearch.org/chi2011-workshop/papers/nickerson.pdf [Accessed June 22, 2013].

Osborn, A.F., 1957. *Applied imagination; principles and procedures of creative problem-solving*, Published by Scribner's.

Stewart, O., Huerta, J. & Sader, M., 2009. Designing crowdsourcing community for the enterprise. *ACM SIGKDD Workshop on Human Computation*, pp.50–53. Available at: http://portal.acm.org/citation.cfm?doid=1600150.1600168 [Accessed July 17, 2013].

Yochai Benkler, 2006. The wealth of networks: How social production transforms markets and freedom. *Yale University Press*. Available at: http://paigrain.debatpublic.net/docs/1_Symposium_PFIE_6_2_web.pdf [Accessed July 17, 2013].

Yu, L. & Nickerson, J., 2011. Cooks or cobblers?: crowd creativity through combination. *CHI 2011, May 7–12, Vancouver, BC, Canada*. Available at: http://dl.acm.org/citation.cfm?id=1979147 [Accessed June 25, 2013].

Yu, L. & Nickerson, J., Generating creative ideas through crowds: An experimental study of combination. In *Thirty Second International Conference on Information Systems, Shanghai 2011*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2039547 [Accessed June 25, 2013].

Yu, L. & Sakamoto, Y., 2011. Feature selection in crowd creativity. *Foundations of Augmented Cognition. Directing the …*, pp.383–392. Available at: http://link.springer.com/chapter/10.1007/978-3-642-21852-1_45 [Accessed June 25, 2013].

CHAPTER 2

LITERATURE REVIEW

I. INTRODUCTION

Increasingly creative design is seen not as a product of an individual but rather the combined efforts of many people. Although such collaborative design is well documented in the literature for design activities carried out by, say teams of professional engineers and architects (Whitfield et al. 2002) less is known about the potential of distributed, anonymous, crowd-based collaboration in creative tasks. In contrast to the established design processes academic research into crowdsourced design has investigated the power of iteration, competition, reward and combination processes. However, to effectively employ these tools, a crowdsourced design task must embed a process that generates an adequate volume and quality of responses in a feasible time.

1. Crowdsourcing Definition

Commercial design tasks are rarely undertaken by individuals, but rather by groups of people with various skills (i.e., marketing, designers, manufacturers, engineers, purchasing managers, salesmen and after-sale service workers, etc.)(Hsiao & Chou 2004)(Cooper & Kleinschmidt 1986)(Dougherty 1992). This collection of people could be regarded as a form of crowd, so in many commercial enterprises, design and creativity has been seen as attributable to an internal crowd for many years. However, over time, the nature of the "crowd" has been changing.

In the reported research papers, these collaborations of human intelligence were defined as many words as follows: radical decentralization, wisdom of crowds, peer production, open innovation, mass innovation, wikinomics and more (T. W. Malone 2004)(Surowiecki 2004)(Yochai Benkler 2006)(Chesbrough 2003)(Leadbeater & Powell 2009)(Tapscott & Williams 2008). In 2006, "crowdsourcing" was defined by Jeff Howe as:

> "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call" (Howe 2006)

However, these people do not have the same composition as the internal company "crowd" discussed earlier. This new type of "crowd" is composed of anonymous, isolated individuals (Yochai Benkler 2006). In this crowd, members do not know each other; and usually work alone on tasks that, if their results are accepted, are rewarded with, typically, small amounts of money. (Kittur et al. 2008). Crowdsourcing groups include online product communities (Brabham 2009)(Jeppesen & Frederiksen 2006)(Kozinets et al. 2008), virtual communities of special interests (Hogue 2011), the general public (Chilton 2009)(Haklay & Weber 2008), and employees who typically would not participant in the tasks to be completed (Stewart et al. 2009).

II. CROWDSOURCING AS AN INNOVATION METHODOLOGY

1. Crowdsourcing as the Scientific Method

Because for crowdsourcing, open innovation is associated with ideas of the process is often described in a business (or innovation) domain, and more recently as a phenomenon in scientific research. Technological advance is usually subdivided into two different categories (Nelson & Winter 2009):

1) Invention – a scientific breakthrough
2) Innovation – commercialization of the invention

With the booming of Internet which enables communities to connect and collaborate, as well as integrating anonymous people work together, creating a virtual world of collective intelligence (Malone et al. 2010) has been successfully done during the last decades. What is more, it has been reported that for any group of users of a technology, a large number of participants will come up with innovative ideas (von Hippel 2005). And the similar fundamental research practices have been suggested by Irwin in 1995 and Shneiderman in

2008 on “Citizen Science” (Irwin 1995) and “Science 2.0” (Shneiderman 2008) to support this point of view that “what began as process in business is also being observed in science”.

An example is provided in the customer factor where a form of crowdsourcing (i.e., Chesbrough described as open innovation) has validated successful and leaded to a practically fundamental research:

“In 1999, Procter & Gamble decided to change its approach to innovation. The firm extended its internal R&D to the outside world through an initiative called Connect and Develop. This initiative emphasized the need for P&G to reach out to external parties for innovative ideas. The company's rationale is simple: Inside P&G are more than 8,600 scientists advancing the industrial knowledge that enables new P&G offerings; outside are 1.5 million.” (Chesbrough 2003)

Furthermore, in 2000, Schrage pointed out that innovation and creation both require improvisation. It is not about following the rules of the game, but more about rigorously challenging and revising them, which is consistent with criticism of any standardization of the Scientific Method. What is more, an expert scientist (or an expert group) needs to manage (and perhaps improvise) the overall process and aggregate potential input from the online crowd participants (Buecheler et al. 2010). However, the crowd doesn’t necessarily have to be composed of experts.

(Maintained) diversity is an essential advantage of crowds (i.e., crowd workers have a wide variety of nationality, education background, experience, gender, etc. – this will be discussed in details in the section of demographic information of crowd workers, in chapter 8). Scott E. Page has created a theoretical framework to explain why groups often outperform experts. The results of several experiments formed the basis for the “Diversity Trumps Ability” Theorem (Page 2008): Given certain conditions, a random selection of problem solvers outperforms a collection of the best individual expert problem solvers due to its homogeneity. The experts are better than the crowd, but at fewer things.

Friedrich von Hayek stated in 1945 that nearly every individual "has some advantage over all others because he possesses unique information of which beneficial use might be made" (von Hayek 1945). Although some academic institutes have been trending towards a more entrepreneurial model for more than two decades, (Etzkowitz 1983)(Etzkowitz et al. 2000), it is still be regard as being in the not-for-profit field, interested in spreading knowledge throughout society. Crowdsourcing has been successfully used in the business environment for creating economic value. And there is no systematic study investigating the applicability of Crowdsourcing in not-for-profit basic research (as conducted in traditional universities) (Buecheler et al. 2010).

Consequently, based on the case studies' results in several industries, it has been indicated that fundamental research potentially benefits from leveraging collective intelligence techniques (i.e., crowdsourcing) (Buecheler et al. 2010). What is more, Buecheler hypothesized that there are "tasks" in the scientific method that can potentially benefit from crowdsourcing as well. Besides, Buecheler also tested a series of applicability of agent design principles from artificial intelligence research to crowdsourcing.

2. Open Innovation and Crowdsourcing

In 2003, Chesbrough determined two definition which were:

1) **Closed innovation:** it is fundamentally inwardly focused, utilizing internal research and development to develop innovation (Chesbrough 2006);
2) **Open innovation:** valuable ideas can come from inside or outside the organization and can go to market from inside or outside the organization (Chesbrough 2006).

What is more, it is important to know the definition of innovation that is to further understand the open innovation. One definition of innovation is:

> Innovation is: production or adoption, assimilation, and exploitation of a value-added novelty in economic and social spheres; renewal and enlargement of products, services, and markets; development of new methods of production; and establishment of new management systems. It is both a process and an outcome (Wikipedia 2016)

During the innovation process, “social media crowdsourcing alleviates some of these potential issues, as participants have interest in the brand, product, or firm, and they actively choose to contribute and be part of the social community regardless of incentives (i.e. competition rewards, fun and self-satisfaction, etc.) (Füller 2006)(Hippel 2002)”. Since then, the crowdsourcing and such social media have been collaboratively used in the context of open innovation.

From the reported research work crowdsourcing applied in open innovation and idea creation area, it has been investigated that because creativity is facilitated by collective action, crowdsourcing is a tool for knowledge storing and sharing. In the open crowdsourcing process innovations can be undertook by the knowledge “combination and recombination”, and “collective intelligence emerges from the many-to-many interactions” supported by crowdsourcing during open innovation activities (Mount 2014).

The use of commercial crowdsourcing platforms could facilitate an open innovation approach to even small design tasks (such as desk lamp or kitchen layout).

3. Linguistic Study

Some of the earliest systematic research into crowdsourcing performance was reported by linguistic researchers. In this section, the application of crowdsourcing in natural language, or translations – the linguistic research area will be described. Because online crowdsourcing has gained popularity in recent years by it’s cheap and easy programmatic access to human intelligence, “researchers have proposed using crowdsourcing for a diverse set of natural language processing in tasks” (Mitchell et al. 2014). These tasks include paired data collection for training machine translation systems (Zaidan & Callison-burch 2011).

In 2011, Zaidan reported that by a crowdsourcing translation process, the professional quality of translation results can be conducted from the non-professionals (i.e., crowd workers)(Zaidan & Callison-burch 2011). Interesting, Zaidan investigated that a requester (i.e., people who posts task on crowdsourcing platform to solve the problem) can only pay as low as $0.01 to crow workers for each solution for a translation task. What is more, Zaidan defined crowd workers as “anonymous non-professionals” and pointed out the risk of poor translation. However a quality control model and the evaluation strategy were established to solve that risk of poor quality results.

Finally by a crowdsourcing process, Zaidan has demonstrated that high quality translations can be collected from non-professional translators (i.e., crowd workers), and the cost is “an order of magnitude cheaper” than the experts.

In the context of crowdsourced design, the above work is interesting because of the process used to review and merge the translations generated by individual workers.

III. CROWDSOURCING PROCESS

1. Crowdsourcing Platforms Used for Creative Design

Like other internet technologies, the crowd needs online platforms to manage the assignment of tasks and payment of rewards.

Over recent years a variety of different approaches to commercial crowdsourcing have appeared. Several of these have been used to crowdsource creative design and the approaches taken can be seen to fall into several distinct categories:

- Public Design Competition
- Multi-stage Community, Competition Design
- Aggregated Anonymous Crowd Design

The flowing sections review the quality assessment methods and reward models used in these platforms.

2. Assessment Methods & Reward Models

2.1 Public Design Competitions

The use of public competition to generate novel designs from anonymous crowds started long before the Internet (e.g., the Longitude Prize, 1730) and is today exemplified by Chinese websites such as Taskcn or Witkey. Taskcn was established in 2006 and Witkey in 2005 (Anon n.d.). Witkey is an online business model in which employers or task requesters can publish their questions/problems in diverse categories, such as: logo design, room interior decoration, website page design, translation jobs, computer coding, copywriting tasks and even competitive bidding, etc., and then participants (i.e., registered users) can select tasks and provide results/solutions to requesters, after which the winners (and only winners) will receive their payment.

Taskcn is famous for its competition tasks posted by companies or organizations. Usually, Taskcn posts tasks as invitations to tender. Workers who are interested in those tasks are required to upload their proposals as the submission of tender. The payment budget could as low as ¥200 (about $33.06), and as high as ¥20,000 (about $3306.44) (in Figure 2 - 1). Presently, the total value of all tasks on Taskcn is ¥36,524,694 (about $6,038,335.46) (Anon n.d.).

Figure 2 - 1 Example tasks of invitation of tenders on Taskcn

By the end of January of 2014, the total number of registered users in Taskcn was 3,429,790 (Anon n.d.). Malone (T.W. Malone 2004) claimed that the future freelance marketplaces will be like the Witkey crowdsourcing module. On different Witkey websites, participants are required to have expertise in particular subject areas (Jiang Yang et al. 2008). The requirements for workers in the crowd to have a level of skill in a specific task are frequently seen. For example, in Galaxy Zoo (Anon n.d.), only users having a basic knowledge of astronomy are able to undertake the image classification tasks. In practice workers who do not have skills in the required areas will be eliminated from tasks by competitors who do have particular professional skills. That is because most Chinese crowd platforms are competition-based, which means, only winners gain monetary rewards. Professional expertise and keen competitiveness are two obvious characteristics of Taskcn. This reward model leads to a series of participants' behaviours. For one thing, expert workers participate in a great number of tasks because by doing so, they increase their experience and reputation (even if some tasks are unrewarded). But in contrast, less skilled individuals who might seek to optimize their strategies to increase the possibility of winning would learn to attempt the solution of problems in the closing stage of the task. Also these workers learn to select tasks which are not popular among other more experience users, or might pick tasks with higher expectations (J Yang et al. 2008). In comparison to other crowdsourcing sites, usually, there are a large number of tasks with high payment in Taskcn.

2.2 Multi-stage Competitions

Although the "winner take all" approach employed in public competition sites is simple for the requester to administer, it limits the number (and so the diversity) of solutions because workers are aware that their chances of payment are small. To mitigate this effect competitions can be conducted in a number of stages in which participants are guaranteed some levels of reward. A good example of this approach is the "GE Jet Engine Bracket Challenge". As one of the largest engineering design companies in the world, GE (General Electric Company) organized an engineering design competition which involved multiple stages of design specifications (in Figure 2 - 2), named as "GE jet engine bracket challenge".

Figure 2 - 2 Screenshot of GE jet engine bracket challenge

In this engineering design challenge, participants were required to design a bracket which should be structurally efficient and cost effective. Traditionally, for engineers, it is difficult to balance the performance requirements for high strength and stiffness against the need for low size and weight. However new additive manufacturing technology provides the chance for engineers to use novel shapes while not decreasing performance. To realize this opportunity, GE crowdsourced this bracket design task online stipulating that "Participants in this challenge will use additive manufacturing as the basis for optimizing an existing aircraft engine bracket".

What is more, requirements of the design specifications were provided to participants. For example, the material should be Ti-6Al-4V, the service temperature was 75°F, and design tools used by participants could be any CAD software as long as it is STEP, or IGES compliant. Furthermore, designs had to satisfy the requirements of given load conditions were raised (i.e., max static linear load of 8,000 lbs vertical up, max static torsional load of 5,000 lb-in horizontal at intersection of centerline of pin and midpoint between clevis arms).

In terms of the prize which motived participants to solve the design problems, in Phase 1, there was a $1,000 reward for each of the top 10 entries; and in Phase 2, there was an additional award for the top 8 entries. In total, there were 661 entries published online (in Figure 2 - 3).

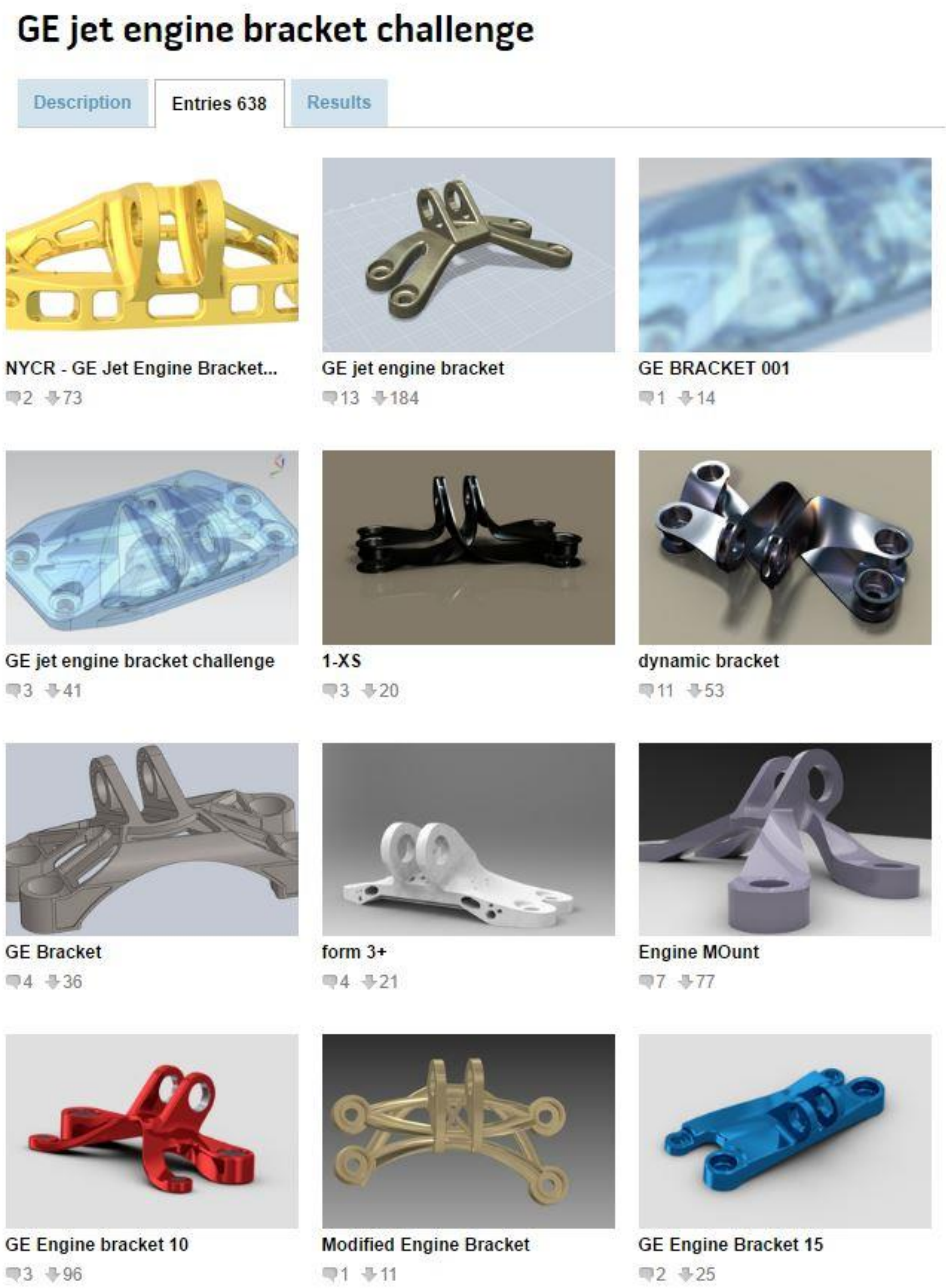


Figure 2 - 3 Examples of submitted bracket

2.3 Integrated Crowdsourced Design

In many ways, the use of Internet crowds to "design by competition" is simply a long establish method updated to exploit cheap global Internet based communication. However, the use of large crowds of anonymous workers whose individual efforts are in some way, combined to create a novel design solution is new. The platform most commonly employed to support this approach is exemplified by Amazon Mechanical Turk (MTurk).

MTurk was established in 2005 for online crowds who would like to earn incomes by solving problems and for client who hope to obtain solutions for their problems. The former called workers, and the latter requesters. Today, MTurk has become one of the most significant crowdsourcing platforms not only for business use, but for experimental research studies. The basic process of how MTurk functions can be summarized as follows:

1. Requesters publish their problems as tasks in MTurk (build up HITs-Human Intelligence Tasks);
    a. People who have registered in MTurk can create HITs describing their main problems.
    b. To avoid misunderstanding, the amount of the rewards (monetary payment/bonus), the task expiry date, how much time for doing the task once started and the detailed task contents (, etc.) should all be given to MTurk workers detailed as clearly as possible in the first instant.

2. Workers search for tasks in which they are interested, or where the payment of tasks is attractive;
    a. All HITs are listed. There are more than 5100 tasks that can be found by MTurk workers.
    b. When selecting HITs, participants could search HITs by entering the least amount of payment they are willing to accept.

3. Before undertaking tasks, it is often necessary for workers to pass "qualification tests", after which workers can scan the whole contents and then try to solve problems;
    a. In some tasks, a "qualification test" (honesty validation) is compulsory for workers to avoid automated or manual "cheating" (e.g., attempting to get paid for no work).

b. Before beginning to do tasks, workers have to press the “accept the HITs” label, so that the timer could start running. Workers must accept HITs first and then figure problems out within the time limit. Otherwise workers’ actions would be treated as invalided.

4. After completing problems successfully, workers submit their solutions. These solutions could be submitted to Requesters directly, or by some other means (i.e., sending by email or sharing by public websites).

5. Requesters validate results, and then only the approved solutions would be paid.

6. Monetary rewards are paid to workers if the workers’ solutions are excellent, then Requesters could pay additional rewards as a bonus.

7. Workers gain payment, and Requesters solve problems.

Some advantages of MTurk was also investigated as follows (Zaidan & Callison-burch 2011):

1. Zero overhead for hiring workers
2. A large, low-cost labour force
3. Short turnaround time, as tasks get completed in parallel by many individuals
4. Access to foreign markets with native speakers of many rare languages

Although the MTurk was used here as a platform for the experiments, there are many other similar services that would work equally well (e.g., ShortTask, Taskcn, etc.). One of the most impressive examples of collaborative, crowdsourced design is the Human-based Genetic Algorithms (HBGA) method that has been used to enable innovation tasks on MTurk (Yu & Nickerson 2011)(Yu & Nickerson n.d.)(Yu & Sakamoto 2011). Based on the research which showed that combinations could develop creativity (Osborn 1957)(Amabile n.d.), a combination process inspired by genetic algorithms was established (Yu & Nickerson n.d.)(Yu & Nickerson 2011)(Bao et al. 2011). Creative design comes from combinations (Amabile n.d.). In the HGBA, new ideas are basically separated into different “generations” – design/evaluate results after each task (i.e., results from the 1st task is the generation 1 and results from the 2nd task is the generation 2). In the first generation, participants from the crowdsourcing platform create the first group of designs. Then a second crowd evaluates the first generation and chooses several pairs for the combination process to construct the second generation (i.e., generation 1 evaluation). In generation 2, some of the ideas were selected directly from the top ranked generation 1 designs, and others were collected by combining pairs chosen from the first generation (i.e., generation 2 combination). Then, the third generation applies the evaluation process to the second generation combination process again to create generation 3 (Yu & Nickerson n.d.)(Yu & Nickerson 2011). So, iteratively, generation after generation, new ideas could be sequentially created.

When a competition of individuals is used to crowdsourced design, there appears little doubt that a high reward results in a higher quality of entry, but in the combination style tasks (i.e., HBGA) the relationship is not clear.

IV. HUMAN-BASED GENETIC ALGORITHMS

Unlike the competition model system (i.e., Taskcn (Anon n.d.; Wu et al. 2014)) where the design work is ultimately done by individuals, collaborative design requires the merging or selective combination of ideas (Jiang Yang et al. 2008; Liu & Yang 2011). One of the most impressive methodologies to emerge for collaborative, crowdsourced design is the Human-based Genetic Algorithms (HBGA) method that has been used for generative innovation tasks (Yu & Nickerson 2011; Yu & Nickerson n.d.; Yu & Sakamoto 2011) (the basic process is shown in Figure 2 - 4). The approach uses selective combinations to develop creativity (Osborn 1957; Amabile n.d.), and has been applied to a number of different applications (Yu & Nickerson n.d.; Yu & Nickerson 2011; Bao et al. 2011). This is a theoretically appealing approach because it has been suggested by some researchers that creative design comes from combinations (Amabile n.d.). In the HGBA, new ideas are basically separated into different generations. In the first generation, participants from the crowdsourcing platform create the first group of designs. Then a second crowd evaluates the first generation and chooses several pairs for the combination process to construct the second generation (i.e., generation 1 evaluation). In generation 2, some of the ideas were selected directly from the top ranked generation 1 designs, and others were collected by combining pairs chosen from the first generation (i.e., generation 2 combination). Then, the third generation applies the evaluation process to the second generation combination process again to create generation 3 (Yu & Nickerson n.d.; Yu & Nickerson 2011). So, iteratively, generation after generation, new ideas could be sequentially created.

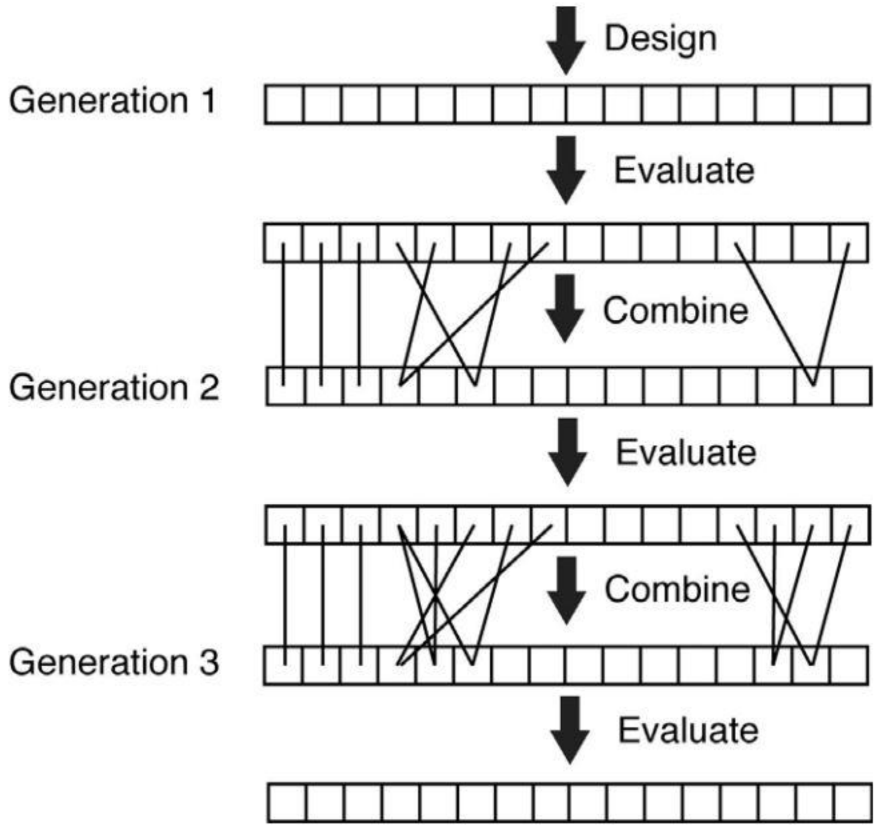


Figure 2 - 4 Yu's HBGA process

The HBGA and competition model (i.e., Taskcn) methodologies are clearly effective, for the design domains they have been applied to free hand sketches or 2D layouts. The objectives of this paper are: 1, to investigate if HBGA can support 3D layout design; 2, to investigate the applicability of the crowdsourced design (cDesign) framework to 3D design tasks. The hypothesis of the result is that the 3D design can be crowdsourced, and the cDesign framework is appropriate to both 2D and 3D design tasks. The next section presents the general cDesign Framework as well as the specifics of its application in 3D design methods.

V. DESIGN METHODOLOGY

In this research, because one of the most important purposes is to improve the design quality (by crowdsourcing technique), the design methodologies need to be reviewed as well. Design methodology was defined as: a set course of action for the design of technical systems that derives its knowledge from design science, cognitive psychology and from practical experience. It contains plans of action that work linking steps, strategies, rules and principles to achieve goals and methods to solve individual design tasks (Ulrich & Eppinger 2011).

This procedure should finally provide a “framework” to guide designers to encourage creativities (Pugh 1991)(Ulrich & Eppinger 2011). And this framework can be widely used in different categories of design activities. Here are many methodologies exist with varying process used in design research area as well as industry. In this section, the following design methodologies will be described as the classical and basic design model.

1. Total Design (Stuart Pugh)

Total design (Pugh 1991) was defined by Stuart Pugh, and widely used in the design education. Total Design splits the process into a central design of core of activities as shown in Figure 2 - 5. The processes start from the market analysis, then a product design specification (PDS) (Pugh 1991) is used to guide designers conducting the next steps (in Figure 2 - 6).

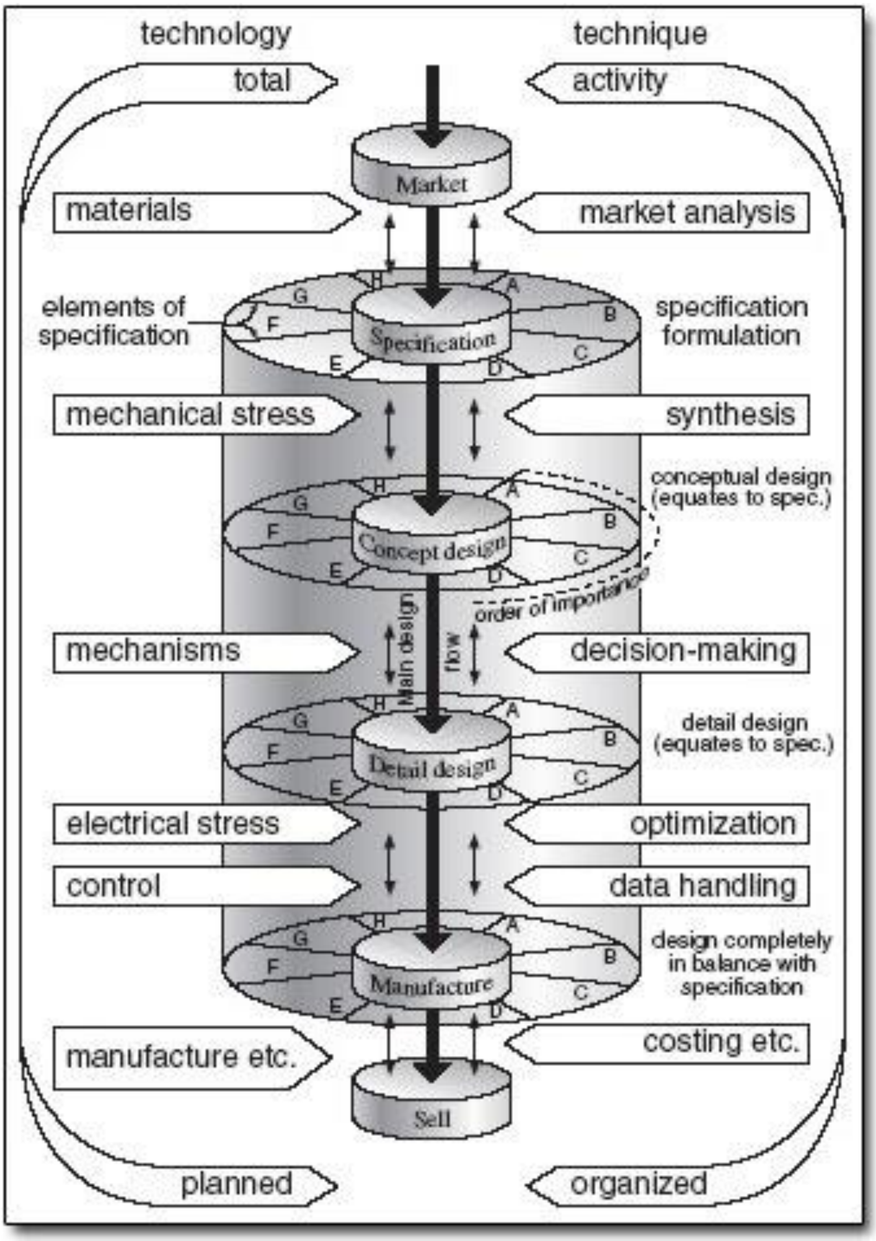


Figure 2 - 5 Total Design model and process

Total Design is an iterative process as shown in Figure 2 - 5, and techniques are used to enable the designer to operate the core activity and carry out the designs.

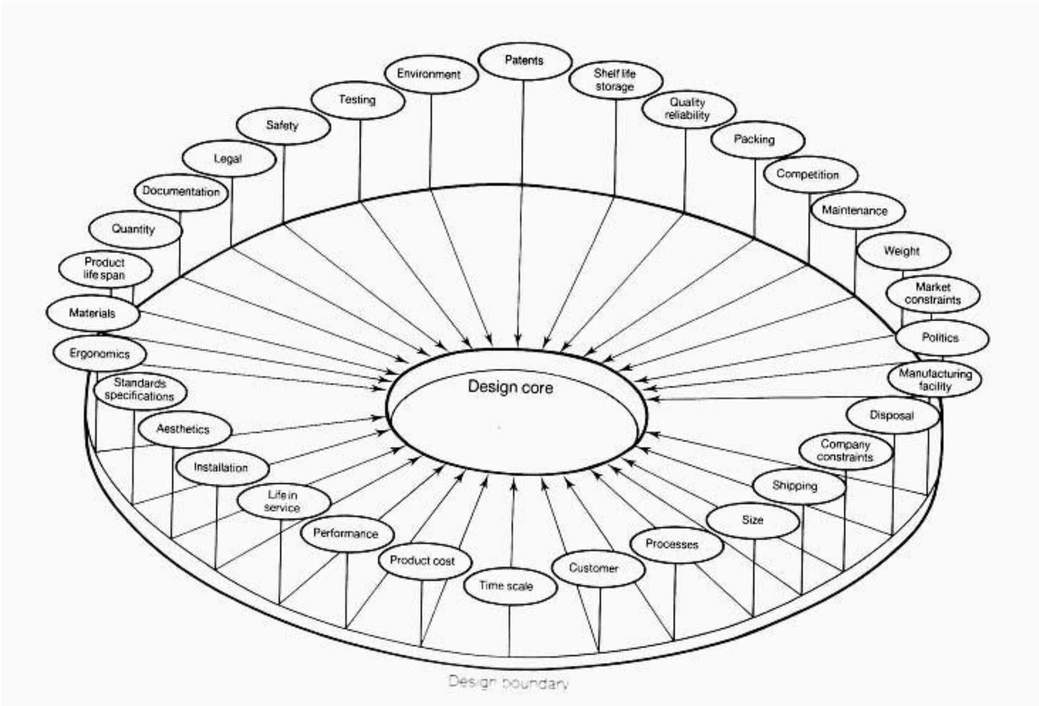


Figure 2 - 6 Product design specification (PDS)

2. Ulrich & Eppinger

Ulrich & Eppinger (Ulrich & Eppinger 2011) is a design technique[1] applied in design education as well. Figure 2 - 7 shows that the design process is structured into six stages: planning, concept development, system-level design, detail design, testing and refinement, and production ramp-up.

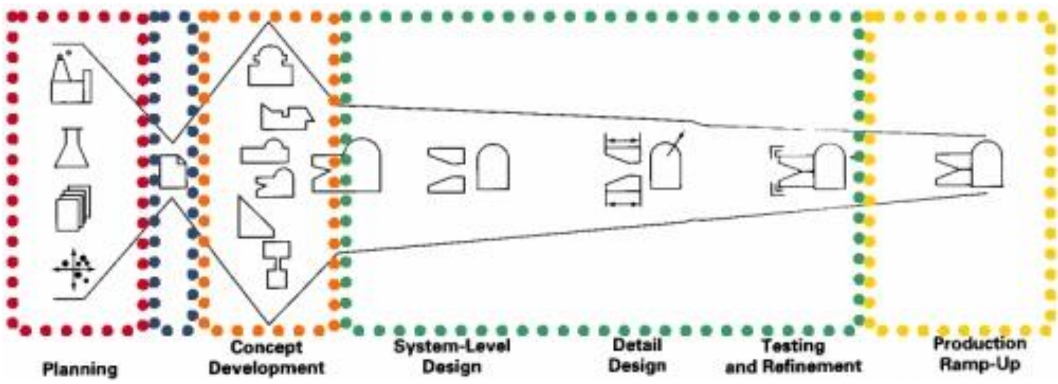


Figure 2 - 7 Ulrich & Eppinger design process

[1] http://www.ulrich-eppinger.net/, 11/2016

The process from the planning of the design task considering such as, corporate strategy, assessment of technological developments and market objectives. Then the concepts are developed. Following this the process is narrow down to the system level where the detailed definition of the product and components are considered till the testing and refinement (i.e., prototype). The final phase is the production ramp-up where product is made.

3. Wideblue: design process

Wideblue[2] is a medical device design company that has a structured approach to design (in Figure 2 - 8). During the design process, the final design requirements are considered at the beginning of the design process. This approach provides strong control with is vital when approaching regulatory compliance and what wide blue feel that "maximises the probability of technical and commercial success"[3].

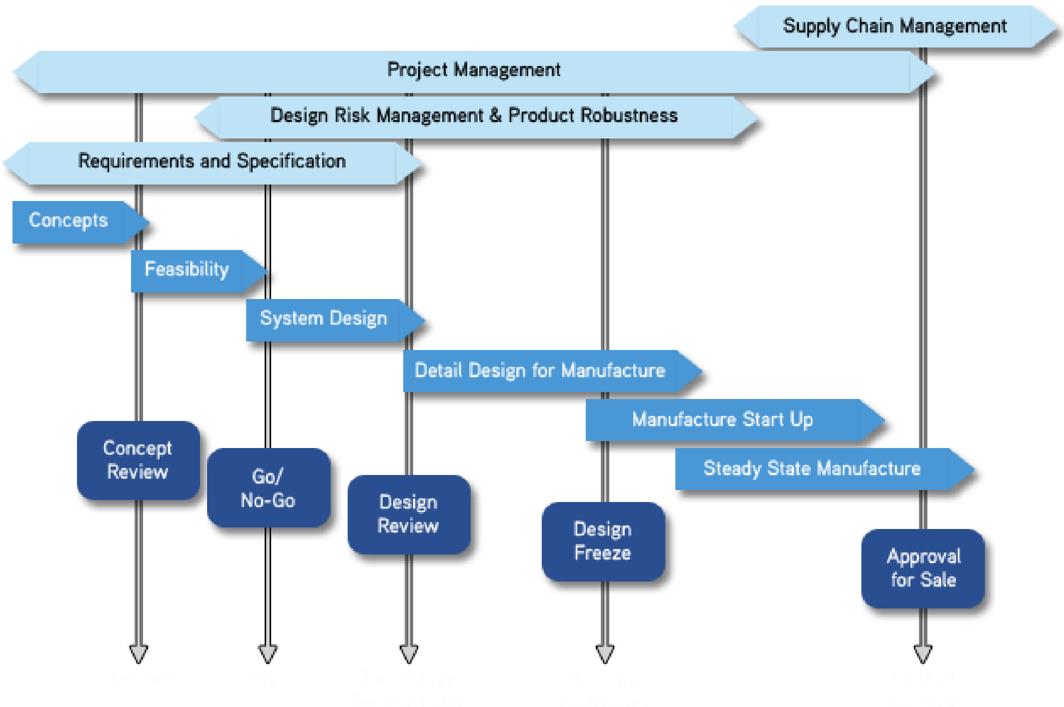


Figure 2 - 8 Wideblue design process

[2] http://www.wide-blue.com/, 11/2016

[3] http://www.wide-blue.com/resources.php, 11/2016

In this section, three different product design methodologies are discussed. In the next section, the knowledge gap of this research will be described.

VI. KNOWLEDGE GAPS

From the review of crowdsourcing, crowdsourced design and the process adopted, it can be observed that although the HBGA and competition methodologies are clearly effective, they are only components of the overall design process. At a high level the creators of crowdsourced design tasks must select the "tools" (i.e., components) they are going to employ and define the parameters (e.g., crowd size, payment level) and workflows needed to implement the task online. The components used to implement the crowdsourced design tasks reported by different researchers are illustrated in Table 2 - 1 and Table 2 - 2. Although the authors' selection of parameter values (e.g., payment level) and components (e.g. iterative design) effectively enables their investigations, none of the papers explicitly enumerate the choices available or the rational for final selection. The lack of an explicit process design model for crowdsourced design creates a barrier to the wide spread adoption of the method. To address this gap, it is proposed to define a framework that explicitly defines the major steps in the creation of a crowd-based design task.

Table 2 - 1 Reported components of crowdsourced design tasks - 1

	Design Generation	**Design Process**		**Evaluation Method**
		Non-Iterative	Iterative	
Lixiu Yu (Yu & Nickerson 2011)	Human-based		√	Quantitative
Amit Banerjee (Banerjee et al. 2008)	Computational-based		√	Quantitative
Chunyan Xu (Xu et al. 2012)	Human-based	√		Quantitative
Lingyun Sun (Sun et al. 2014)	Human-based		√	Quantitative
Kurt Luther (Luther et al. 2015)	Human-based	√		Quantitative
Kazjon Grace (Grace et al. 2014)	Human-based		√	Quantitative & Qualitative

Table 2 - 2 Reported components of crowdsourced design tasks - 2

	Evaluation Criteria	Payment Level	Cheating Strategy	Task Design Rational
Lixiu Yu (Yu & Nickerson 2011)	Set by Requester	Fixed	no	no
Amit Banerjee (Banerjee et al. 2008)	Set by Requester	Fixed	no	no
Chunyan Xu (Xu et al. 2012)	Set by Requester	Fixed	no	no
Lingyun Sun (Sun et al. 2014)	Set by Requester	Fixed	no	no
Kurt Luther (Luther et al. 2015)	Set by Requester	Fixed	no	no
Kazjon Grace (Grace et al. 2014)	Set by Requester	Fixed	no	no

VII. CHAPTER SUMMARY

In this chapter, the development of crowdsourcing and the crowdsourced design is described. And the research gap has been discussed. It was investigated that at present, there is no crowdsourced design mode that is generic. Based on the Introduction and Literature Review chapters, the final aim of this research is to establish a complete and effective general crowdsourced design framework to design tasks. But before the experiments in crowdsourced design, in the next chapter, the methodologies and research design will be discussed. These will provide the scientific fundamental thoughts to design the research.

References

Amabile, T., *Creativity in Context: Update to the Social Psychology of Creativity*, Westview Press, June 7, 1996.

Anon, Galaxy Zoo. Available at: http://www.galaxyzoo.org/ [Accessed January 20, 2014a].

Anon, TaskChina. Available at: http://weike.taskcn.com/.

Anon, Witkey.com. Available at: http://www.witkey.com/ [Accessed January 20, 2014c].

Banerjee, A., Quiroz, J. & Louis, S., 2008. A model of creative design using collaborative interactive genetic algorithms. *Design Computing and Cognition'08*, pp.397–416.

Available at: http://link.springer.com/chapter/10.1007/978-1-4020-8728-8_21 [Accessed June 25, 2013].

Bao, J., Sakamoto, Y. & Nickerson, J., 2011. Evaluating Design Solutions Using Crowds. In *Seventeenth Americas Conference on Information Systems, August 4th-7th, Detroit, Michigan*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2201651 [Accessed June 25, 2013].

Brabham, D., 2009. Crowdsourced advertising: how we outperform Madison Avenue. *Flow: A Critical Forum on Television and Media Culture*. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Crowdsourced+Advertising:+How+We+Outperform+Madison+Avenue#0 [Accessed July 17, 2013].

Buecheler, T. et al., 2010. Crowdsourcing , Open Innovation and Collective Intelligence in the Scientific Method : A Research Agenda and Operational Framework Why Crowdsourcing in the Scientific Method. In *The Alife XII Conference*. Odense, Denmark, pp. 679–686.

Chesbrough, H.W., 2003. *Open innovation. The new imperative for creating and profiting from technology*, Boston: Harvard Business School Press.

Chesbrough, H.W., 2006. *Open Innovation: The New Imperative for Creating and Profiting from Technology*, Harvard Business Press.

Chilton, S., 2009. Crowdsourcing is radically changing the geodata landscape: Case study of OpenStreetMap. *24th International Cartographic Conference*. Available at: http://raptor1.bizlab.mtsu.edu/S-Drive/DMORRELL/Mgmt 4990/Crowdsourcing/22_6.pdf [Accessed July 17, 2013].

Cooper, R.G. & Kleinschmidt, E.J., 1986. An Investigation into the New Product Process: Steps, Deficiencies, and Impact. *Journal of Product Innovation Management*, 3(2), pp.71–85. Available at: http://doi.wiley.com/10.1111/1540-5885.320071.

Dougherty, D., 1992. Interpretive Barriers To Successful Product Innovation In Large Firms. *Organization Science*, 3.

Etzkowitz, H., 1983. Entrepreneurial scientists and entrepreneurial universities in American academic science. *Minerva*, 21(No. 2/3), pp.198–233. Available at: http://www.jstor.org/stable/41820527.

Etzkowitz, H. et al., 2000. The future of the university and the university of the future : evolution of ivory tower to entrepreneurial paradigm. *Research Policy*, 29(2), pp.313–

330. Available at: http://www.sciencedirect.com/science/article/pii/S0048733399000694.

Füller, J., 2006. Why Consumers Engage in Virtual New Product Developments Initiated by Producers. *Advances in Consumer Research*, 33, pp.639–646.

Grace, K. et al., 2014. A Process Model for Crowdsourcing Design: A Case Study in Citizen Science. In *Design Computing and Cognition ’14*. Springer International Publishing Switzerland 2015, p. 18.

Haklay, M. & Weber, P., 2008. Openstreetmap: User-generated street maps. *Pervasive Computing, IEEE*, pp.12–18. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4653466 [Accessed July 17, 2013].

von Hayek, F.A., 1945. *The Use of Knowledge in Society*,

von Hippel, E., 2005. *Democratizing innovation*, MIT Press.

Hippel, E. Von, 2002. Open Source Software and the “Private-Collective” Innovation Model: Issues for Organization Science. *Organization Science*, 14(2003), pp.208–223.

Hogue, C., 2011. Crowdsourcing for science. *Chemical & Engineering News*. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:CROWDSOURCING+FOR+SCIENCE#7 [Accessed July 17, 2013].

Howe, J., 2006. The rise of crowdsourcing. *Wired magazine*, (14), pp.1–5. Available at: http://sistemas-humano-computacionais.wikidot.com/local--files/capitulo:redes-sociais/Howe_The_Rise_of_Crowdsourcing.pdf [Accessed June 22, 2013].

Hsiao, S.-W. & Chou, J.-R., 2004. A creativity-based design process for innovative product design. *International Journal of Industrial Ergonomics*, 34(5), pp.421–443. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0169814104000873 [Accessed May 29, 2013].

Irwin, A., 1995. *Citizen Science: A Study of People, Expertise and Sustainable Development*, Psychology Press.

Jeppesen, L.B. & Frederiksen, L., 2006. Why Do Users Contribute to Firm-Hosted User Communities? The Case of Computer-Controlled Music Instruments. *Organization Science*, 17(1), pp.45–63. Available at: http://orgsci.journal.informs.org/cgi/doi/10.1287/orsc.1050.0156 [Accessed May 22, 2013].

Kittur, A., Chi, E.H. & Suh, B., 2008. Crowdsourcing user studies with Mechanical Turk. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, p.453. Available at: http://portal.acm.org/citation.cfm?doid=1357054.1357127.

Kozinets, R. V., Hemetsberger, a. & Schau, H.J., 2008. The Wisdom of Consumer Crowds: Collective Innovation in the Age of Networked Marketing. *Journal of Macromarketing*, 28(4), pp.339–354. Available at: http://jmk.sagepub.com/cgi/doi/10.1177/0276146708325382 [Accessed May 28, 2013].

Leadbeater, C. & Powell, D., 2009. *We-think. mass innovation, not mass production*, London: Profile.

Liu, T. & Yang, J., 2011. Crowdsourcing with all-pay auctions: A field experiment on Taskcn. *ASIST 2011, October 9-13, New Orleans, LA, USA*. Available at: http://onlinelibrary.wiley.com/doi/10.1002/meet.2011.14504801298/full [Accessed June 22, 2013].

Luther, K. et al., 2015. Structuring, Aggregating, and Evaluating Crowdsourced Design Critique. In *CSCW 2015, March 14–18, 2015, Vancouver, BC, Canada*. ACM, p. 13.

Malone, T.W., 2004. *The future of work. How the new order of business will shape your organization, your management style, and your life.*, Boston: Harvard Business School Press.

Malone, T.W., 2004. *The Future of Work: How the New Order of Business Will Shape Your Organization, Your Management Style, and Your Life*, Harvard Business Press.

Malone, T.W., Laubacher, R. & Dellarocas, C., 2010. Harnessing Crowds : Mapping the Genome of Collective Intelligence. *MIT Sloan Research Paper No. 4732-09*.

Mitchell, M. et al., 2014. Crowdsourcing Language Generation Templates for Dialogue Systems. In *Proceedings of the INLG and SIGDIAL 2014 Joint Session*. pp. 172–180. Available at: http://www.anthology.aclweb.org/W/W14/W14-44.pdf#page=180.

Mount, M., 2014. Social Media: A Tool for Open Innovation. *California Management Review*, 56(4), pp.124–143.

Nelson, R.R. & Winter, S.G., 2009. *An Evolutionary Theory of Economic Change*, Harvard University Press.

Osborn, A.F., 1957. *Applied imagination; principles and procedures of creative problem-solving*, Published by Scribner's.

Page, S.E., 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton Univ. Press.

Pugh, S., 1991. *Total Design: integrated methods for successful product engineering*, Addison-Wesley Pub (Sd).

Shneiderman, B., 2008. Copernican challenges face those who suggest that collaboration, not computation are the driving energy for socio-technical systems that characterize Web 2.0. *Science*, 319(1349–1350). Available at: http://www.umiacs.umd.edu/publications/copernican-challenges-face-those-who-suggest-collaboration-not-computation-are-driving.

Stewart, O., Huerta, J. & Sader, M., 2009. Designing crowdsourcing community for the enterprise. *ACM SIGKDD Workshop on Human Computation*, pp.50–53. Available at: http://portal.acm.org/citation.cfm?doid=1600150.1600168 [Accessed July 17, 2013].

Sun, L. et al., 2014. Collaborative sketching in crowdsourcing design: a new method for idea generation. In *Int J Technol Des Educ (2015) 25:409–427*. Springer Science+Business Media Dordrecht 2014, p. 19.

Surowiecki, J., 2004. *The wisdom of crowds. Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations* 1st ed., New York: Doubleday.

Tapscott, D. & Williams, A.D., 2008. *Wikinomics. How mass collaboration changes everything*, New York: Portfolio.

Ulrich, K.T. & Eppinger, S.D., 2011. *Product Design and Development: A Resource for Students and Professionals in the Field of Product Design and Development*, McGraw-Hill Education.

Whitfield, R. et al., 2002. Distributed design coordination. *Research in Engineering Design*, 13, pp.243–252. Available at: http://link.springer.com/article/10.1007/s00163-002-0021-1 [Accessed October 14, 2013].

Wikipedia, 2016. Innovation. Available at: https://en.wikipedia.org/wiki/Innovation [Accessed November 7, 2016].

Wu, H., Corney, J. & Grant, M., 2014. Crowdsourcing Measures Of Design Quality. In *34th Computers and Information in Engineering Conference (CIE)*.

Xu, C., Qin, S. & Xiao, Z., 2012. Crowdsourcing based product innovation design service model for small-and medium-sized enterprises. *Automation and Computing (ICAC), 2012 …*, (September). Available at:

http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6330506 [Accessed July 17, 2013].

Yang, J., Adamic, L. & Ackerman, M., 2008. Competing to Share Expertise: The Taskcn Knowledge Sharing Community. In *ICWSM*. Available at: http://www.aaai.org/Papers/ICWSM/2008/ICWSM08-027.pdf [Accessed July 26, 2013].

Yang, J., Adamic, L. & Ackerman, M., 2008. Crowdsourcing and knowledge sharing: strategic user behavior on taskcn. *EC'08, July 8–12, Chicago, Illinois, USA.* Available at: http://dl.acm.org/citation.cfm?id=1386829 [Accessed July 19, 2013].

Yochai Benkler, 2006. The wealth of networks: How social production transforms markets and freedom. *Yale University Press*. Available at: http://paigrain.debatpublic.net/docs/1_Symposium_PFIE_6_2_web.pdf [Accessed July 17, 2013].

Yu, L. & Nickerson, J., 2011. Cooks or cobblers?: crowd creativity through combination. *CHI 2011, May 7–12, Vancouver, BC, Canada*. Available at: http://dl.acm.org/citation.cfm?id=1979147 [Accessed June 25, 2013].

Yu, L. & Nickerson, J., Generating creative ideas through crowds: An experimental study of combination. In *Thirty Second International Conference on Information Systems, Shanghai 2011*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2039547 [Accessed June 25, 2013].

Yu, L. & Sakamoto, Y., 2011. Feature selection in crowd creativity. *Foundations of Augmented Cognition. Directing the …*, pp.383–392. Available at: http://link.springer.com/chapter/10.1007/978-3-642-21852-1_45 [Accessed June 25, 2013].

Zaidan, O.F. & Callison-burch, C., 2011. Crowdsourcing Translation: Professional Quality from Non-Professionals. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, pp. 1220–1229. Available at: http://delivery.acm.org/10.1145/2010000/2002626/p1220-zaidan.pdf?ip=130.159.247.113&id=2002626&acc=OPEN&key=C2D842D97AC95F7A.78A0E221B184F35C.4D4702B0C3E38B35.6D218144511F3437&CFID=691357193&CFTOKEN=37735314&__acm__=1478536727_917ab1ef325b1c2acc6ff25e2e0985f1.

CHAPTER 3

RESEARCH DESIGN &

METHODOLOGY

I. INTRODUCTION

Research Design is defined by Yin[4] as:

"the logical sequence that connects the empirical data to a study's initial research questions and, ultimately, to its conclusions, involving an action plan to getting from the initial research question to conclusions."

To discuss how to manage the research design, choose methodologies and techniques, (appropriate for the research aims, objectives and research questions), this chapter is structured as follows. Firstly, the research paths created by Beech is presented and adopted. Then based on this model, the author briefly describes a number of different research paradigm, before determining the research methodology to be used by this investigation. This methodology works throughout the experiments, and guides each stage of the research, and finally leads to the contribution of the knowledge.

II. RESEARCH PARADIGM

In 2005, Beech presented a theoretical framework of research design which provided a taxonomy of activities ranging from ontology (at the highest level) and progressing down to techniques, the bottom. between these highest and lowest levels, there are two middle levels: Epistemology and Methodology (in Figure 3 - 1).

---

[4] Robert K. Yin, 'Case Study Research: Design and Methods', Sage Publications, 18/03/1994 – 170 pages.

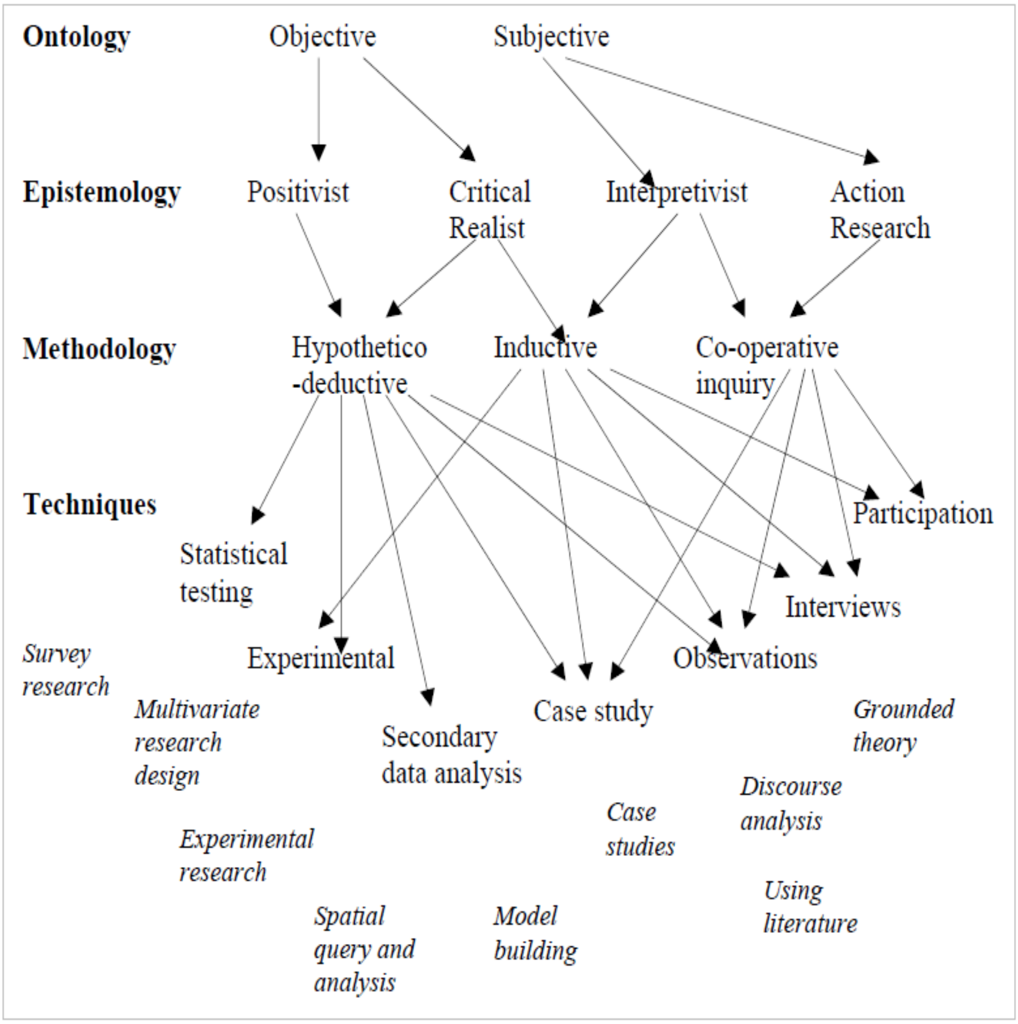


Figure 3 - 1 Beech's research design map (2005)[5]

Because the ontological issue at the highest level addresses the nature of reality, researchers generally fall into two broad groups: firstly, the objective (i.e., quantitative research), and secondly, the subjective (i.e., qualitative research). The following figure shows the choice of research methods related to ontology (in Figure 3 - 2).

[5] N. Beech, 2005. Research Methodology Course Notes: Strathclyde Business School.

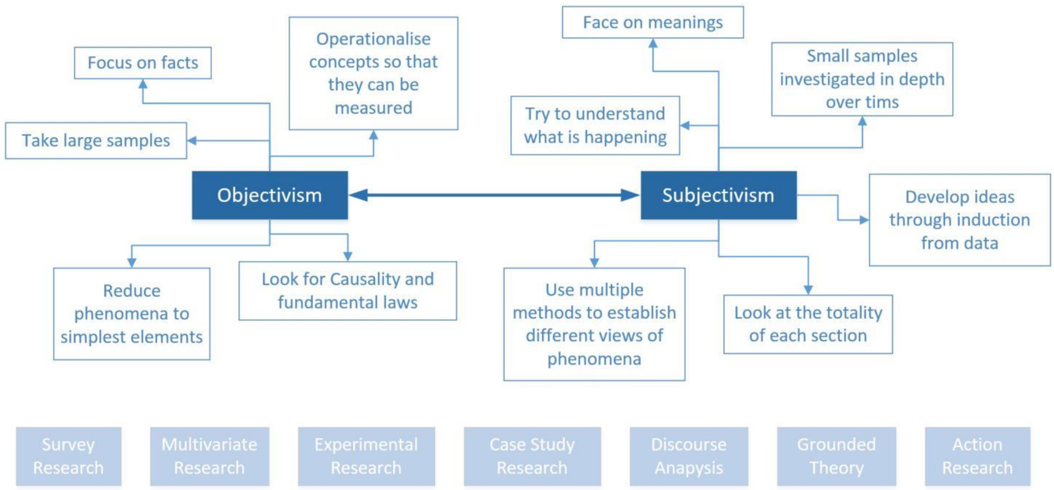


Figure 3 - 2 The choice of research methods related to ontology, Beech (2005)[6]

Choosing the methodologies and techniques for the research requires the details of the research questions to be considered. In Yin's research design structure, the following four research questions need to be answered. Firstly, what question on study? Secondly, what data are relevant? Thirdly, what data should be collected? Fourthly, how to analyse the results? The next section describes the answers and the solving methodologies to those questions.

[6] N. Beech, 2005. Research Methodology Course Notes: Strathclyde Business School.

III. RESEARCH QUESTIONS & APPROPRIATE METHODOLOGIES

1. What Question to Study?

Although the literature reports some methods for investigating crowdsource design for specific tasks, several limitations are obvious. For example, because on commercial crowdsourcing platforms, the first priority after designing a task is to set the level of payment to crowd workers. It is still unknown that whether rewarding the crowd with higher payment improves the design quality (or conversely low payment can decrease the quality). Moreover, the limits of the approach and supporting technologies are not clear, for example, some 2D design tasks have been crowdsourced, but will the same approach work for 3D designs, and is the feasibility determine by the representation (i.e., sketches or 3D models)? Furthermore, different researchers reported different crowdsourced methods to create and optimise designs, which suggests that the fundamental and general process of creating crowdsourced design tasks has not been established. This contrasts with other areas of engineer design were explicit methodologies have been reported. For instance, in engineering design area, Pugh's Total Design model (in chapter 2) shows the basic stages for successful execution of a general engineering design project (Pugh 1991). However, in the area of crowdsourced design, such model (i.e., crowdsourced design framework) has not been reported.

These observations identified the following questions:

1) What is the relationship between the payment to crowd workers and design quality?

2) Can 3D designs be crowdsourced as well as 2D designs using similar processes?

3) After investigating the above questions and summarising the research results, can a general crowdsourced design framework be defined, and, if so, what are the specific stages in the framework?

4) Can this framework be applied in other different design task, and can the framework improve the design quality?

Given these research questions, the key elements of the research were identified. Regarding the most fundamental observation is that the questions concern a form of design so any investigation will include:

1) design data
2) design data collection (including a design process)
3) design data analysis

In other words, in a design research activity, the design process and the resulting data, (its collection and analysis) are all key elements.

Importantly, the nature of the crowd, the human (i.e., the online crowd) also needs to be understood. Because all designs in crowdsourced design tasks are done by participants, the crowd (i.e., participants in the tasks) directly influences the design quality in a crucial way. Consequently, it is important to control and organise the crowd to produce higher quality. In the meantime, the nature of the crowd behaviour needs to be measured, recorded and analysed. For example, what gender are the workers, and how should cheating behaviour be avoided?

2. What Data are Relevant and What Data to Collect?

To answer these two questions, the first thing is to understand exactly what data is relevant to the research. Given that both design data and online crowd data are relevant to the research. The design data includes design drawings, design concepts layouts (in both 2D and 3D), design specifications, design evaluation criteria, design evaluation results, etc. In terms of the online crowd, for example, considering the variations in performance of the crowd participating in design tasks on crowdsourcing platforms, it is important to characterize the individual participants who form the crowd. So, to investigate the form of crowd, participants' individual information needs to collected (i.e., age, gender, nationality, design experience, education background, etc.).

Given the above, in this research, the following data will be initially considered to represent a design process and so will be collected. In the first place, the design data (i.e., sketches, drawings, layouts, 3D models, 3D layouts, etc.), plus design specifications (i.e., a detailed document providing information about the characteristics of a project to set criteria the developers will need to meet[7]), and any design evaluation criteria (i.e., design quality judging measurements to participants). What is more, the design data will be quantitatively transformed (by a process of relative ranking and voting) into numerical data which will be used to judge design quality by statistical methods.

3. How to Analyse the Results?

Statistical methods will be used to measure the evaluation of design quality. Because in this research, designs will be evaluated by an online crowd, and the design quality will be characterized as numerical data. The numerical design evaluation data will facilitate the measurement of design quality.

Consequently, considering the above discussions on research questions, in terms of the nature of the study and the chosen methodology at the ontology level, both objective and subjective aspects are included. Table 3 - 1 shows the research methodologies and techniques selected for this investigation. The key decisions are:

1) Design will be evaluated in a quantitative way (i.e., statistical analysis of design evaluation data).
2) Because design tasks as well as design evaluation tasks will be done by an online crowd, the management and analysis of the crowd's responses use qualitative method (i.e., coding method to collect design specifications and evaluation criteria).

Because it combines both quantitative and qualitative research method, the methodology of this research can be classified as 'Multiple Experimental Case Study'.

[7] https://en.wikipedia.org/wiki/Design_specification, version on 16/06/2016

IV. RESEARCH VALIDATION

Finally, the research results need to be validated. As reported by Yin, the key criteria to judge the quality of the research design are as follows:

1) Construct Validity
2) Internal Validity
3) External Validity

In this section, these criteria are further discussed, and their application in this research are explained.

1. Construct Validity

It has been reported that the phenomena of the focused elements in the research need to be first defined, and then used to identify the operational measures that match these elements. For instance, a focus of this crowdsourced design research is how to improve the design quality generated by a crowdsourced design process on the Internet (i.e., the crowdsourced platform, the payment to crowd workers, Human-based Genetic Algorithm, etc.). And the operational measures applied to investigate them are done through structuring the crowdsourced design framework.

2. Internal Validity

Internal Validity was defined as 'predominantly a concern only for explanatory type case studies where the researcher is striving to identify causal relationships'. In terms of the framework in this research, each stage and the detailed element need to be explained – what is it? And why is it done? For example, what is the reward for crowd workers, how to set the payment level to them, and why (i.e., why pay higher or lower?)?

3. External Validity

'A theory must first be tested by replicating the findings in a separate case' (Yin). External Validity is to ensure the generalizability of the research findings, which requires to measure the effectiveness of the findings across other cases. So, in this research, an experimental case study design will be used to test the generality of the crowdsourced design framework.

Table 3 - 1 Research methodology and technique

Crowdsourced Design	Technique
Crowdsourced Design Specification	Subjective – Qualitative – Coding
Crowdsourced Design Evaluation Criteria	Subjective – Qualitative – Coding
Design Data Analysis	Objective – Quantitative – Statistical Analysis
Crowd	
Management of Crowd Workers (i.e., cheats avoiding)	Subjective – Observation & Participation
Payment to Workers	Objective – Quantitative – Statistical Analysis

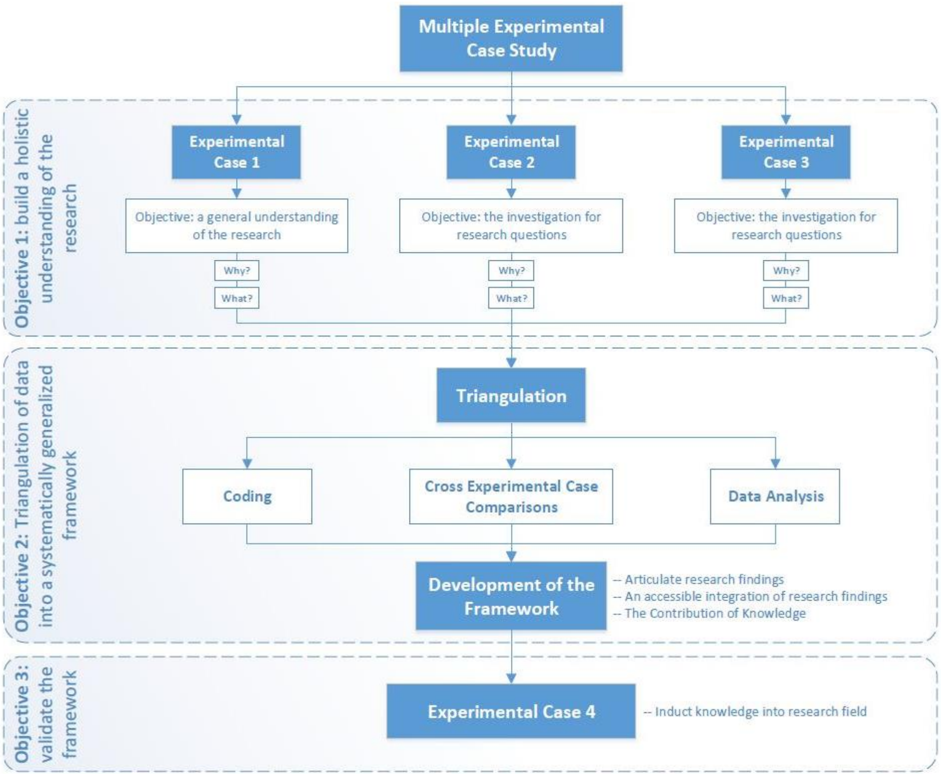


Figure 3 - 3 Research methodology illustration

V. CHAPTER SUMMARY

In this chapter, the research design and its associated methodology/technique are discussed. It is concluded that three crowdsourced design experimental case studies will be done to build a holistic understanding of the process. Then using this experiment, a systematically generalized framework will be created. Finally, this framework will be validated by an online design experimental case study (shown as Figure 3 - 3). The following chapters and sections will be guided by Figure 3 - 1. In the first stage, three experimental cases will be investigated. Especially, the first experiment is a pre-experiment of the experimental case 2. This experiment is to test the usability of the reported crowdsourced design research method and the crowdsourcing platform, and give a basic understanding of the execution of crowdsourced design. Then in the second experiment, the relationship between the payment to online workers and the design quality will be investigated. Then, based on the experience of the previous, two, experiments, a third experiment that provides a 3D design task will be explored, compared with the 2D design tasks.

After three cases of crowdsourced design experiments, a triangulation method will be used for generalising a crowdsourced design framework will be discussed. The coding method will be used to show all key elements in different crowdsourced design stages, the cross experimental case comparisons and the previous experiments data analysis will also provide a clear view of the developing the framework. This framework is a general crowdsourced design framework being appropriately applied in different crowdsourced design tasks.

In the last chapter of the experimental cases, the framework will be validated by a design task. Each of the elements and stages in the structure of the framework will be validated, and finally the framework will be optimised.

CHAPTER 4

CASE STUDY A: A DESK LAMP DESIGN TASK &

CASE STUDY B: A LIVING ROOM LAYOUT DESIGN TASK

I. INTRODUCTION

The aim of case study A is to understand the process of HBGA and the combination system in the crowdsourced design process, a lamp design task (Case A) was implemented by the author that was posted on MTurk. The design and design evaluation method drew on the experience of the Children Chair design experiment (Yu & Nickerson 2011). The workflow of this experiment is shown below in Figure 4 - 1.

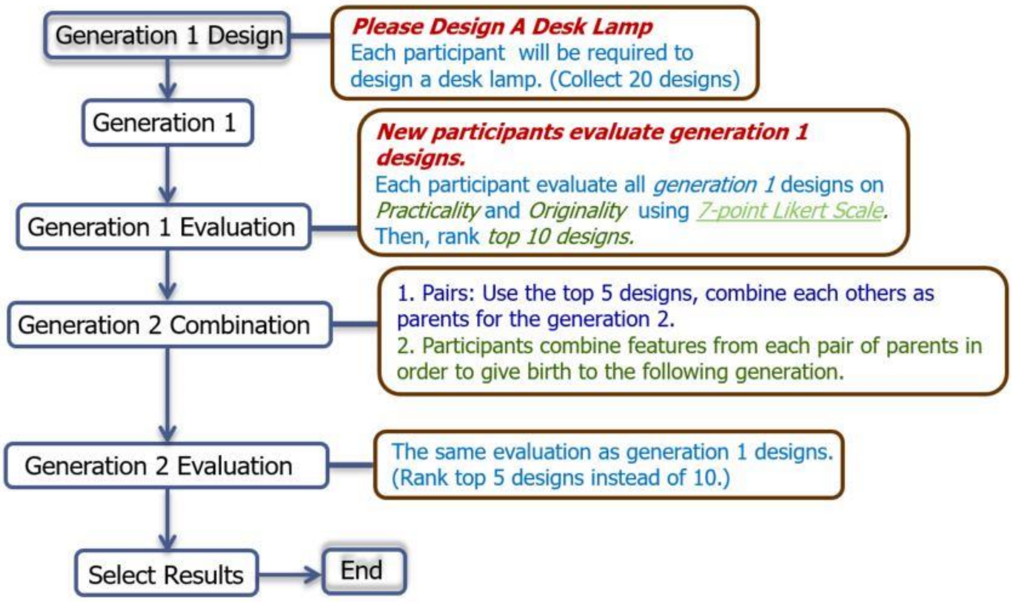


Figure 4 - 1 Desk lamp design task workflow

Case Study A
1st generation design task description

Please design a desk lamp. You could create or innovate a lamp which is the best or the perfect one in your mind. After finishing your design, please upload an image showing your design results. At least, ONE DESIGN RESULT need to be provided.

Besides, if you agree, would you please provide the following information attached in your design results, which could be very helpful for us:

1. Your gender
2. Your age
3. Are you a native English speaker?
4. Your education level, i.e., undergraduate diploma
5. Do you have any design experience? i.e., one-year design experience

MTurk Task A - 01

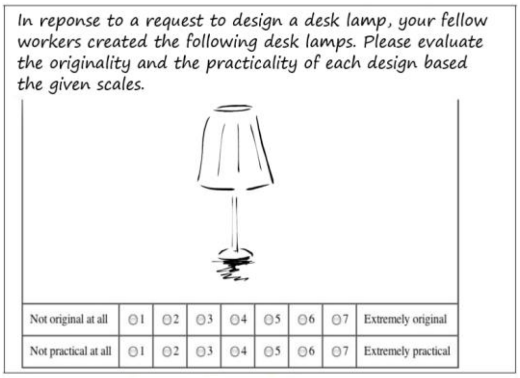
In the first stage, crowd workers were required to create a sketch of a design for a lamp as the first generation. The task description is shown above in the text-box. In this first generation creation task, the payment for each approved submission was $0.25 (any rejected submission get zero reward). Figure 4 - 2 shows the approved submissions for the first generation creation.

In the second stage, the results of the first generation were evaluated by the crowd workers

Case Study A

1st generation desk lamp evaluation task description:

Please evaluate the given 21 desk lamp designs by using the 7-point Likert Scales shown below. (When evaluating designs, you could use "O" to replace "Originality", use "P" to stand for "Practicality") Then please rank the highest 10 designs which are the best or the perfect lamps in your mind. When finished, please upload your result as a Microsoft Word document. The evaluation matrix of “O & P” method is illustrated in the figure below.



“O & P” evaluation method

For Example:
1. Design Evaluation:
No.1: O-5, P-5
No.2: O-4, P-6
No.3: O-3, P-3
No.4: O-5, P-6
... ...
No.21: O-6, P-4
2. Design Ranking:
1. No.3,
2. No.10,
3. No.7,
... ...
10. No.17

Besides, if you agree, would you please provide the following information attached in your design result, which could be very helpful for us:
1. your gender
2. your age
3. are you a native English speaker?
4. your education level, i.e., Bachelor Degree
5. do you have any design experience? i.e., one-year design experience

MTurk Task A - 02

instead of expert designers. The task description is shown below in the textbox:

After the evaluation process, the five best desk lamp designs judged by the crowd workers to be No. 18, No.5, No. 1, No.6, and No. 7 (shown in Figure 4 - 3):

In the third stage, MTurk workers created new lamps (second generation) by combining the good features from the best five designs in the first generation. Figure 4 - 4 illustrates the second generation designs. To assess if the combination process had improved the quality of the designs the set including the original top five (i.e., parents) designs from the first generation. (Figure 4 - 5 shows three desk lamps judged best by the crowd workers).

The demographic information revealed that only 26.89% (32/117) of the workers had design experience. One for the conclusions of the case study A was that even though most participants in MTurk have no design experience, designs can be created by an HBGA method and the combination system on MTurk. It was also noted that because the topic was based on worker common knowledge, there was no need for an explicit description of the design specification or evaluation criteria. Having established the feasibility of the process, case study B investigated the first research question 'what is the relationship between the payment to crowd workers and design quality?' using a living room layout design task.

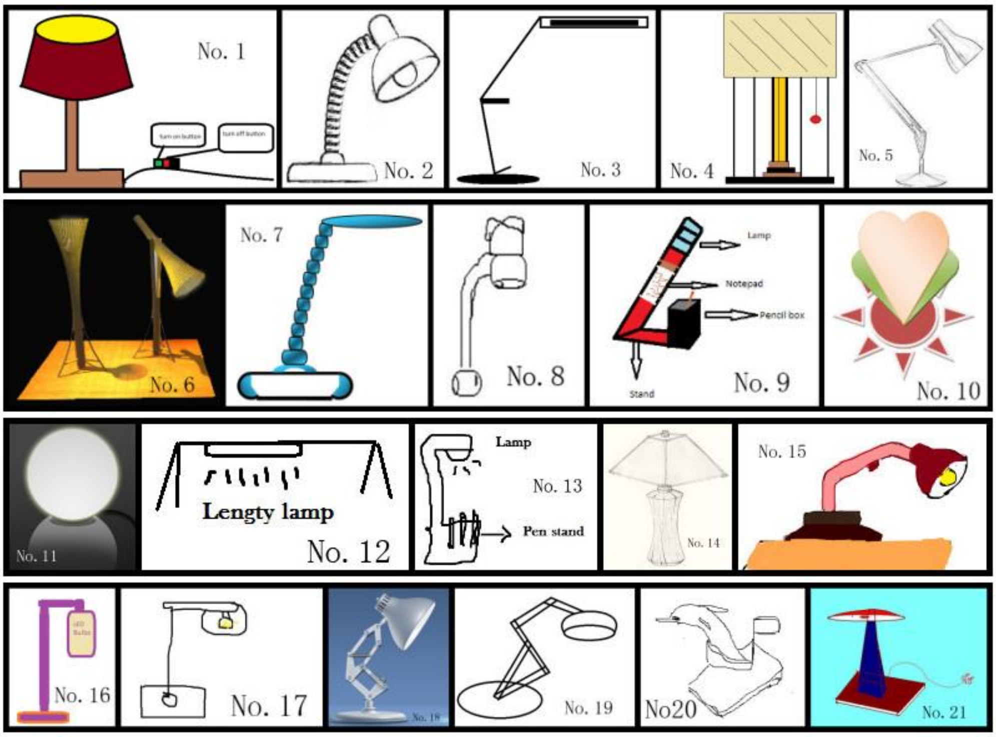


Figure 4 - 2 Case study A: examples of crowdsourced designs from the first generation

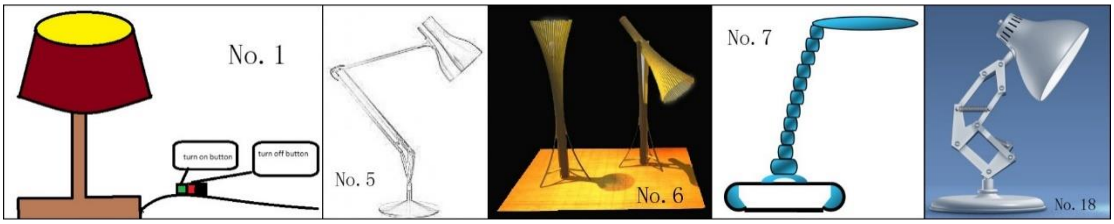


Figure 4 - 3 Case study A: best five desk lamp designs from the first generation

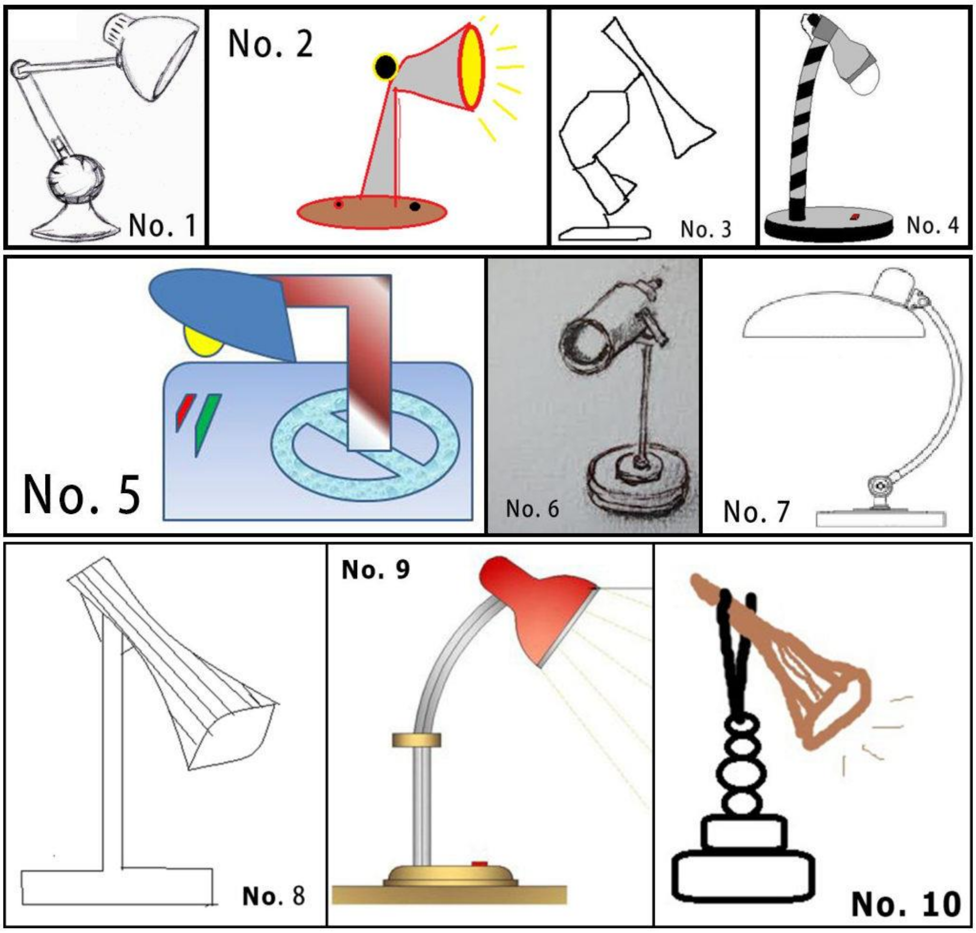


Figure 4 - 4 Case study A: second generation desk lamp designs

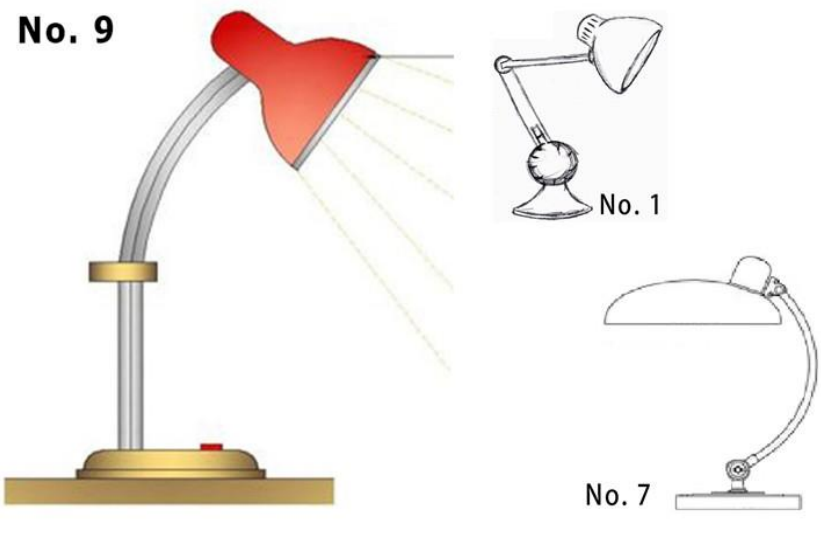


Figure 4 - 5 Case study A: the best 3 desk lamps

II. Case Study B: Relationship between Design Quality and Payment to Workers

Although it is well documented that crowds can design (Yu & Nickerson 2011)(Yu & Nickerson n.d.)(Nickerson et al. n.d.), and crowds can evaluate designs (Bao et al. 2011), the parameters of the process and their sensitivity are unclear.

To investigate the relationship between payment and the quality of design, a 2D interior design (i.e., furniture layout) experiment using MTurk was arranged. The main focus of the experiment was to investigate the relationship between quality of design and the payment for workers. The literature review (chapter 2) noted that little has been reported with regard to that relationship. The work is motivated by the fact that if a relationship between cost and quality could be established, “requesters” may be able to spend less but gain more innovative designs. In terms of the hypothesis, it is believed that with increasing rewards, the quality of design will improve. It is also expected that payment will have a correlation with the speed and quantity of designs submitted.

On MTurk, participants search tasks by their interests as well as by the remuneration. They were not informed that they would undertake crowdsourced design experimental tasks. With the objective of understanding how the level of payment affected the quality of the resulting design, different level of payment for Turkers were correlated with design outcomes. The lowest payment was \$0.15 and the highest \$1.00. Between these two extremes, the levels of payment were set at \$0.35, \$0.50 and \$0.75. This is a generous payment in comparison to other research: the payment as low as \$0.01 (Paolacci et al. 2010), \$0.10 (Kosinski & Bachrach 2012)(Paolacci et al. 2010) (this has been reported). However, in consideration of the experiment’s level of difficulty, the lowest rate was fixed as \$0.15 (Jagadeesan et al. 2009)(Corney & Torres-Sanchez 2010). The reason for setting \$1.00 as the highest reward was that the payment in MTurk is rarely over \$1.00 (Paolacci et al. 2010), only some translation jobs might be paid as much as \$1.40 per hour (Horton & Chilton 2010). After determining these two payment limits, the remaining degrees of payment were set as previously mentioned.

The experiment was divided into two steps- First the Design task: Turkers used “Google Drawing” to design a living room layout within a specified floor plan. The parameters of the MTurk task (e.g., time) were constant with only the payment varying. The second step is the Evaluation task: participants ranked and voted on the designs created by the MTurk workers.

2.1 Living Room Layout Design Task (for Creating Designs)

The design task description is shown in the textbox below:

Case Study B

Design creating task

"Please use Google Drawing to design a livingroom plan. In this livingroom (the plan outline and main size are shown as image 1 below), first, you need to insert the image 1 (shown below) into your new Google Drawing document. Second, draw some appliances and furniture (at least, a TV, a TV bench, Hifi devices, a tea table, a set of sofa are required; specifically, the more the better, the more detailed the better) which fit the outline and main size. Additionally, the position of a window (any size fitting the plan outline and the main size) needs to be fixed. It is also required that you need to add texts into your graphics to explain what are in your drawings."

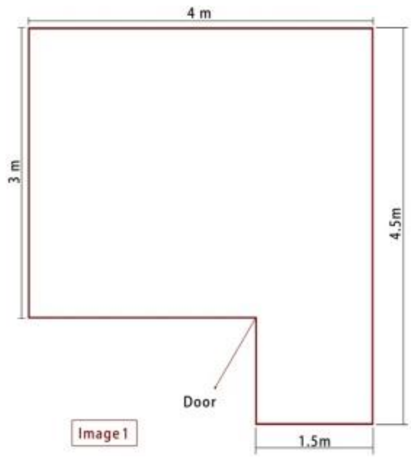


Image 1: download website-link: https://www.opendrive.com/files?Ml8xNTk0MzczOF9idVJEMA)

Result Uploading:

1. Please directly use the snipping tool to upload your design drawing result (screen shot) as a JPEG document (at least 300 x 300 pixels, 96 dpi) for MTurk task submission.

2. Please name your drawing document in Google Drawing as your MTurk worker ID. Then, please share your design drawing result with "h.wu.strath@gmail.com" via Google Drawing sharing tool.

MTurk Task B-01

Additionally, the participants' (i.e., Turkers) individual information was requested to continue building a demographic data set for the people who carried out the design task.

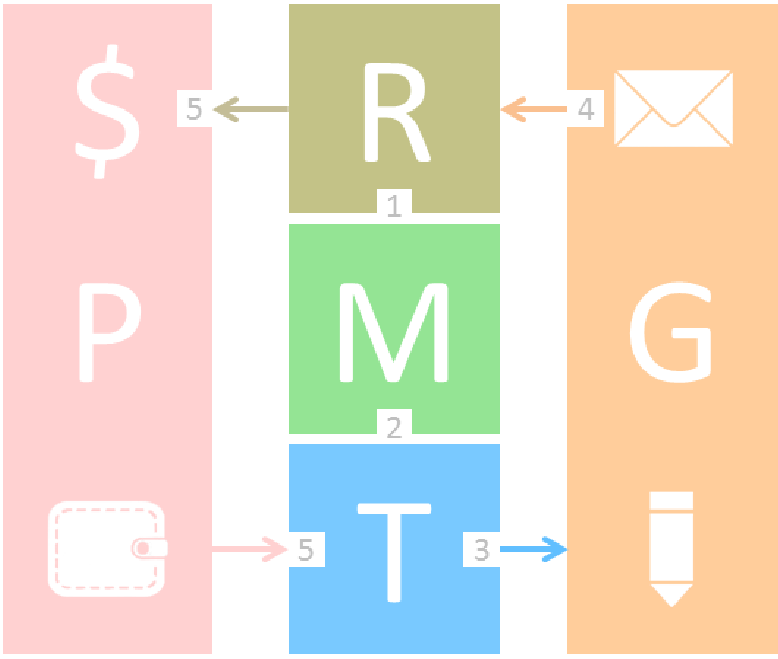


Figure 4 - 6 Case Study B: Main stages of the experiment: 1. Requesters upload problems to MTurk; 2. Turkers find tasks via MTurk; 3. Google Drawing is the tool for participants to draw the living room layout; 4. Workers upload their results to Requesters as well as share the drawings via Google Drawing; 5. Once solutions are approved, Payment will be given to Turkers by MTurk.

Figure 4 - 6 illustrates the main processes involved in the living room layout task performed by the Turkers. The original intention had been to develop an experiment that places no restriction on the innovation of the MTurk workers. However, it was observed that the results of case study A which placed no restriction on the type of CAD tool used were too uneven in their appearance to ensure that they were being objectively compared, and furthermore the solution files were difficult to organize because of their non-uniform format. After reflecting on what kind of platform can be best used as a design tool for crowdsourcing. It was concluded from a review of previous research that this platform should contain at least the following features:

1) free & public to use
2) be easily learnt (if never used before by those undertaking the task)
3) use a standardized file format (to allow easy organization of flies)

Given these requirements, it was determined that the “Google Drawing” system (in Google Drive Applications[8]) would be an appropriate choice. In an early trial task on MTurk, Google Drawing performed well and it was also noted that L. Yu’s Children Chair design experiment (Yu & Nickerson 2011) had also reported that the Google Drawing could work well as a tool for investigating crowdsourced design.

The conditions of the design task were all fixed. Participants had 1 hour to draw the layout (enough time to finish the drawing work), and a series of tasks (each with a different payment) was posted for seven days (after one week the task would expire). The quantity of solutions for each task was unlimited so both the volume of responses and their quality could be studied simultaneously. Previously reported work has only explored a limited range of payments (in Kazai’s experiment for example (Kazai 2011), there were only two levels of payment: $0.10 or $0.25). In contrast, in case study B the design task was repeated five times with different degrees of remuneration. Each time the same task was posted, but with different payment levels being set at: $0.15, $0.35, $0.50, $0.75 and $1.00.

2.2 Qualifying Design Quality: Living Room Layout Evaluation Task (for Design Evaluation)

To allow MTurk workers to evaluate the design quality relative to each other, it was judged important that this was done in groups small enough to be displayed on a single screen to allow easy comparison. To facilitate this, design results were mixed randomly into different groups. This was one using a spread sheet function that randomly generated an integer number between 1 and 8. This enabled the 83 layouts to be separated into 7 groups (12*6+11*1=83) randomly. Each group was posted on MTurk as a separate evaluation task.

After all five living room layout design tasks at the different payment levels had been completed, there were a total 83 drawings created by the MTurk workers and approved for payment. The evaluation process used to judge the design quality comprised the following 3

[8] https://docs.google.com/drawings/, 2014

steps. First: the 83 results were randomly separated into 7 groups randomly. In groups 1-6, there were 12 drawings; in the group 7, there were 11 drawings. Second: each group was posted on MTurk and workers asked to rank them from best to worst. Third: aggregate the results.

Case Study B: Evaluation task

"Please evaluate the 12 different room layouts shown below. The dark red color figure in the lower-left corner in each room layout is the number of the different layout. Please give marks from 0 to 100 to judge the quality of each design. "0" means impractical or not original at all, and "100" means perfect or the best design.

After the layouts design evaluation, please choose the best 3 designs from those 12 layouts, and provide a short description illustrating the reason why you choose them."

MTurk Task B-02

III. RESULTS OF CASE STUDY B

The raw number of layout designs submitted by the MTurk workers (including the approved solutions as well as the rejected designs) rises strongly with the increasing payment level. The time for collecting designs submitted at each payment level is one week. For $0.15 payment, only 3 layouts were received. However, when the payment increased to $1.00, the number of designs submitted was 93.

The 83 "approved" (i.e., accepted for payment by the requesters) room layout designs were assessed for their quality by asking the MTurk worker to assign a mark between 1 and 100 to each member of seven groups. Figure 4 - 7 shows the examples of the approved submissions (all living room layout designs collected in the design task are shown in Appendix 1 - 1, Appendix 1 - 2, Appendix 1 - 3, Appendix 1 - 4, Appendix 1 - 5 and Appendix 1 - 6).

3.1 Demographic Information

Only the designs generated by workers who provide their personal and professional details were analysed.

At the end of the design and evaluation tasks, over one hundred and ten pieces of individual information were collected. The design creation task allowed some limited of validation of the demographic information of this because layouts were submitted by using the “sharing” facility of the Google Drawing tool, allowing individual information could be shown by Google+. Even though a number of participants did not provide their information directly, their Google+ account allowed their genders and nationalities/locations to be found. In some cases, Turkers’ education backgrounds were also shown. By collecting individual information, the composition of the MTurk crowd could be estimated and a comparison made with previous research (Ross et al. 2010)(Khanna et al. 2010)(Yu & Nickerson 2011). The following individual information was collected: gender, age, mother tongue, education level, and whether participants have design experience or not.

The data gathered suggests that the crowd had the following composition, numbers of males and females in layout design and evaluation tasks are similar: females – 47.66%, males – 52.34%. As for the age, the youngest participant is only 19, and the oldest worker is 72. The average age of participants is 29 years old. 25 and 30 are the most frequent ages which appeared 8 times respectively. The second most frequent age is 23 appearing 7 times, which is followed by 26 years old (6 times). In terms of language, 68.35% were English speakers (over 2/3). Further, 72.50% of participants are college/university students or obtain at least associated degrees. But perhaps the most striking result is the worker’s design experience, from the participants’ submission, 74.67% (almost three quarters) of them do not have any design experience (which supports the assertion that crowds can create design, and evaluate design). However, there are 6.67% of people having more than 4 years’ experience of designing.

3.2 Design Evaluation Results

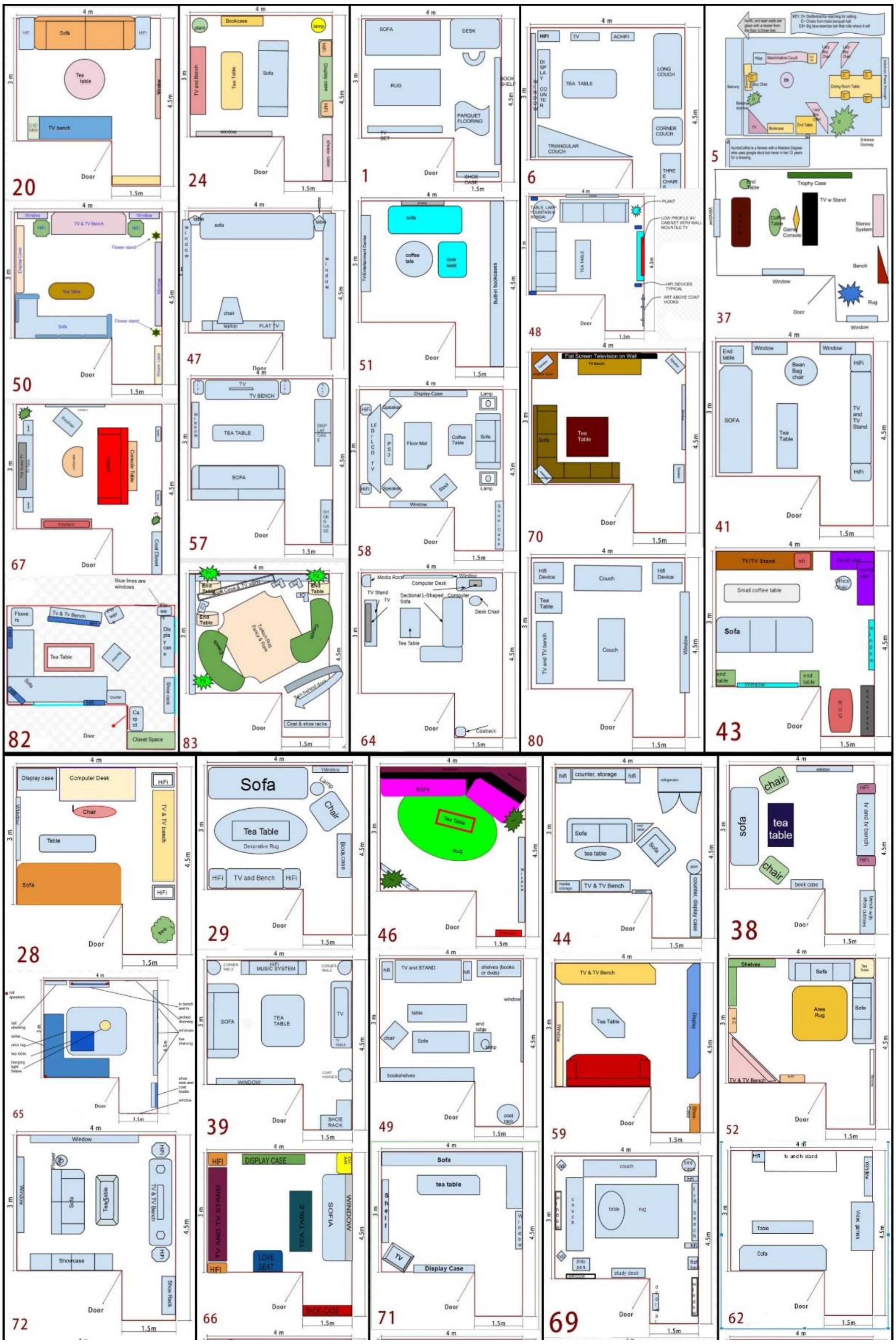


Figure 4 - 7 Case Study B: examples of the approved living room layout designs

The overall average score for each member of each group was calculated and the following observations were made. The highest average mark 59.4 comes from the $1.00 payment and the lowest 53.7 points to $0.75 (rather than $0.15, in Figure 4 - 9). In addition to marking designs, Turkers doing the assessment were also asked to identify the best 3 drawings in each group. Analysis of these votes showed that some of the layouts were not selected even once. Whereas, one layout (in the $1.00 payment group) was nominated 8 times (the most votes). However, reviewing the methodology, it was realized that marking and voting did not provide easily comparable information, because the numerical values of the marks varied with worker personality and cultured background. To allow a non-numerical comparison, a ranking process was performed as discussed in the following section.

IV. DISCUSSION

The results show that the quantity of designs generated by Turkers increased significantly when the payment rose. But for each individual level of payment, the average quality of submissions decreased gradually (Figure 4 - 9) until the payment increased to $1.00 when the number of results rebounds visibly. One possible explanation for that phenomenon is: in MTurk, although there are over 500,000 workers from over 190 countries (Anon 2013), the number of Turkers who are interested in particular categories of work is fixed. When a new task published, workers find it, complete it and gain the payment from requesters after their solution is approved. Then they hunt for new tasks rather than the same one. As a consequence, the first 24 hours always generates the most number of submissions.

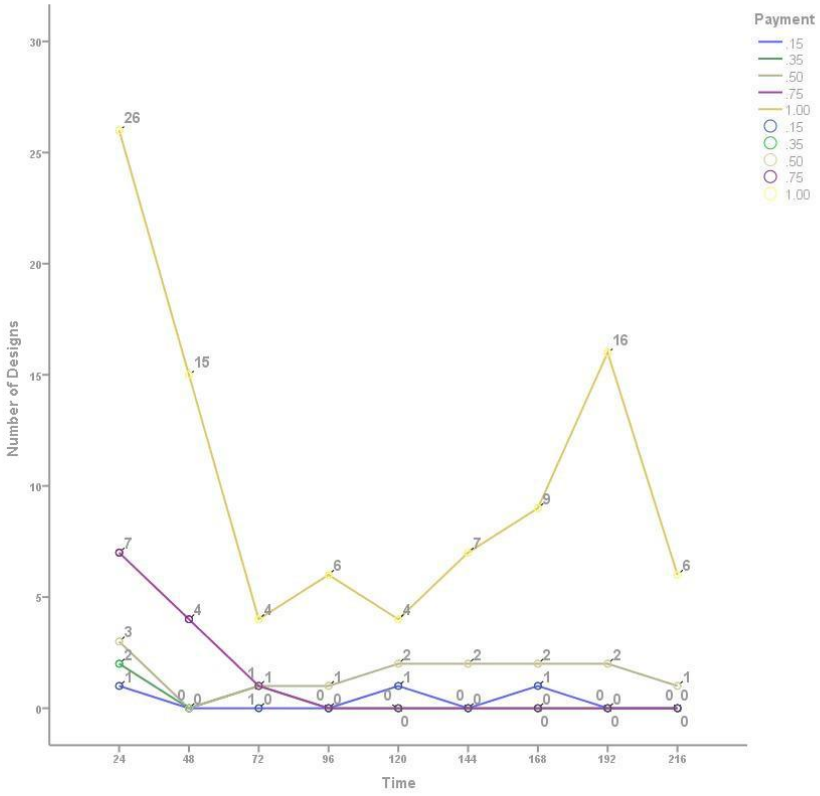


Figure 4 - 8 Relationship between the number of designs v.s. time

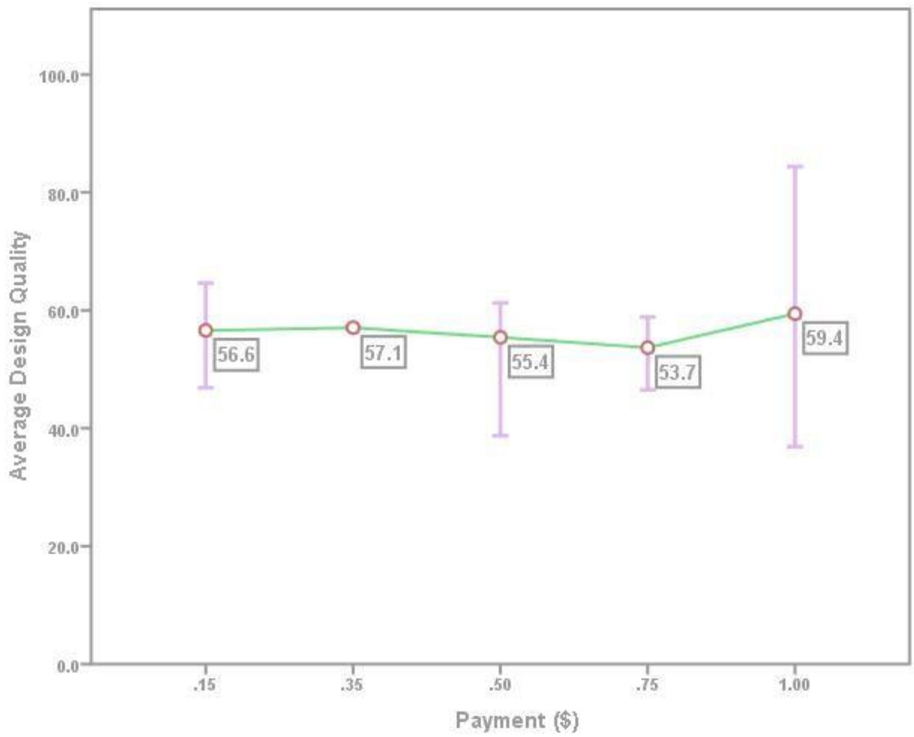


Figure 4 - 9 Spread of average design marks v.s. payment level

In terms of the quality of design, Table 4 - 1 shows the Pearson Correlation between the payment and the average scores of all layouts. It can be seen from the Table that r=0.138>0 (Sig=0.215), and |r|=0.138. The data suggests that although there is a positive correlation between payment and average marks, the correlation is weak.

Table 4 - 1 correlations between payment and the average marks

Correlations			
		Payment	*Average*
Payment	Pearson Correlation	1	0.138
	Sig. (2-tailed)		0.215
	N	83	83
Average	Pearson Correlation	0.138	1
	Sig. (2-tailed)	0.215	
	N	83	83

However, the authors also investigated if there was a stronger correlation with the relative ranking of designs rather than the assignment of absolute percentage value (the scale of which is subjective and may not be consistent from one Turkers to the next). The method used to assign a ranking was: first, in each group, the relative position of each design was calculated based on the average percentage marks from each participant (for each group, there were 10 Turkers marking designs, so each design in a group got 10 different ranks). Second, calculated the average rank for each individual layout, so a smaller number implies higher ranking within the assessment group (i.e., 1 = best, 10 = worst). Finally, the first two highest ranked designs in each group were identified. In this way, the relative quality of the individual designs could be established independent of the percentage scale used by individual MTurk workers.

Table 4 - 2 Ranking Results

Group Number	Highest Ranked Design	Payment for Turker (Highest)	2nd Highest Ranked Design	Payment for Turker (2nd Highest)
1	No. 49	$1.00	No. 72	$1.00
2	No. 50	$1.00	No. 52	$1.00
3	No. 70	$1.00	No. 57	$1.00
4	No. 23	$1.00	No. 35	$1.00
5	No. 63	$1.00	No. 22/No. 11	$1.00/$0.50
6	No. 54	$1.00	No. 08	$0.50
7	No. 60	$1.00	No. 31	$1.00

From Table 4 - 2 above, it can be seen that only designs No.8 and No.11 layout drawing were paid lower than $1.00 ($0.50) (both the highest and lowest ranked designs came from the $1.00 payment, Figure 4 - 10). Figure 4 - 11 shows that both the highest and lowest rankings were generated from the $1.00 payment level. The best fit line in Figure 4 - 11 proves that there is no strong relationship between ranking and payment (R2 Linear=0.008). Figure 4 - 12 illustrated the Fit Line for the best ranking and payment. The highest two rankings come from $0.50 and $1.00 payment (R2 illustrates that the relationship between ranking and payment of the best quality layouts is much stronger than that of the average quality).

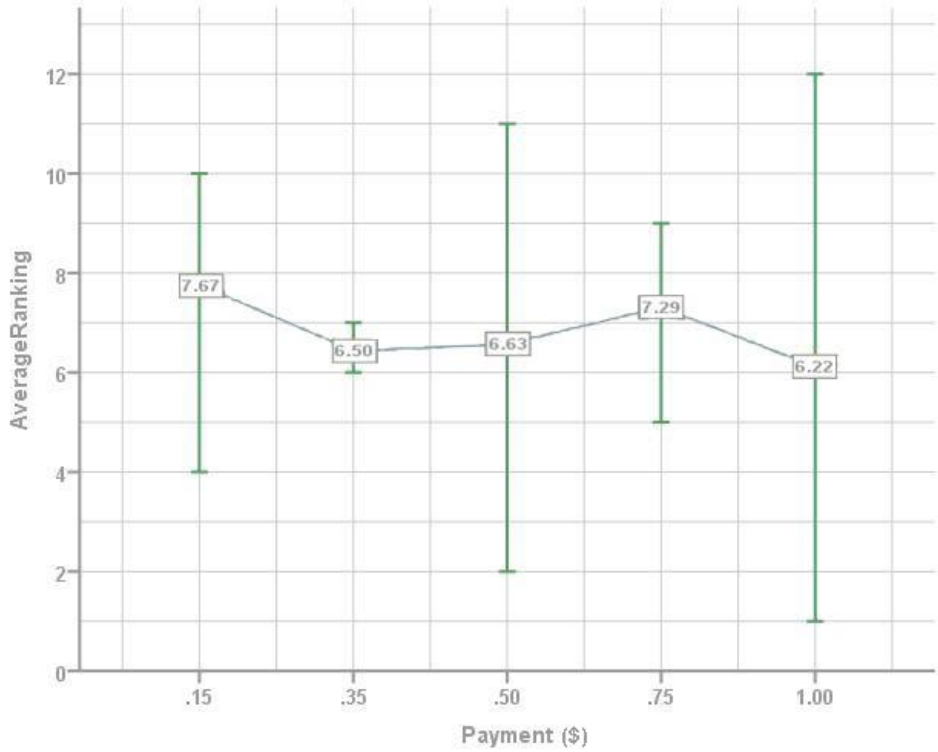


Figure 4 - 10 Spread of ranking v.s. payments where best designs have low ranks (i.e., 1st or 2nd)

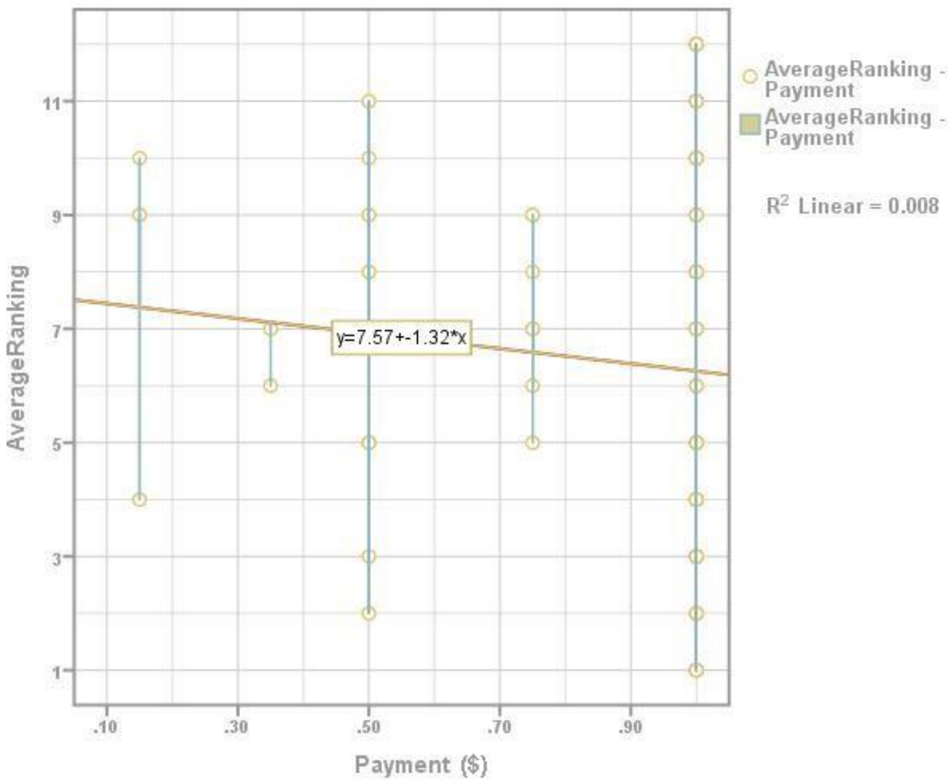


Figure 4 - 11 Best Fit line: average ranking v.s. payment

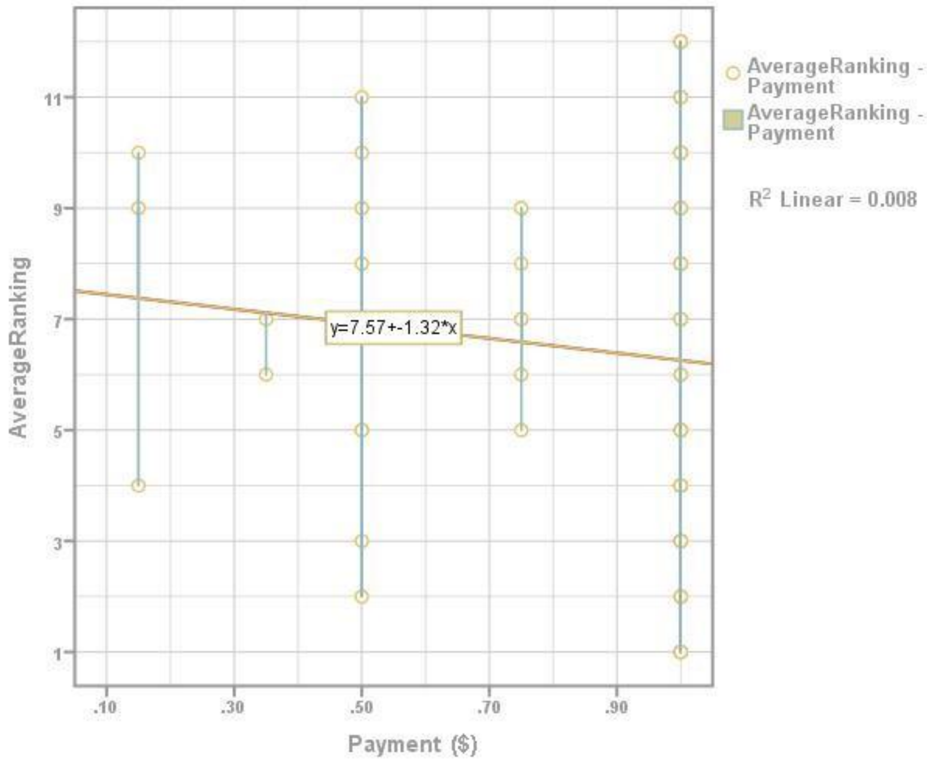


Figure 4 - 12 Best Fit Line: most highly ranked designs v.s. payment

Inspection of the designs judged best in their groups by Turkers (Figure 4 - 13) showed they do indeed have good features relating to the layouts (compared with the low ranking layouts): reasonable utilization of space, diversified types of furniture, practical, and meeting the functional requirements of a living room.

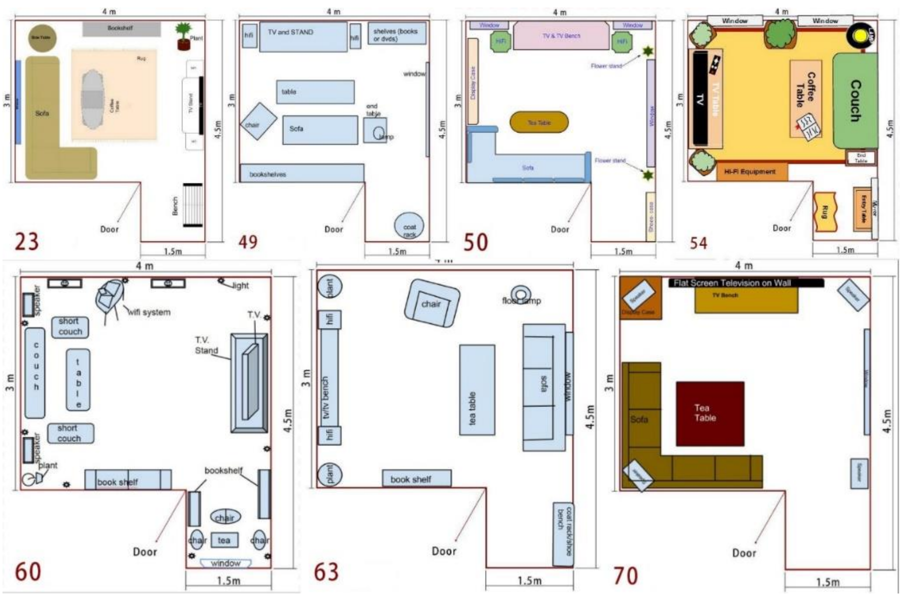


Figure 4 - 13 Examples of the best ranking layouts

V. CHAPTER SUMMARY

In this chapter, a living room layout design and evaluation experiment was presented. The work suggests that the quality of the best layouts increased significantly when the remuneration improved. However, it is interesting that there was no strong evidence that the average quality of designs increases. Indeed, the results suggest that a higher payment generates a larger spread of design quality rather than simply leading only good designs. Figure 4 - 12 shows the highest ranked design of each payment group plotted against price. The best fit line shows a clearly improving quality with price. The results also record a strong increase in the quality of results generated. Consequently, if requesters want to increase the numbers of solutions, they can either one way is to wait for a much longer time or offer a larger payment. Indeed, from the experiments' experience in these case studies, the first 24 hours of any new task beginning posted produces the majority of submissions.

However, the experiment has limitations. In the design task, the number of layout solutions generated is very low for low payment. For instance, only 3 drawing were received when paid \$0.15/\$0.35. In design evaluation and statistics stages, such a number of samples definitely limits the accuracy of any analysis. If the number of layout drawings were increased, the HITs would be given more time than one week. As a result, in future work if more drawings are required, then tasks could be restarted in MTurk. Additionally, participants' personal factors (i.e., career, interest) may influence the design quality in the meantime.

This work suggests that increasing the number of designs generated is key to ensure a high quality work is created. One can hypotheses that this is because the distribution of design quality remains constant and by increasing payment a larger sample size increases the extreme values generated. The next chapter will describe a 3D kitchen room layout design experiment to investigate the second research question 'can 3D designs be crowdsourced as well as 2D designs?'.

Reference

Anon, 2013. Amazon Mechanical Turk_Wikipedia. Available at: http://en.wikipedia.org/wiki/Amazon_Mechanical_Turk#cite_note-awsdevforum-6.

Bao, J., Sakamoto, Y. & Nickerson, J., 2011. Evaluating Design Solutions Using Crowds. In *Seventeenth Americas Conference on Information Systems, August 4th-7th, Detroit, Michigan*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2201651 [Accessed June 25, 2013].

Corney, J.R. & Torres-Sanchez, C., 2010. Putting the crowd to work in a knowledge-based factory. *Advanced Engineering Informatics*, 24(3), pp.243–250. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1474034610000388 [Accessed June 22, 2013].

Horton, J. & Chilton, L., 2010. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on EC'10, June 7–11*. Cambridge, Massachusetts, USA, pp. 209–218. Available at: http://dl.acm.org/citation.cfm?id=1807376 [Accessed September 26, 2013].

Jagadeesan, a. P. et al., 2009. Geometric reasoning via internet CrowdSourcing. *2009 SIAM/ACM Joint Conference on Geometric and Physical Modeling on - SPM '09*, p.313. Available at: http://portal.acm.org/citation.cfm?doid=1629255.1629296.

Kazai, G., 2011. In Search of Quality in Crowdsourcing for Search Engine Evaluation. , pp.165–176.

Khanna, S. et al., 2010. Evaluating and improving the usability of Mechanical Turk for low-income workers in India. *ACM DEV'10, December 17–18, 2010, London, United Kingdom*. Available at: http://portal.acm.org/citation.cfm?doid=1926180.1926195 [Accessed June 22, 2013].

Kosinski, M. & Bachrach, Y., 2012. Crowd IQ: Measuring the intelligence of crowdsourcing platforms. In *WebSci '12 Proceedings of the 3rd Annual ACM Web Science Conference, NY, USA*. pp. 151–160. Available at: http://dl.acm.org/citation.cfm?id=2380739 [Accessed June 22, 2013].

Nickerson, J., Sakamoto, Y. & Yu, L., Structures for creativity: The crowdsourcing of design. In *CHI 2011 Workshop on Crowdsourcing and Human Computation: Systems, Studies, and Platforms, May 8, 2011, Vancouver, BC, Canada.* Available at: http://crowdresearch.org/chi2011-workshop/papers/nickerson.pdf [Accessed June 25, 2013].

Paolacci, G., Chandler, J. & Ipeirotis, P., 2010. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5), pp.411–419. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1626226 [Accessed June 22, 2013].

Ross, J., Irani, L. & Silberman, M., 2010. Who are the crowdworkers?: shifting demographics in mechanical turk. *Proceedings of the 28th …*, pp.2863–2872. Available at: http://dl.acm.org/citation.cfm?id=1753873 [Accessed June 25, 2013].

Yu, L. & Nickerson, J., 2011. Cooks or cobblers?: crowd creativity through combination. *CHI 2011, May 7–12, Vancouver, BC, Canada*. Available at: http://dl.acm.org/citation.cfm?id=1979147 [Accessed June 25, 2013].

Yu, L. & Nickerson, J., Generating creative ideas through crowds: An experimental study of combination. In *Thirty Second International Conference on Information Systems, Shanghai 2011*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2039547 [Accessed June 25, 2013].

CHAPTER 5

EXPERIMENTAL CASE STUDY C

3D KITCHEN LAYOUT DESIGN

I. INTRODUCTION

The results presented in the last chapter and literature (Yu & Nickerson 2011)(Yu & Nickerson n.d.)(Bao et al. 2011) have demonstrated the effectiveness of various forms of 2D design crowdsourcing, its application in 3D design has been less investigated. The aim of the research work presented in this chapter is to investigate if similar processes of creation and evaluation processes for 3D designs can be crowdsourced via an open commercial crowdsourcing site. The objectives implicit in this goal are to answer the second research question 'can 3D designs be crowdsourced using the same methodology as 2D designs?'. This chapter first describes an integrated crowdsourced design process developed from the previous experiment which will be used to structure the creation of a 3D design task on a commercial crowdsourcing platform, and then presents the results of an experiment to test the effectiveness of the crowdsourced design process's application.

Kitchen designing has often been used as a vehicle for academic research (Fischer et al. 1989; Nomura et al. 2001; Fukuda et al. 1997), because it offers a creative task that is both "open" to many (and so suitable for public crowdsourcing) and accessible (i.e., the results can be objectively quantified).

II. EXPERIMENT PROCESS

From the living room layout design experiment in chapter 4, a basic crowdsourced design process can be summarised as follows:

1. the crowdsourced design task preparation stage (stage 1) – Specification Stage

   In the living room layout design experiment, the first decision for the crowdsourced design process was to select: the crowdsourcing platform (i.e., Amazon's Mechanical Turk), design tools (i.e., Google Drawing), the crowd (i.e., Turkers), which method to use (i.e., HBGA), design specifications (i.e., how to design?), design evaluation criteria (i.e., how to evaluate?), etc.
2. the crowdsourced design task prototype stage (stage 2) – Prototype Stage

After the task preparation stage, a prototype of the crowdsourced design task is required to test the practicability of the task. For example, whether the design tool is appropriate for workers to use? Whether the platform is suitable for the task? Whether the payment to worker is need to decrease or increase? Whether the solution submitting method is appropriate for transferring?

3. the crowdsourced design task execution stage (stage 3) – Execution Stage
   If the crowdsourced design task prototype meets the requirements, this design task can be executed on the crowdsourcing platform. For instance, in the living room layout design experiment, a large number of layout designs needed to be collected based on the preparation in the first stage.
4. the crowdsourced design results evaluation stage (stage 4) – Evaluation Stage
   During the design process, the requesters collect results, and the results need to be evaluated to generate the next generation following the HBGA process. In this design process, the evaluation criteria could also be crowdsourced to crowd workers as tasks (i.e., this task is posted to collect the criteria), and these criteria were used to evaluate designs.

These four stages are described in more details.

1. Stage 1

The Specification Stage comprises tasks such as: Platform Selection, Design Tool Selection, “Crowd” Selection, Methodology Selection and Design Workflow. Every design task needs a crowdsourcing platform to host the process and the choice of crowdsourcing platform will reflect the nature of the task: some of the design work can be attempted by anyone regardless of education or background, whereas other tasks require specific experience or education. For example, Amazon’s Mechanical Turk (MTurk) and ShortTask involve workers from all over the world. In contrast, some platforms are only for workers from one country, for example the Taskcn platform has workers mostly from China. After selection of the platform the choice of design representation or tool is the second most important step. Design tools need to be selected for workers as a consideration of the task itself (i.e., 2D design task – 2D design tools or 3D design task – 3D design tools). There are several considerations of design tool selection which are discussed in the ‘Experimental Design’ section. Furthermore, the "crowd" provided by a given platform needs to be selected and consideration given to any specialist skills they might require. In parallel to the fundamental decisions on platform, tool and crowd, the methodology to be adopted in the execution process must be determined at this initial stage. For example, the design task processes can be iteratively, or non-iteratively, executed. Finally, once the methodology is specified the design workflow needs to be discussed (i.e., results' file transfer, shared access to a representation held in the cloud, etc.).

2. Stage 2

Without prior experience of running similar tasks many of the choices made in the specification stage will be educated guesses whose effectiveness is uncertain. There are 6 implementation decisions that need to be specified and validated in Stage 2: the payment for participants; time to undertake the task; clarity of the task instruction; results submission method and the manner in which workers who attempt to scam, or cheat, the system should be handled.

The design of the crowdsourcing task is refined through the process of prototype testing until the require Quantity and Quality (Q & Q) of results are being produced. At which point the process moves to the Execution.

3. Stage 3 & 4

Execution is essential a scaling up of the task for presentation to a larger crowd. The length of the execution stage will be determined by the method set in Stage 1. A competition might last many weeks whereas a Human-based Genetic Algorithms (HBGA) will often cycle through generations of design every few days. In terms of the Evaluation process, regardless of the mechanism used the process ends, with a review of the generated design by a panel of experts who review the crowd's work and select the best outputs. At both the validation and execution stages the ability to accurately evaluate designs is crucial to tasks such as the setting of payment levels (Stage 2) or selecting the best design for iterative improvement

(Stage 3). The next sub-section describes an experiment, in terms of the cDesign framework, that was created to investigate the framework's application in 3D design area.

III. EXPERIMENTAL DESIGN - 3D INTERIOR DESIGN EXPERIMENT

1. Stage 1: Specification

The nature of the design brief will determine the platform, design tools, crowd type, methodology and workflow. In this case, a public crowdsourcing platform (MTurk) was selected rather than a specialise site (e.g., GrabCAD for engineering, or Taskcn for graphic design experts). The literature suggests that MTurk can be selected as an effective tool to get work done quickly and at minimal cost. What is more, all people using the internet and having an account on the crowdsourcing platform would be welcome to participate in the design as well as the subsequent evaluation experiments.

In terms of the design tool, from prior experience of crowdsourced design tasks (Wu et al. 2014b; Wu et al. 2014a) a review of previous research concluded that the CAD tools used for public crowdsourcing sites should contain the following features:

1) they should have minimal barriers to use (i.e., low cost or free, little or no installation, no registration)
2) be easily learnt (so workers who have never used the tool before can still undertake the task)
3) use a standardized file format (to enable easy processing of results and organizing files).

In the process of creating the design experiment, the following free & cloud based 3D design tools were considered: "Build with Chrome TM", TinkerCAD TM and Homestyler TM. Finally, the Homestyler was selected as the modelling tool. The reasons can be summarized as follows: although "Build with Chrome" is extremely easy to use, the results could not be saved, download and shared between different accounts (i.e., workers and solution seekers).

Secondly, after running the pre-experiments of the last two tools, it was obvious that TinkerCAD is a more open-ended platform, which means participants could easily submit nominal work (i.e., a simple and crude geometrical solid, instead of a 3D kitchen model) in an attempt to cheat the system. What is more, in Homestyler, there are already a large number of kitchen utensils/decorations/appliances (i.e., microwaves, cooking bench, tables, chairs and range hood, etc.). So participants could focus on designing the layout of kitchen model rather than the creation of geometry. Additionally, the layout design model could be saved as "public" so that participants could share the results by the URL links, which provide the possibility of design collaboration.

When considering the methodology selection, there are two generic crowd design methodologies namely 1) linear competition (non-iterative) and 2) iterative improvements. Compared with the linear competition, iterative improvement can range from the very structured HBGA process to a looser process, where workers compete for bonus payments by improving on previous solutions. The experiment's process is illustrated in Figure 5 - 1, and described as follows:

Stage 1: Specification

In Stage 1, it is sufficient that the high level methodology is fixed. This choice will allow the workflow to be defined. In the case of the 3D interior design layout task, it was decided that an iterative process would be suitable since the objective is to generate designs and importantly, use the crowd to improve designs as well as evaluate them.

Participants on MTurk create designs marked as the 1st Generation (G1) (design task).

1. G 1 designs were evaluated on MTurk to select the top 3 designs (evaluation task).

The top 3 designs are then combined with each other to create the combination generation (C1= No.1 combines No. 2, C2=No. 2 combines No. 3, C3=No. 1 combines No. 3, 'C' means combination), and each combination design collects three results (combination task). For each 3 combination design groups, participants would evaluate them to select the best combination designs (3 best designs: C1B, C2B and C3B) (evaluation task).

Final evaluation: integrate the best combination designs (C1B, C2B and C3B) and the top 3 designs from G1, and then evaluate them (evaluation task).

Figure 3 shows examples of the kitchen layouts developed in "Homestyler".

2. Stage 2: Validation (Prototype)

Generally, when posting a task on a crowdsourcing platform, the parameters required are: 1), the payment for workers; 2), how much time should be given to workers; 3) how they submit their solutions; 4) how to avoid cheats. Then the task instruction can be integrated.

The last chapter has discussed that there is a weak correlation between the level of payment and the average quality of results (Wu et al. 2014b). This results was obtained using generous payment levels in comparison to other reported research studies (which could be as low as $0.01 (Paolacci et al. 2010), $0.10 (Kosinski & Bachrach 2012; Paolacci et al. 2010)). And the payment on the platform used is rarely over $1.00 (Paolacci et al. 2010), only some translation jobs might be paid as much as $1.40 per hour (Horton & Chilton 2010). However, although the relationship with quality is weak the correlation with quantity is very strong and after the pre-experiment (in which participants were paid $0.50, resulting in only six results (including cheats) were collected), the payment was increased to $1.00.

Additionally, based on observation and prior experience of working on MTurk, it was found that for even a simple task, it was important that requesters give enough time to participants for undertaking the task. In the previous 2D layout design experiment (Wu et al. 2014b), participants had an hour to submit their work. Because the 3D layout design task is more complicated than the previous 2D tasks reported by the authors, crowd workers were 90 minutes. Similarly in the later tasks such as the design combination task, workers were also paid $1.00 for design combining within 90 minutes to complete the work.

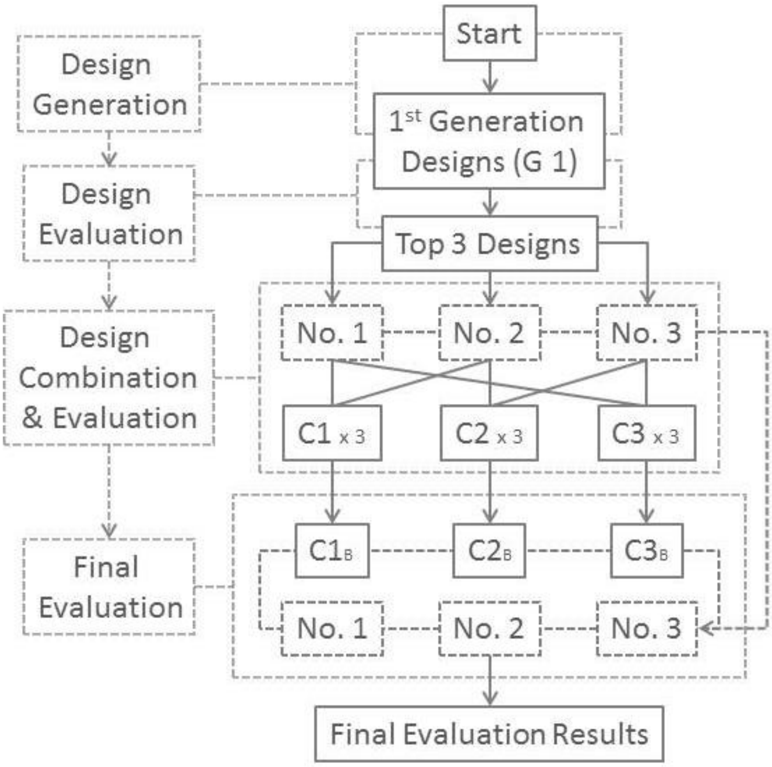


Figure 5 - 1 The workflow of the 3D kitchen layout design experiment

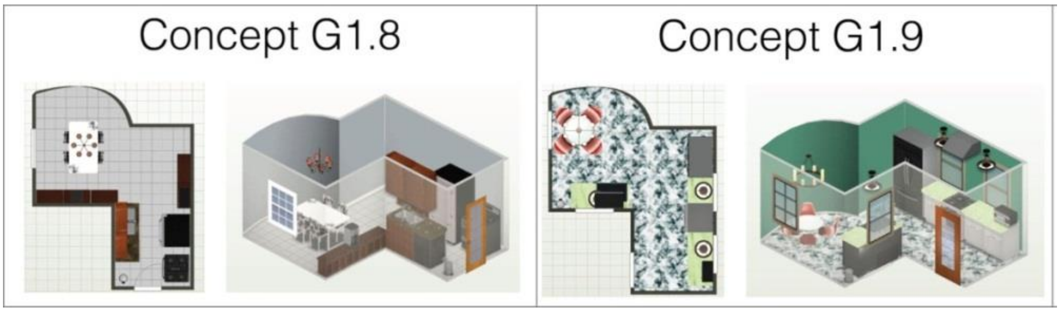


Figure 5 - 2 Examples of the kitchen designs by Homestyler

The adoption of the Homestyler system as a design tool to require some changes to the way the crowd workers returned their work. Native Homestyler files could not be submitted via MTurk directly, so instead when workers finished their design (or finished design combinations), they "shared" the results (saving the result as "public" in Homestyler and sending an email to the requester) and also submitting a screen-shot of the results via MTurk (so that they could get paid).

In addition, in crowdsourcing platforms, a number of workers always attempt to subvert or cheat the system (Eickhoff & Vries 2012)(Wang et al. 2012)(Little et al. 2009). In this experiment, the following methods were used to validate submissions. Based on the fixed design tool, participants would need to share a URL to their Homestyler design with the requester (instead of just submitting results on MTurk), which would then require workers to design the layouts by themselves (and so avoiding people simply submitting random images).

Once the above decisions had been made the focus turned to the writing of the task instruction. Several drafts were reviewed to make sure the text was easy to understand and as clear as possible. The design instructions are detailed as follows:

Case Study C: kitchen layout design task description

Title of the task: Kitchen layout design

Instructions: Using HomeStyler, (www.homestyler.com) create a layout for a kitchen. You have to use this floor plan, http://www.homestyler.com/designprofile/1ff7fd76-2f0c-4b74-b8b2-05e313e96b50 , as the base model for your design. To use this model first click on "Explore" and then click on "Start your own design".

If it is your first time using this online tool you can watch this YouTube tutorial to learn the basics steps: https://www.youtube.com/watch?v=3cWJADWp-aM

Be sure to fulfil ALL THE REQUIREMENTS available in the following document: https://www.dropbox.com/s/5eylwaj34wtxgc2/Specifications%20.pdf

Your kitchen SHOULD INCLUDE, but not limited to:

- Refrigerator
- Dishwasher
- Cooker
- Cupboards
- Basin/Sink
- Waste bin
- Microwave
- Work surfaces
- Sitting/Eating area

When finished your project you should save and mark it as PUBLIC in HomeStyler.

Duration: 6 days (max 90min)

How to share the results: Send an email to h.wu.strath@gmail.com with/message the subject title "HomeStyler project _your MTurk account ID" and copy and paste the link of your project inside the body text. Also, could you please submit a screenshot of your result in mTurk so that you could get paid.

MTurk Task C - 01

After the design workflow was fixed, the task was made available on the Crowdsourcing platform for a small number of workers to test the job's design. This allowed prototype results to be judged in terms of their quality and quantity (i.e., Q & Q). As a result, after the evaluation showed that the results satisfy the Q & Q, the design and its workflow was deemed suitable to move to the Execution Stage. If this had not happened the prototype would needs to be corrected, until it reaches the Q & Q threshold.

3. Stage 3: Execution

Based on the cDesign methodology framework, there are two main choices in the design execution: Non-iterative Design Task (NIDT) and Iterative Design Task (IDT). From the section above it can be seen that this experiment was applied to the iterative design method. After the Prototype Design stage, the design method, platform, payment and the design tool were all validated such that the task could collect design results of acceptable quality. The following sections discuss the design results, design evaluation method and the evaluation results. Since the main purpose of the design experiment was to investigate how crowdsourcing as a tool could be applied to 3D design, an iterative design method was employed. The interior 3D design task was completely (after validation of the design prototype) and in total 10 approved 3D layout designs were collected as the 1st generation designs (G1).

4. Stage 4: Evaluation

After collecting the 10 results as the 1st generation, the evaluation task was then posted on MTurk (also, the combination designs' evaluation task applied to the same evaluation method). In the evaluation process, in total, 10 criteria were provided to workers using a 5-Point Likert Scale for the assessment (where 1 means very bad and 5 means excellent) (the 10 criteria are shown in Figure 5 - 3). The Google Drive platform was used for workers to participate in the evaluation task. A form with all the criteria could be created and shared via the MTurk system with a link and the answers would be interrogated on an Excel spread sheet at the same time as the criteria were being judged.

QUESTIONS	
ERGONOMICS	Rate the following designs according to how well you can circulate around the kitchen (e.g. when you move, are you going to hit any object?)
SAFETY	Rate the following designs according to how well the position of the oven and other dangerous objects is displayed - thinking about the safety of the user.
	Rate the following designs according to how well the cabinets are displayed if the user of the kitchen wants to rest a hot pan after taking it out of the oven or the microwave.
EFFICIENCY	Rate the following designs according to how well this kitchen would fit in your daily life.
	Rate the following designs according to how well the sink is placed (e.g. near the food preparation area).
	Rate the following designs according to how well eating area is displayed (e.g. in a place where everyone feels comfortable, can move the chairs and can leave the table without disturbing the others).
CRIATIVITY and APPERENCE	Rate the following designs according to how appealing and attractive the cooking area looks.
	Rate the following designs according to how appealing and attractive the eating area looks.
	Rate the following designs according to how well the kitchen was decorated with ornaments and other objects.
	Rate the following designs according to how well the colours work together in order to create a pleasant look.
	Answer this question with the same answers that you gave in the question number 2.
	Prove you're not a robot! Copy what is written in the image.

Figure 5 - 3 Ten evaluation criteria

Based on the prior experience of a living room layout design experiment (Wu et al. 2014b), in the evaluation task, participants were paid $0.25 in the pre-experiment. However, considering that participants were required to judge 10 concepts, and for each concept they needed to evaluate 10 criteria (so in total each worker had to answer 100 questions). Consequently the payment was increased to $0.50.

To avoiding participants cheating in the evaluation process (i.e., giving scores randomly), a CAPTCHA (Completely Automated Public Test) image was used. Additionally it was easily detect this kind of random answers in the Excel spread sheet summarizing the results, so they were eliminated and not used in any further analysis. The next section discusses how designs were combined after the evaluation of the 1st generation designs as well as the final design and evaluation results.

IV. DESIGN COMBINATION AND RESULTS

1. 1st Generation Layout Design and Evaluation Results

Based on the design process from the experiment workflow, in the first stage, participants submitted 10 approved 3D kitchen layout designs named from G1.1 to G1.10 (in figure 4). Then workers evaluated them on the basis of 10 criteria (e.g., easy to move around, relative location of key objects, etc.) with the scores ranging from 1 to 5. In total, 36 approved evaluation results were collected.

As a result, in terms of the layout design ranking, the score of the sum of the average for each criterion was ranked (1) (SG1 means the average score of each criterion for the 1st generation designs, SS1 means the sum of all evaluation scores for one design, N means the number of participants who evaluated design. To get an average score the results were summed and divided by 10 (i.e., there are 10 criteria for each evaluation).

$$S_{G1} = (S_{S1} / N) / 10$$

The results show that the top ranked designs from generation 1 are: G1.7 (3.755 scores), G1.5 (3.664 scores), and G 1.10 (3.536 scores) (in table 1). So these three designs from the 1st generation were selected to be combined in the next step in the hope of producing better designs.

2. Design Combination and Evaluation Results

2.1 1st generation designs combination

Once the best three designs from the 1st generation were selected, the combination process started. Previously from the 2D sketch experiment (Yu & Nickerson 2011) it had been reported that better designs come from combinations, it was assumed that in 3D design experiment, then similar results would probably emerge. Consequently, the best three 1st generation designs were combined by Turkers (i.e., workers on MTurk) (the task description is shown below).

Case Study C: design combination task

COMBINATION HIT 1

Title of the task: Kitchen layout combination design Task 1

Description: Combine two kitchen designs to create a new one using HomeStyler

Keywords: kitchen, design, furniture, housing, combine, interior design

Instructions: Please use HomeStyler (www.homestyler.com) to create a new layout for a kitchen by combining the best features (you think) of the following two layouts.

LAYOUT 1: http://www.homestyler.com/designprofile/3f7b92c8-19ad-4008-8b6a-5fcbe731b5d7

LAYOUT 2: http://www.homestyler.com/designprofile/70003ffc-7503-4b1a-b159-356fdd96c48b

You could only use the same floor template plan, http://www.homestyler.com/designprofile/1ff7fd76-2f0c-4b74-b8b2-05e313e96b50 as the base model for your design.

DO NOT CHANGE ALL THE DESIGN. It's expected that the final result must be a COMBINATION of the TWO given layouts.

To use this model first click on "Explore" and then click on "Start your own design".

If it is your first time using this online tool you can watch this YouTube tutorial to learn the basics steps: https://www.youtube.com/watch?v=3cWJADWp-aM

Your kitchen MUST INCLUDE, but not limited to:

- Refrigerator
- Dishwasher
- Cooker
- Cupboards
- Basin/Sink
- Waste bin
- Microwave
- Work surfaces
- Sitting/Eating area

When finished your project you should save and mark it as public in HomeStyler.

Duration: 2 days (90 min)

How to share the results: Send an email to h.wu.strath@gmail.com with/message the subject title "HomeStyler Combination_your MTurk account ID" and copy and paste the link of your project inside the body text. Also, could you please submit a screenshot of your result in mTurk so that you could get paid.

MTurk Task C - 02

Case Study C: design combination task

COMBINATION HIT 2

Title of the task: Kitchen layout combination design Task 2

Description: Combine two kitchen designs to create a new one using HomeStyler

Keywords: kitchen, design, furniture, housing, combine, interior design

Instructions: Please use HomeStyler (www.homestyler.com) to create a new layout for a kitchen by combining the best features (you think) of the following two layouts.

LAYOUT 1: http://www.homestyler.com/designprofile/3f7b92c8-19ad-4008-8b6a-5fcbe731b5d7

LAYOUT 2: http://www.homestyler.com/designprofile/b73e647b-eeef-459e-b70a-e7c22088a041 (someone changed the design, it has to be changed as: http://www.homestyler.com/designprofile/061a0ed6-17c7-45cc-bbdd-7c3cfdad09b6)

You could only use the same floor template plan, http://www.homestyler.com/designprofile/1ff7fd76-2f0c-4b74-b8b2-05e313e96b50 as the base model for your design.

DO NOT CHANGE ALL THE DESIGN. It’s expected that the final result must be a COMBINATION of the TWO given layouts.

To use this model first click on “Explore” and then click on “Start your own design”.

If it is your first time using this online tool you can watch this YouTube tutorial to learn the basics steps: https://www.youtube.com/watch?v=3cWJADWp-aM

Your kitchen MUST INCLUDE, but not limited to:

- Refrigerator
- Dishwasher
- Cooker
- Cupboards
- Basin/Sink
- Waste bin
- Microwave
- Work surfaces
- Sitting/Eating area

When finished your project you should save and mark it as public in HomeStyler.

Duration: 2 days (90 min)

How to share the results: Send an email to h.wu.strath@gmail.com with/message the subject title “HomeStyler Combination _your MTurk account ID” and copy and paste the link of your project inside the body text. Also, could you please submit a screenshot of your result in mTurk so that you could get paid.

MTurk Task C - 02

Case Study C: design combination task

COMBINATION HIT 3

Title of the task: Kitchen layout combination design Task 3

Description: Combine two kitchen designs to create a new one using HomeStyler

Keywords: kitchen, design, furniture, housing, combine, interior design

Instructions: Please use HomeStyler (www.homestyler.com) to create a new layout for a kitchen by combining the best features (you think) of the following two layouts.

LAYOUT 1: http://www.homestyler.com/designprofile/70003ffc-7503-4b1a-b159-356fdd96c48b

LAYOUT 2: http://www.homestyler.com/designprofile/b73e647b-eeef-459e-b70a-e7c22088a041 (someone changed the design, it has to be changed as: http://www.homestyler.com/designprofile/061a0ed6-17c7-45cc-bbdd-7c3cfdad09b6)

You could only use the same floor template plan, http://www.homestyler.com/designprofile/1ff7fd76-2f0c-4b74-b8b2-05e313e96b50 as the base model for your design.

DO NOT CHANGE ALL THE DESIGN. It's expected that the final result must be a COMBINATION of the TWO given layouts.

To use this model first click on "Explore" and then click on "Start your own design".

If it is your first time using this online tool you can watch this YouTube tutorial to learn the basics steps: https://www.youtube.com/watch?v=3cWJADWp-aM

Your kitchen MUST INCLUDE, but not limited to:

- Refrigerator
- Dishwasher
- Cooker
- Cupboards
- Basin/Sink
- Waste bin
- Microwave
- Work surfaces
- Sitting/Eating area

When finished your project you should save and mark it as public in HomeStyler.

Duration: 2 days (60min)

How to share the results: Send an email to h.wu.strath@gmail.com with/message the subject title "HomeStyler Combination _your MTurk account ID" and copy and paste the link of your project inside the body text. Also, could you please submit a screenshot of your result in mTurk so that you could get paid.

MTurk Task C - 02

The combination process is illustrated in Figure 5 - 5. G1.7 was combined with G1.5 (C1: G2.1, G2.2 and G2.3); G1.5 with G1.10 (C2: G2.4, G2.5 and G2.6); G1.7 with G1.10 (C3: G2.7, G2.8 and G2.9) ('C' means combination).

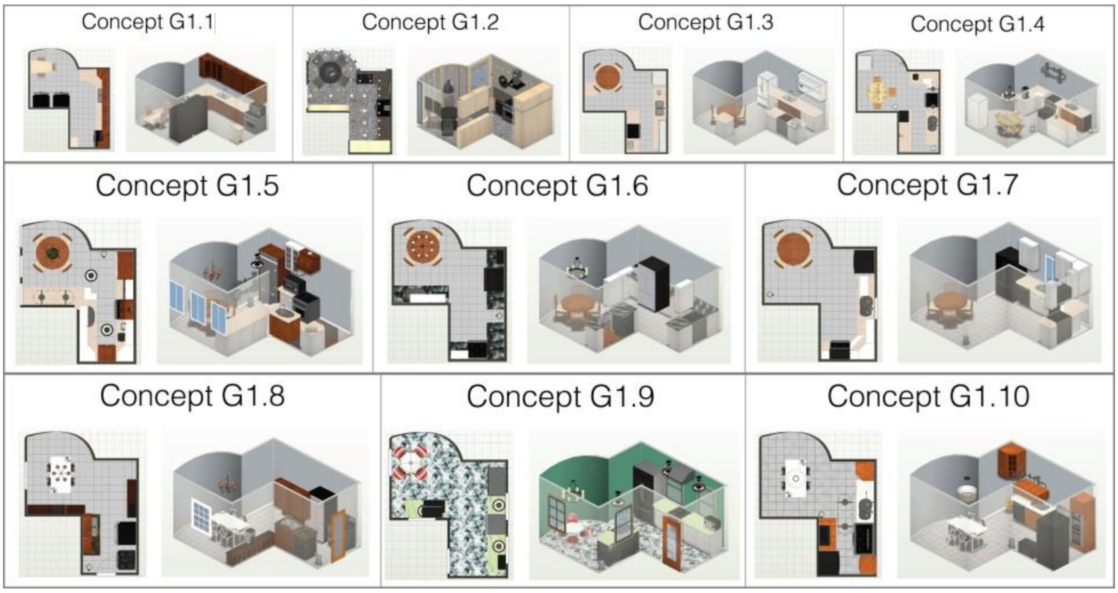


Figure 5 - 4 Ten 1st generation layout designs, link for high quality image

Table 5 - 1 Scores for the 1st generation designs

Generation 1 Evaluation				
	Total	Average	**Total of**	**36 Answers**
G1.1	1221	3.392		9 Brazilians
G1.2	1217	3.380		8 Indians
G1.3	1163	3.231		1 Irish
G1.4	1175	3.264		18 Americans
G1.5	1319	3.664	**Average Age**	32.80
G1.6	1131	3.142		
G1.7	1352	3.755		
G1.8	1140	3.167		
G1.9	1117	3.105		
G1.10	1273	3.536		

Table 5 - 2 Scores for the combination designs

Scores for G1.7 and G1.5				
	Total	Average	**Total of**	**16 Answers**
G2.1	**567**	**3.544**		6 Indians
G2.2	565	3.531		10 Americans
G2.3	511	3.194	**Age Average**	32.87
Scores for G1.5 and G1.10				
	Total	Average	**Total of**	**15 Answers**
G2.4	548	3.650		6 Indians
G2.5	**564**	**3.760**		9 Americans
G2.6	518	3.450	**Age Average**	30.20
Scores for G1.7 and G1.10				
	Total	Average	**Total of**	**15 Answers**
G2.7	547	3.647		5 Indians
G2.8	**577**	**3.847**		8 Americans
G2.9	568	3.787		2 Russian
			Age Average	28.00

2.2 Combination designs evaluation

From the combination process, each pair created three combination layouts. Because only one of those three designs for each pair would be selected for the next final evaluation stage, an evaluation task for them was published on MTurk which applied the same evaluation method – 5-Point Likert Scale as in the 1st generation evaluation stage. From Table 5 - 2, it can be seen that for each pair, the winners are C1B: G2.1, C2B: G2.5 and C3B: G2.8 ('B' means the best).

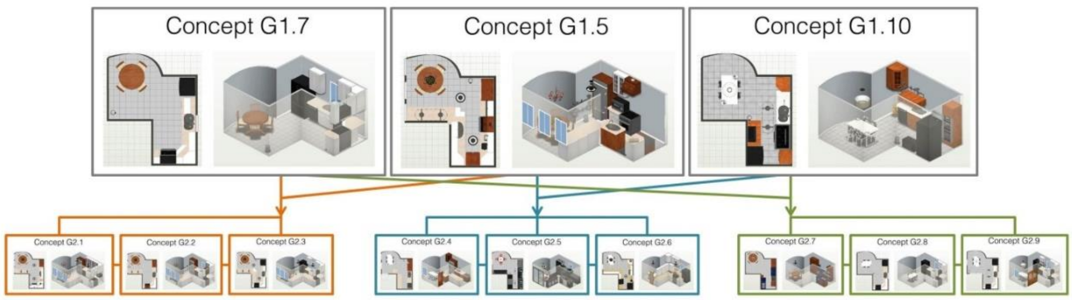


Figure 5 - 5 Layout Combination Progress

V. DISCUSSION

A detailed comparison of the results of this experiment with those of the other studies is presented in the discussion chapter. But here the results will be discussed in isolation. The first observation is that the design combination process worked and the quality of the combined designs (as assessed by the crowd) was improved enormously (Yu & Nickerson 2011). The scores of the best designs in Table 5 - 1 and Table 5 - 2 are illustrated in Figure 5 - 6by means of the lowest and the highest scores in the combination generation where it can be seen that 2nd generation results (pink) are all higher than those of the 1st generation (blue). Additionally, the average score of the best designs in the combination generation (3.717) is also higher than those of the 1st generation (3.652), which means the quality of designs did increase in the later generation.

The crowd's evaluation was validated by inspection of all the generated layouts by two professional architects. The architects confirmed that the quality of generation two (combined designs) was superior to generation one, and selected design G2.1 as the best overall. This judgement by expert practitioners confirms that the crowd's ability to evaluate designs (verifying that the best features had been effectively combined).

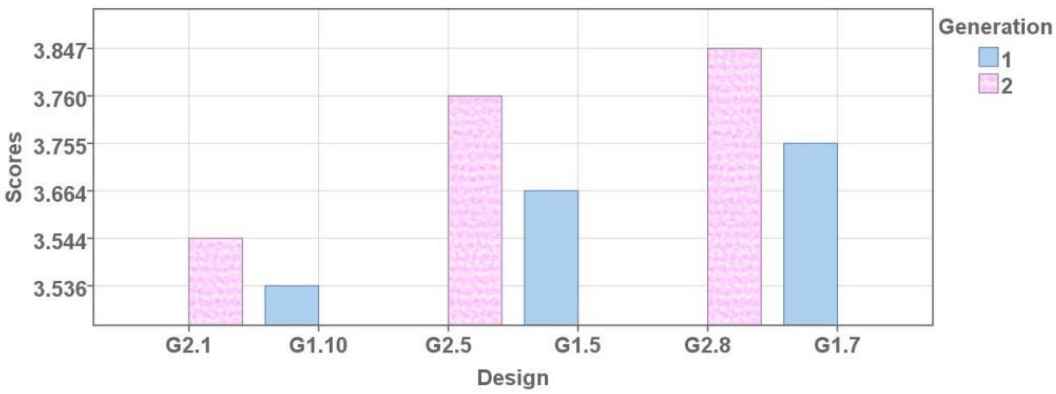


Figure 5 - 6 Scores of the best 3 designs from generation 1 and the generation 2 (combined designs)

VI. CHAPTER SUMMARY

These results demonstrate that a 3D cloud based CAD tool can effectively support an iterative crowdsourced design and evaluation tasks. The results further support the assertion that this type of tool (i.e., open, cloud based CAD) is appropriate for both 2D and 3D crowdsourced design tasks. Furthermore, the basic crowdsourced design process presented in this chapter in the second section has proved flexible enough to support 3D design tasks. However, because the number of the experimental results especially in the design combination stage was relatively limited the results can only weakly support the assertion that the use of the crowdsourced design method and strategy (i.e., HBGA) is effective for 3D work. In the next chapter, based on the experiment results in chapter 4 and chapter 5, a Crowdsourced Design (cDesign) Framework will be inducted and illustrated.

Reference

Bao, J., Sakamoto, Y. & Nickerson, J., 2011. Evaluating Design Solutions Using Crowds. In *Seventeenth Americas Conference on Information Systems, August 4th-7th, Detroit, Michigan*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2201651 [Accessed June 25, 2013].

Eickhoff, C. & Vries, A. de, 2012. Increasing cheat robustness of crowdsourcing tasks. *Information Retrieval*, 16(2), pp.121–137. Available at: http://link.springer.com/10.1007/s10791-011-9181-9 [Accessed June 22, 2013].

Fischer, G., McCall, R. & Morch, A., 1989. Design environments for constructive and argumentative design. In *Proceedings of the SIGCHI conference on Human factors in computing systems Wings for the mind - CHI '89*. New York, New York, USA: ACM Press, pp. 269–275. Available at: http://portal.acm.org/citation.cfm?doid=67449.67501.

Fukuda, T. et al., 1997. Networked VR System : Kitchen Layout Design for Customers. In *Proceeding VRML '97 Proceedings of the second symposium on Virtual reality modeling language*. pp. 93–100.

Horton, J. & Chilton, L., 2010. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on EC'10, June 7–11*. Cambridge, Massachusetts, USA, pp. 209–218. Available at: http://dl.acm.org/citation.cfm?id=1807376 [Accessed September 26, 2013].

Kosinski, M. & Bachrach, Y., 2012. Crowd IQ: Measuring the intelligence of crowdsourcing platforms. In *WebSci '12 Proceedings of the 3rd Annual ACM Web Science Conference, NY, USA*. pp. 151–160. Available at: http://dl.acm.org/citation.cfm?id=2380739 [Accessed June 22, 2013].

Little, G. et al., 2009. TurKit : Tools for Iterative Tasks on Mechanical Turk. *IEEE Symposium on Visual Languages and HumanCentric Computing VLHCC*, (Figure 1), pp.252–253.

Nomura, J. et al., 2001. Virtual reality technology and its industrial applications. *Annual Reviews in Control*, 25, pp.99–109.

Paolacci, G., Chandler, J. & Ipeirotis, P., 2010. Running experiments on amazon mechanical turk. *Judgment and Decision Making*, 5(5), pp.411–419. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1626226 [Accessed June 22, 2013].

Wang, A., Hoang, C.D.V. & Kan, M.-Y., 2012. Perspectives on crowdsourcing annotations for natural language processing. *Language Resources and Evaluation*, 47(1), pp.9–31. Available at: http://link.springer.com/10.1007/s10579-012-9176-1 [Accessed June 19, 2013].

Wu, H., Corney, J. & Grant, M., 2014a. Crowdsourcing Measures Of Design Quality. In *34th Computers and Information in Engineering Conference (CIE)*.

Wu, H., Corney, J. & Grant, M., 2014b. Relationship between quality and payment in crowdsourced design. In *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. Ieee, pp. 499–504. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6846895.

Yu, L. & Nickerson, J., 2011. Cooks or cobblers?: crowd creativity through combination. *CHI 2011, May 7–12, Vancouver, BC, Canada*. Available at: http://dl.acm.org/citation.cfm?id=1979147 [Accessed June 25, 2013].

Yu, L. & Nickerson, J., Generating creative ideas through crowds: An experimental study of combination. In *Thirty Second International Conference on Information Systems, Shanghai 2011*. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2039547 [Accessed June 25, 2013].

CHAPTER 6

CDESIGN FRAMEWORK

I. INTRODUCTION

In chapters 4 and 5, the research questions about the relationship between payment to crowd workers, the evaluation process of the quality of design, and the feasibility of 3D crowdsourced design tasks were discussed. This process resulted in the identification of two of the objectives of this research. In this chapter the findings from the two experimental cases are used to systematically generate a crowdsourced design framework (cDesign Framework). Despite its apparent diversity the process of mechanical design has been formalized by models such as Pugh's "Total Design" (Pugh 1991) or Pahl and Beitz's method (Pahl et al. 2007). These models of the design process provide a reference framework which enumerate the critical steps and allow previously "ad hoc" activities to be structured and managed. Similarly, the cDesign model presented in this section is a synthesis of reported academic and commercial work and is motivated by the desire to provide a generic structure for the process of creating crowdsourced design tasks. The structure of this chapter is as follows:

In the first section, the triangulation of the case studies is presented to induce the key elements and stages of the cDesign framework. A coding method is used to indicate the most relevant and frequent words/phrases which can define a general framework structure from the specific experimental cases in the previous chapters. Furthermore, a method called 'cross experimental cases comparisons' is applied, which can compare all findings side by side in the analysis sub-section. The conclusion of the resulting analysis of the experimental cases is the crowdsourced design framework.

In the second section, the application of the cDesign Framework is illustrated. Based on the results from the first section, the structure and the detailed elements of the framework are fixed. This systematic framework standardises the crowdsourced design process and its application.

In other words, the purpose of this chapter is to detail how the results reported in previous chapters have been triangulated into a number of general classes, or categories of activities.

In the cDesign Framework, these categories form the component elements which define the process of crowdsourced design task ranging from the design task preparation to the evaluation of design quality, and ultimate the results assessment. The next section starts with the cross case analysis of the experimental cases' results.

II. CROSS CASE ANALYSIS

In this section, all findings from the experimental cases will be compared and be matrixed by a qualitative research method named cross case analysis. The cross case analysis is defined as:

"Cross-case analysis is a research method that facilitates the comparison of commonalities and difference in the events, activities, and processes that are the units of analyses in case studies." (Khan & Van Wynsberghe 2008)

Figure 6 - 1 shows the guide for the following research work. The research findings will be integrated into a triangulation which will lead to a conclusion, followed by the framework establishment. In the triangulation, experiment results are compared and then used to define into a common language for the crowdsourced design study.

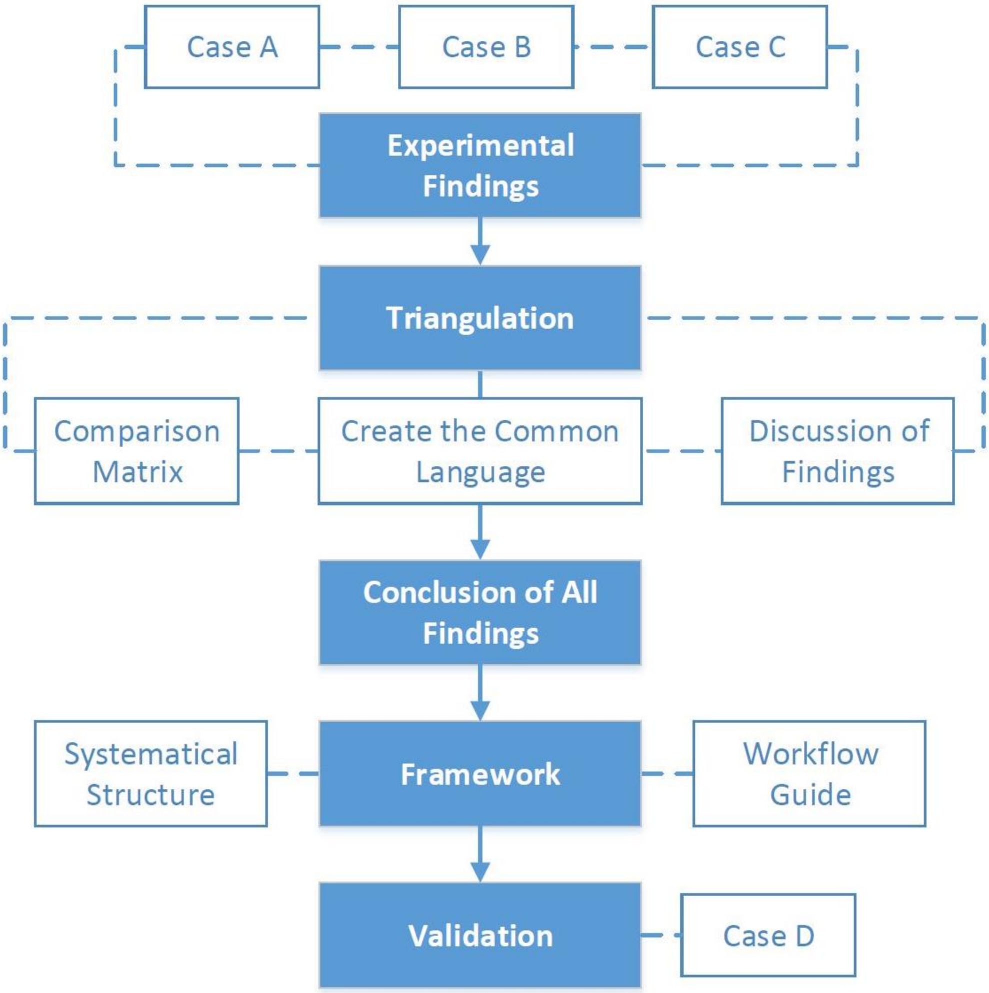


Figure 6 - 1 Integration and triangulation of research findings in case studies

III. COMPARISONS OF EXPERIMENTAL CASES FINDINGS

To compare the findings of each case, the objectives of the experiments are reviewed to understand which key elements need to be compared. The purposes of each experimental case are as follows:

3.1 Experimental Case One: Desk Lamp Design Task

Purpose: to understand the mechanism of crowdsourced design, and to investigate whether the reported crowdsourced design platforms/tools can be applied in different design tasks; to self-approve the reported crowdsource design process from the reported work (Yu & Nickerson 2011)(Yu 2011).

3.2 Experimental case two: living room layout design task

Purpose: to investigate the relationship between the payment to the crowd workers and the quality of the design task output; to test the performance of the crowd worker in the crowdsourced design process on the commercial platform; to investigate the potential design tool, experimental method, etc. in crowdsourced design process.

3.3 Experimental case three: 3D kitchen room layout design task

Purpose: to investigate whether 3D designs can be crowdsourced using similar crowdsourced design methods as previously reported; to investigate the potential design tools, and the methods for the design process; also to measure the performance of the crowd workers.

To integrate the experiment findings, from each experiment, all the key elements of the task were collected. The results of this process are shown in Table 6 - 1, the key elements comparisons are basically divided into three groups by the process of the experiment. Because from the cases, the basic process of the experiments is: firstly, to prepare the experiment. During this stage, requesters on the commercial crowdsourcing need to prepare the essential basis for the task and the workers. For example, decide the reward to workers, write the task description. The second step is to execute the task. In this stage, because the experiments in the last chapters are all carried by iterative process, the task execution (task design workflow) and the evaluation both happened in the same stage. So in Table 6 - 1 Key elements preparation, basically, the key elements identified will be divided into two parts: design task preparation and design task execution & evaluation.

Table 6 - 1 Key elements preparation

Key elements	Experimental Case 1: Desk Lamp Design Task	Experimental Case 2: Living Room Layout Design Task	Experimental Case 3: 3D Kitchen Room Layout Design Task
Design Task Preparation			
Task instruction writing	√	√	√
Task workflow design	√	√	√
Crowd selection	x	√	√
Payment strategy	x	√	√
Time strategy	x	√	√
Design tools selection	x	√	√
Cheats avoiding strategy	x	√	√
Results submission strategy	x	√	√

Key elements	Experimental Case 1: Desk Lamp Design Task	Experimental Case 2: Living Room Layout Design Task	Experimental Case 3: 3D Kitchen Room Layout Design Task
Design Task Preparation			
Design Specification	x	x	√
Design Evaluation Criteria	x	x	√
Task Prototyping – Task Test	√	√	√
Design Task Execution & Evaluation			
Human-based Genetic Algorithm	√	√	√
Design evaluation	√	√	√
Cheats avoiding strategy	x	√	√
Demographic information collection	√	√	√
Task delivery	x	√	√

Based on the common elements identified in Table 6 - 1, the general model is shown schematically in Figure 6 - 2 and consists of four main stages: Specification, Prototype, Execution and Evaluation. The framework is used in this paper to establish the context of the authors' investigations (rather than being, say, a provable optimum model for crowdsourced design). The following sections provide a qualitative description of each stage before the experimental work in support of the design evaluation process used in Stage 2, 3 and 4 is presented.

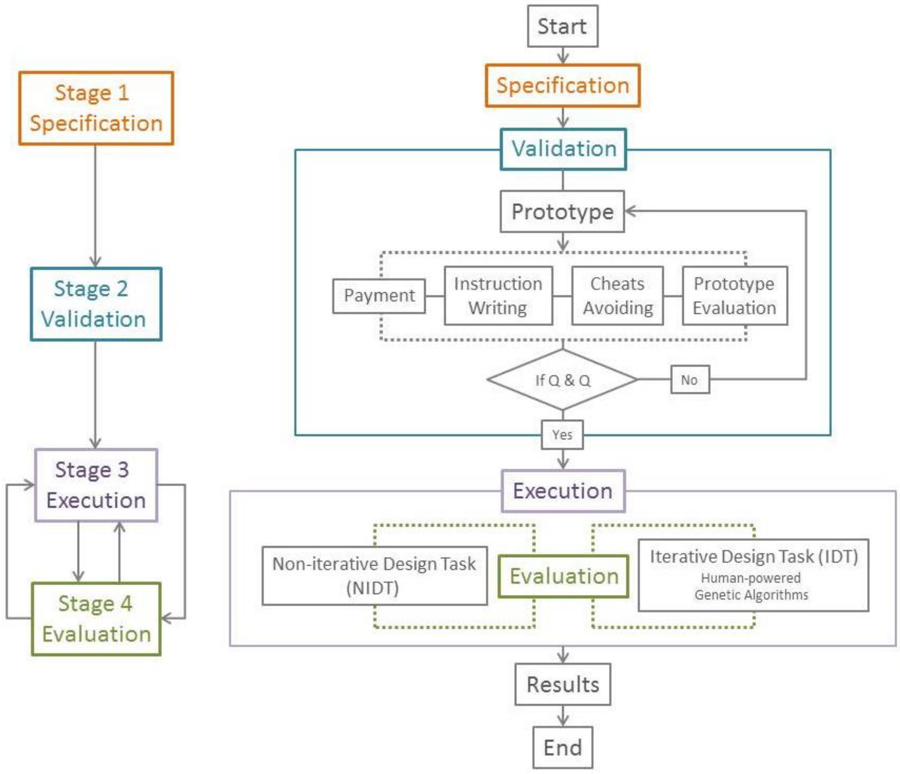


Figure 6 - 2 The cDesign methodology

IV. GENERAL CDESIGN FRAMEWORK AND DETAILED PROCESSES

Each of these stages can be expanded into a specific checklist of issues and options that must be addressed by the creators of crowdsourced design tasks. Table 6 - 2 illustrates the components of the Specification Stage.

Table 6 - 2 Specification Tasks

Stage 1 - Specification	
Issues	***Illustrative Options***
Platform Selection	Amazon's Mechanical Turk (MTurk)
	ShortTask
	Task China (Taskcn)
	……
Design Tool Selection	Cloud CAD Tool (i.e., Google Drawing)
	Specific CAD Package (Solidworks, Atudesk)
	Open (i.e., any format)
	……
"Crowd" Selection	Anyone, Anywhere
	Graphic Designer
	Engineer
Methodology Selection	Iteration
	Non-iteration
Design Workflow	File Transfering
	Data Access
	…

1. Stage 1

Every design task needs a crowdsourcing platform to host the process and the choice of crowdsourcing platform will reflect the nature of the task: some design work can be attempted by anyone regardless of education or background, whereas other tasks require specific experience or education. For example, Amazon's Mechanical Turk (MTurk) and ShortTask involve workers from all over the world. In contrast, some platforms are only for workers from one country, for example the Taskcn platform has workers mostly from China. After selection of the platform the choice of design representation and associated tool is the second most important step. Design tools need to be selected for workers as a consideration of the task itself (i.e., 2D design task – 2D design tools or 3D design task – 3D design tools). Having selected the platform and representation the skills of the "crowd" provided by a given platform needs to be considered (i.e., will the task be open to all are require specific expertise such as CAD experience. In parallel to the fundamental decisions on platform, tool and crowd, the methodology to be adopted in the execution process must also be determined at this initial stage. For example, the design task processes can be iteratively or non-iteratively executed. Finally, once the methodology is specified the design workflow needs to be discussed (i.e., results' file transfer, shared access to a representation held in the cloud, etc.).

2. Stage 2

Without prior experience of running similar tasks many of the choices made in the specification stage will be educated guesses whose effectiveness is uncertain. Stage 2 validates the choice made in Stage 1 by trialing prototype versions of the task. There are 6 implementation decisions (identified in Table 6 - 3) that need to be specified and validated in Stage 2: the payment for participants (per person per task); time to undertake the task; clarity of the task instruction; results submission method and the manner in which workers who attempt to scam, or cheat, the system should be decided.

Table 6 - 3 Validation Tasks

Stage 2 – Prototype Validation	
Issues	***Illustrative Strategies***
Payment	Payment Strategy (flat rate or bonus)
Time	Time Strategy (how long to do the task?)
Results Submission	The required format for file submission
Cheats Avoiding Avoidance	Qualication task (before participants accept the prototype task)
	Objective/Subjective questions and answers
Task Instruction Writing	Written instructions
	Illustrration of typical outputs
Evaluation for Prototype Results	How are the results judged?

The design of the crowdsourced task is refined through the process of prototype testing until the require Quantity and Quality (Q & Q) of results are being produced. At which point the process moves to the Execution stage.

3. Stage 3 & 4

Execution is essential a scaling up of the task for presentation to a larger crowd. The length of the execution stage will be determined by the method set in Stage 1. A competition might last many weeks whereas an HPGA will often cycle through generations of design every few days. So a crowdsourced design task could be characterised by the nature of the designing process as either Iterative Design Tasks (IDT) or Non-iterative Design Tasks (NIDT). Regardless of the mechanism used, the process ends with Evaluation task which reviews the crowd’s work and select the best outputs.

At both the validation and execution stages the ability to accurately evaluate designs is crucial to tasks such as the setting of payment levels (Stage 2) or selecting the best design for iterative improvement (Stage 3). Design Execution and Design Evaluation are regarded as separate stages because although tightly coupled they are distinct activities which occur sequentially (i.e., the process alternates between designing and evaluating). The next section describes an experiment, in terms of the cDesign framework, that was created to investigate the effectiveness of two different approaches to Crowdsourced design assessment.

4. Stage 3&4, Non-iterative Design Task (NIDT)

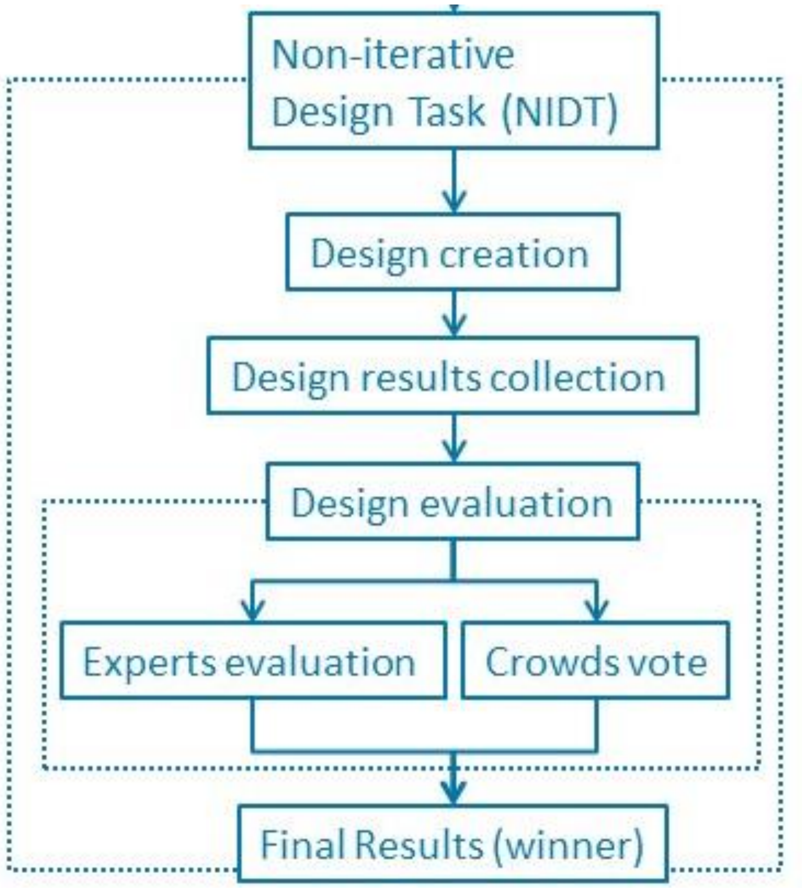


Figure 6 - 3 The workflow and detailed process for Non-iterative Design Task (NIDT)

As described in chapter 2, on a large number of crowdsourcing sites, the design tasks are categorised into the NIDT (the basic NIDT workflow is shown in Figure 6 - 3). Generally, problem requesters post tasks on the platform, then participants take part in the tasks and submit their results. If the results are approved by requesters (or mostly voted by crowd themselves), the winner will get rewarded. The basic steps for NIDT are:

1) Requesters who have problems set up design tasks on the platform
2) Crowd workers solve problems and submit their results
3) Requesters evaluate results or crowd votes
4) Winner participants get rewards (usually 'winners take all')

5. Stage 3&4, Iterative Design Task (IDT)

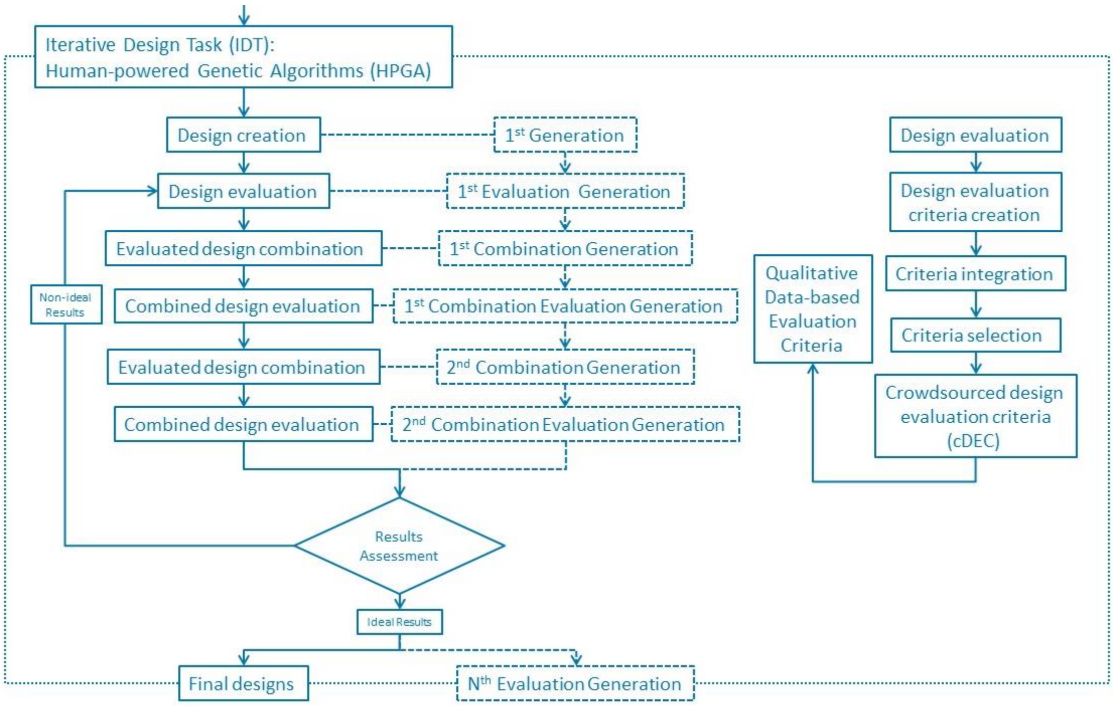


Figure 6 - 4 The workflow and detailed process for Iterative Design Task (IDT) by using HBGA (to create designs) and cDEC (to evaluate designs)

Except for the NIDT, IDT framework illustrates how design tasks are crowdsourced guided by the Human-based Genetic Algorithms (in Figure 6 - 4). Because usually in IDT, several generations of designs will be created, evaluated and combined, the basic stages of IDT are:

1) design creation: crowd creates the first generation designs;
2) design evaluation: crowd evaluates designs from the previous generation so that provides pairs of designs to create the following generations;

3) design combination: crowd combines the pairs of design results from the design evaluation for the new generation of designs.

The following paragraphs describe the details of the stages in IDT.

Iterative Design Tasks Process

In IDT, the first stage is to create the first generation of designs. During this stage, participants submit their design results based on the requesters' design requirements. This stage is similar with the first stage of NIDT. The first generation of designs in the previous chapters usually needs a large number of design results to ensure the quantity of designs, which can influence the quality of designs in the following stages.

In the second stage, crowd workers need to evaluate designs in the most recent, or previous generation by the Crowdsourced Design Evaluation Criteria (cDEC). In cases B and C, the evaluation criteria were created by crowd as well. Participants used 7-pointed likert scale to evaluate designs and also rank designs. Based on the task requirements, the high-ranked designs will be used to generate the next generation.

In the design combination stage, participants combine the best features of the chosen designs from the evaluation stage. Because design combination can bring better quality of designs (Yu & Nickerson 2011) (Nickerson et al. 2011)(Wu et al. 2014)(Hao; Wu et al. 2015)(Hao Wu et al. 2015), the purpose of this stage is to generate the better quality of designs by crowd design combination.

V. CHAPTER SUMMARY

In this chapter, the creation and the illustration of Crowdsourced Design (cDesign) Model are described. The author firstly presents the workflow illustrating how cDesign is generated. Than the cDesign framework and the detailed stages are discussed. This cDesign model results from the cases A, B & C, and can be used in the general design tasks. As a conclusion, there are four stages in the cDesign model: stage 1, Specification; stage 2, Validation; stage 3, Execution and stage 4, Evaluation. The next chapter will validate this crowdsourced design model by a new design task.

References

Khan, S. & Van Wynsberghe, R., 2008. Cultivating the Under-Mined: Cross-Case Analysis as Knowledge Mobilization. *Forum: Qualitative Social Research*, 9(No.1), p.Art. 34. Available at: http://www.qualitative-research.net/index.php/fqs/article/view/334/729#footnoteanchor_1.

Nickerson, J., Sakamoto, Y. & Yu, L., 2011. Structures for creativity: The crowdsourcing of design. *… Workshop on Crowdsourcing …*, pp.1–4. Available at: http://crowdresearch.org/chi2011-workshop/papers/nickerson.pdf [Accessed June 22, 2013].

Pahl, G. et al., 2007. *Engineering design: a systematic approach*, Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0 [Accessed October 3, 2014].

Pugh, S., 1991. *Total Design: integrated methods for successful product engineering*, Addison-Wesley Pub (Sd).

Wu, H., Corney, J. & Grant, M., 2015. An evaluation methodology for crowdsourced design. *Advanced Engineering Informatics*, 29(4), pp.775–786. Available at: http://dx.doi.org/10.1016/j.aei.2015.09.005.

Wu, H., Corney, J. & Grant, M., 2014. Relationship between quality and payment in crowdsourced design. In *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. Ieee, pp. 499–504. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6846895.

Wu, H., Corney, J. & Grant, M., 2015. THE APPLICATION OF CROWDSOURCING FOR 3D INTERIOR LAYOUT DESIGN. In C. Weber et al., eds. *Proceedings of the 20th International Conference on Engineering Design (ICED 15)*. pp. 123–134. Available at: https://www.designsociety.org/publication/37769/the_application_of_crowdsourcing_for_3d_interior_layout_design.

Yu, L., 2011. Crowd creativity through combination. *Proceedings of the 8th ACM conference on Creativity and cognition - C&C '11*, p.471. Available at: http://dl.acm.org/citation.cfm?doid=2069618.2069756.

Yu, L. & Nickerson, J., 2011. Cooks or cobblers?: crowd creativity through combination. *CHI 2011, May 7–12, Vancouver, BC, Canada*. Available at: http://dl.acm.org/citation.cfm?id=1979147 [Accessed June 25, 2013].

CHAPTER 7

VALIDATION OF THE CDESIGN FRAMEWORK

CASE STUDY D:

CAR KEY FOB DESIGN EXPERIMENT

I. INTRODUCTION

The cDesign framework was established by triangulation of three case studies (i.e., desk lamp design task, living room layout design task and 3D kitchen layout design task) after which chapter 6 discussed in detail of the structure and the mechanism of the general framework. This chapter validates the cDesign framework using a different design experiment (i.e., car key fob) to assess its completeness and functionality.

A car key fob or, called a keyless entry system (remote keyless system) is “an electronic lock that controls access to a building or vehicle without using a traditional mechanical key”. The reasons why car key fob design was chosen as a topic for the framework validation experiment are as follows:

a. At a conceptual level, car key fobs can be designed based largely on user experience rather than requiring professional engineering knowledge.
b. A good car key fob design will combine functionality and aesthetic properties.
c. Because the main purpose of this experiment is to validate the iterative process of the cDesign framework, the car key fob design task can fit the Human-based Genetic Algorithm.

This chapter is divided into the following sections:

Section 1, the methodology and tools of the experiments

Section 2, the description of the car key fob experiment

a. The generation 1 creation task
b. The evaluation of the generation 1 drawings
c. The combination of the top 3 ranked drawings – generation 2
d. The evaluation of the generation 2 drawings
e. The combination of the top 2 drawings of the generation 2

Section 3, results of the experiments

a. Generation 1 drawings
b. Evaluation results for the generation 1 drawings

c. Combination of the top 3 generation 1 drawings – generation 2

d. Evaluate generation 2 drawings

e. Combine Best 2 or 3 drawings from generation 1 – generation 3

f. Evaluation of generation 3 – the best generation 3 drawing

Section 4, discussions of the experiment

Section 5, conclusion

II. METHODOLOGY AND TOOLS

Given a design brief to "produce a conceptual design of car key fob that incorporate the most popular features (i.e., with the best features of a car key fob)", the cDesign framework was used to create a process to deliver this design. The process of the experiment will be followed by each step of the framework till the results.

In the living room layout design experiment, participants created designs by the Google Drawing ™ online tool, and in the 3D kitchen room layout experiment, a public 3D tool was used. As for this car key fob design experiment, it was decided that participants can draw car key sketches by free-hand drawing. The reasons are discussed as follows. Firstly, although Google Drawing ™ has been accepted by the design academia as an effective tool to create designs, somehow it limits the thoughts of participants. For example, there are several excited tool bars including different polygons and some stickers, once participants consider that those functions could decrease their workload, the new creative shapes may under covered by that behaviour. Besides, it spends time for participants to learn a new tool during underrating tasks. On the contrary, free-hand drawing can show ideas without limits, and does not spend extra time to learn a new tool.

III. CAR KEY FOB DESIGN EXPERIMENT

Similarly, the car key fob design tasks are basically divided into two parts: drawings creating task and drawings evaluation task. Following the IDT (Iterate Design Task) framework, creating and evaluation tasks are set alternately till the results. The basic workflow of the experiment is shown in Figure 7 - 1. The first stage is to create generation 1 designs.

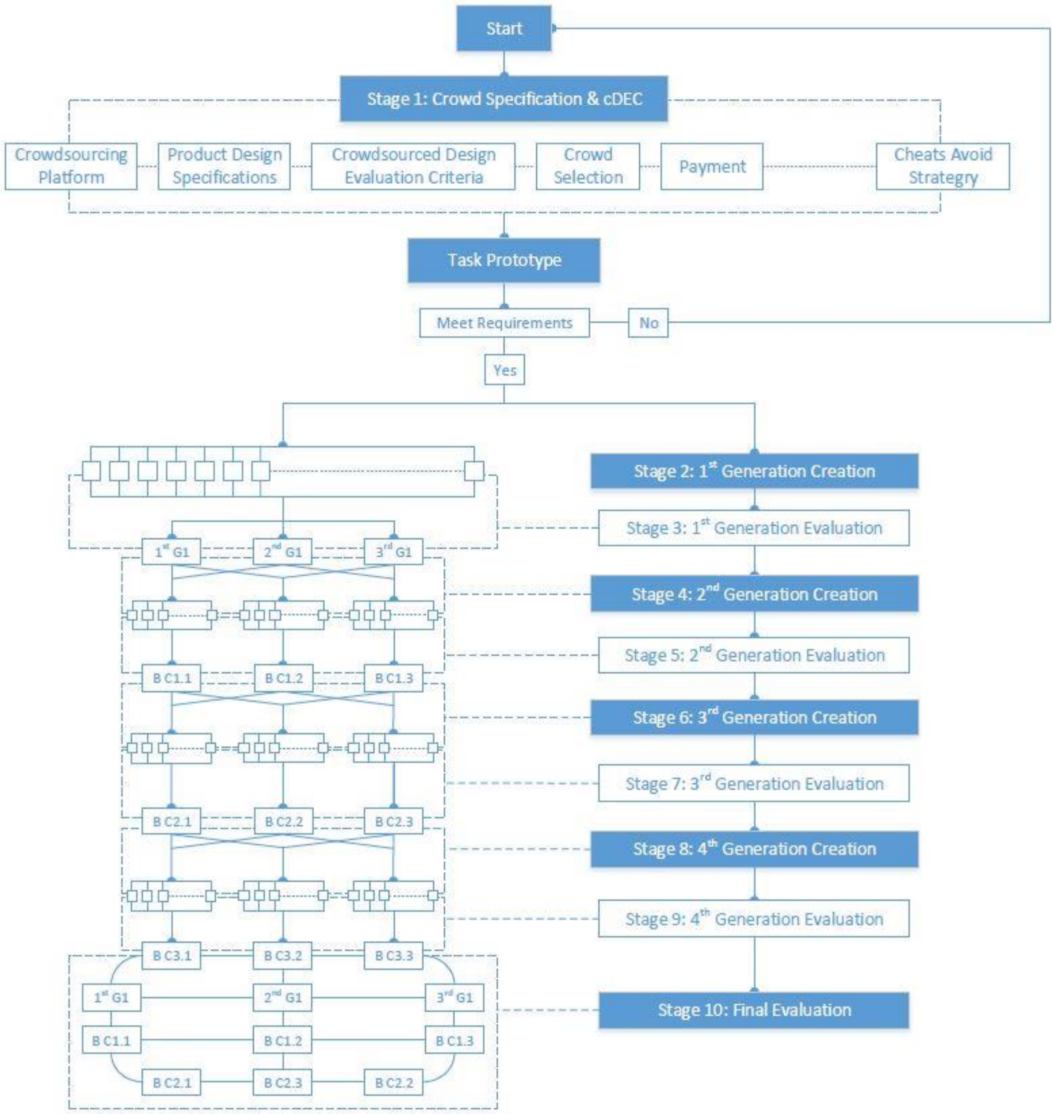


Figure 7 - 1 The workflow of the car key fob design experiment

Generation 1 creating drawings task

The first design task of the experiment was posted on mTurk as: car key fob drawing task. In the task, participants were required to generate a car key fob drawing by free-hand drawing. However, before the first task, a design specification/evaluation criteria creation task was given to the participants. As described in the IDT framework, the criteria throughout the creation and evaluation processes.

Design specification/evaluation

To create the evaluation criteria (that will also be used as the design specification), participants are to provide answers for the following questionnaire (questions were put to the MTurk crowd workers):

Case Study D

Design specification/evaluation criteria collection task

1. Could you please suggest 5 features, or functions, that a remote car key fob should have to be suitable for elderly users/drivers?
2. Could you please suggest 3 further features, or additional functions, that a remote car key fob should have?

MTurk Task D - 01

Then after collecting evaluation/specification criteria, participants were asked to design a car key fob which meets the specifications.

1st generation creation task

In terms of the design task of the experiment, participants are required to draw the car key fob design by free-hand drawing, and then copy (i.e., take a picture or scan) their drawings to the requester. During the drawing, the specifications are illustrated to participants, they are required to follow the specifications to show their idea. The task's instruction is shown below:

Case Study D

1st generation design task description

Please draw (by free-hand drawing) a labelled sketch of a car key fob which incorporates controls for a specific list of functions. Then upload an electronic version (e.g., scanned, jpeg, photo) of your drawing as your final submission.

Process of your participation:

1. Do a free-hand drawing of a car key fob with the following specifications which must including at least controls to:

1) doors lock and unlock

2) engine start

3) active car alarm

4) GPS based function to locate your car

5) Display of car information (e.g., fuel level, temperature)

2. Label the functions in your sketch of the car key fob
3. Take a photo, or scan your sketch into a jpeg format.
4. Name your photo/scanned drawing with your MTurk worker ID number.
5. Upload your electronic sketch

Please Note:

1. Only Free-hand drawing is permitted.
2. Any CHEAT WORK will be REJECTED, and you MTurk account will be flagged as well.

Additionally, please provide some of your personal information by the following questions

1. Gender
2. Age
3. Nationality
4. Education Background (i.e., Bachelor's Degree)
5. Do you have any design experience before (i.e., no experience, one-year experience)
6. Do you drive

Tips: You can add your answers and your drawings into a Microsoft Word document. Or just write down your personal information beside your drawing.

MTurk Task D - 02

During this task, because in the author's previous living room layout design task, to investigate the relationship between payment and design quality, different payment levels were set. In the car key fob design task, different payment levels were established as well: \$0.35, \$0.50, \$0.75 and \$1.00.

1st generation drawings evaluation task

In cDesign framework, after the design creation task, all generation 1 designs need to be evaluated by crowd under the cDEC (crowdsourced design evaluation criteria). Participants evaluate drawings from the generation 1 by the cDEC collected from the crowd. Firstly, a 7-Point Likert Scale is provided to participants to rate drawings from a range of 1 (worst drawings) to 7 (best drawing). Secondly, based on their rating scores, drawings need to be ranked in the group. Thirdly, participants are required to provide reasons for the top three rankings. Any step missed in their submissions, the results will be rejected. In the same time, the demographic information is collected as well. Based on the evaluation results of the 1st generation car key fob drawings, the top three drawings will be combined by their best features to generate the next generation designs following the cDesign framework.

2nd generation creation task – combination design task

In the 2nd generation car key fob creation task, participants need to combine the best features from the 1st generation drawings based on the ranking results. In the cDesign framework, following the Human-based Genetic Algorithm, new generation comes from its last generation by human (i.e., crowd). The task instruction is shown as follows (the demographic information is required as well for the participants):

Case Study D

2nd generation creation task of car key fob designs (combination task)

1. Please draw (by free-hand drawing) a labelled sketch of a car key fob which combines the best features from the two car key fobs illustrated below, and incorporates controls for a specific list of functions given below.
2. Then upload an electronic version (e.g., scanned, jpeg, photo) of your drawing as your final submission.

The link for a high resolution version of the two car key fobs: (the link if shown here.)

Process of your participation:

1. Do a free-hand drawing of a car key fob combining the best features from the two car key fob drawings shown below.

2. During your sketching/drawing, you are required to follow the specifications which must include at least controls to (but not limited to):

1) doors lock and unlock
2) engine start
3) active car alarm
4) GPS based function to locate your car
5) Display of car information (e.g., fuel level, temperature)

2. Label the functions in your sketch of the car key fob
3. Take a photo, or scan your sketch into a jpeg format.
4. Name your photo/scanned drawing with your MTurk worker ID number.
5. Upload your electronic sketch

MTurk Task D - 03

In the drawing combination tasks, any completely new drawings will be rejected. Because the combination task requires participants to combine the best features from the given pair of drawings, the 'gene' of the 1st generation should be clearly visible in the 2nd generation. The best three drawings from the 1st generation designs are combined in this task, so in total, three groups (C_3^2 = 3) of combination drawings are required, and for each group, ten drawings need to be collected:

Group 1: 1st ranking drawing and 2nd ranking drawing: ten results

Group 2: 1st ranking drawing and 3rd ranking drawing: ten results

Group 3: 2nd ranking drawing and 3rd ranking drawing: ten results

Consequently, the 2nd generation drawing creation task (combination task) generates thirty combined drawings which will be evaluated in the next stage – the 2nd generation evaluation task.

2nd generation evaluation task (same as the 1st generation evaluation)

3rd generation creation task (same as the 2nd generation creation)

3rd generation evaluation task (same as the 1st generation evaluation)

IV. EXPERIMENT RESULTS

1. 1st Generation Creation Drawings

In the 1st generation drawing task, a hundred and seventy approved drawings are collected from the mTurk (examples are shown below Figure 7 - 2, all 1st generation car key fob designs are shown in Appendix 2 - 1, Appendix 2 - 2, Appendix 2 - 3, Appendix 2 - 4, Appendix 2 - 5 and Appendix 2 - 6). The $0.35 payment group received eight drawings, $0.50 payment group collected 38 drawings, $0.75 payment group was submitted 48 drawings and in $1.00 payment group, there were 77 drawings. Not surprisingly, higher payment results in larger numbers of submissions, but in order to assess "value of money", it is necessary to determine the design quality.

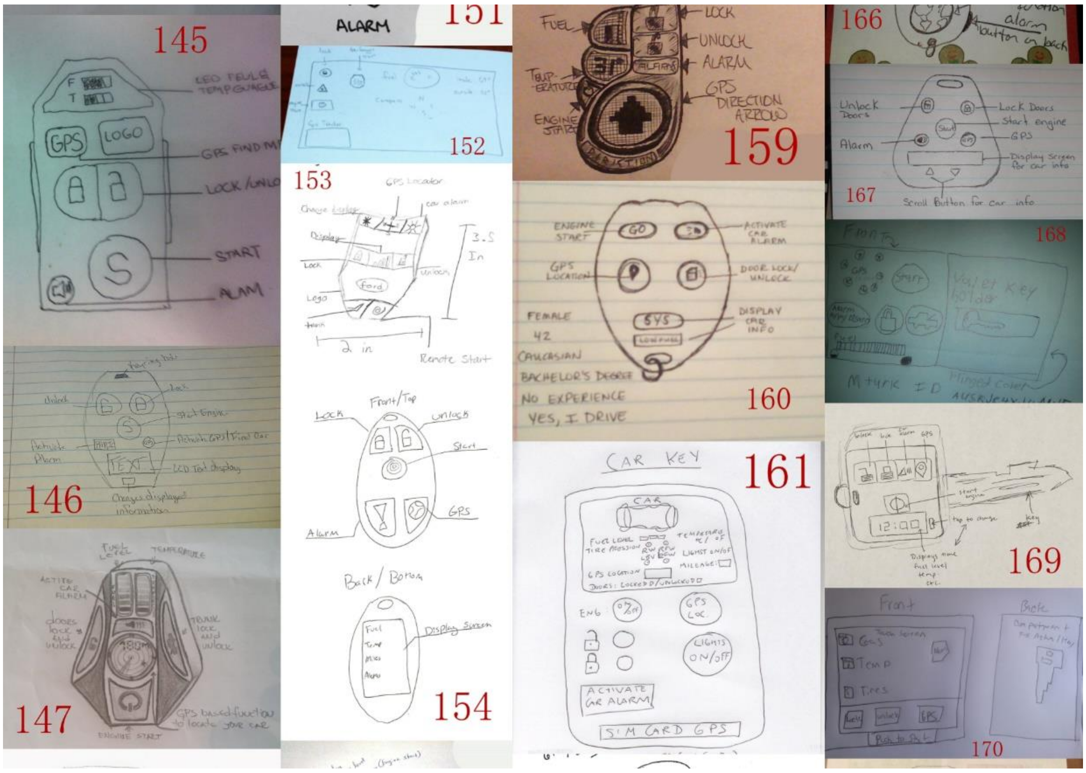


Figure 7 - 2 Examples of the 1st generation drawings

2. 1st Generation Evaluation Task

After collecting all 1st generation drawings, the sketches were separated into seventeen groups each having ten drawings. This evaluation task description is shown as follows:

Case Study D

Car Key Fob Design Evaluation Task

In this task, you need to:

Step 1, evaluate 19 drawings shown below (No. 50, No. 159, No. 95, No. 17, No. 147, No. 78, No. 89, No. 61, No. 73, No. 68, No. 130, No. 163, No. 107, No. 51, No. 36, No. 142, No. 28, No. 64 and No. 82) by the 7-Point Likert Scale illustrated below. You need to give a score to each drawing from the range of 1 (the worst drawing) to 7 (the best drawing). The dark-red colour figure in each drawing is the number of different car key fob drawings. You can download the high resolution of those 10 designs by the link:
https://onedrive.live.com/redir?resid=52BF98E2FB838931!656&authkey=!ALyEbh1LbQvm1po&v=3&ithint=photo%2cjpg

7-Point Likert Scale

Please evaluate the car key fob sketches by the following evaluation criteria:

The "quality" and "appropriateness" of the functions for:

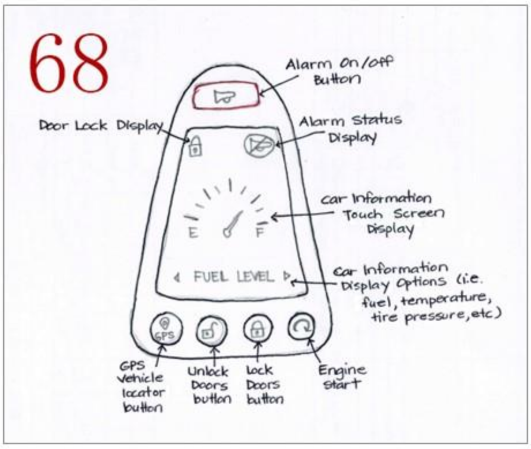
1) Locking and unlocking the doors

2) Starting the engine

3) Activating the car alarm

4) GPS based location of the car

5) Display of car information (e.g., fuel level, temperature)



Rubbish/Poor	○1	○2	○3	○4	○5	○6	○7	Excellent/Best

MTurk Task D - 04

Step 2, please rank all 19 designs from the 1st place (the highest) to the 19th place (the lowest).
Step 3, please give some reasons for your top 3 ranked drawings. (i.e., how the drawings match the evaluation criteria, why you rank those drawings as the best 3, etc.)

Submission template

(JUST AN EXAMPLE, not the real rating and ranking results):

Ranking	Design No.	Score
1	No. 50	7
2	No. 159	7
3	No. 95	7
4	No. 17	6
5	No. 147	6
6	No. 78	6
7	No. 89	6
8	No. 61	5
9	No. 73	5
10	No. 68	5
11	No. 130	5
12	No. 163	4
13	No. 107	4
14	No. 51	4
15	No. 36	3
16	No. 142	3
17	No. 28	2
18	No. 64	2
19	No. 82	1

You can copy the template to evaluate and rank the designs.

Your work is ONLY permitted by uploading a text or Microsoft Word document as your submission. Please Note:
1. Cheating work will influence your reputation in MTurk (i.e., requester would lock your worker ID once your solution is proved as cheating).
2. If you do not provide reasons for the best 3 layouts, you will NOT be paid.
Besides, would you please provide the following information attached in your design result, which could be very helpful.
1. Your gender 2. Your age 3. Your Nationality
4. Your education level, i.e., Bachelor Degree
5. Do you have any design experience? i.e., one year design experience
6. Do you drive?

MTurk Task D - 04

Crowd workers rated each drawing by the 7-Point Likert Scale measurement, and then ranked ten drawings for each group. At the end of the process all the designs will have scores as well as rankings in the groups. Because only the top three ranked drawings will be combined to create the 2nd generation car key fobs, the 1st place ranking drawings in each group will be integrated into one group (seventeen drawings) posted as a new evaluation task on mTurk. Turkers evaluated them by the same method judging all 1st generation drawings. Consequently, the top three ranking drawings are: No. 78 ranked in the 1st place followed by No. 64 as the 2nd, and No. 147 and No. 17 are tied in the 3rd place. As described in the last section that only three drawing will be combined as the 2nd generation, the tied drawings No. 147 and No. 17 were evaluated by the same method again. Finally, No. 147 won more votes than No. 17. Consequently, No. 78, No. 64 and No. 147 will be the 'pairs' of the next generation drawings. Examples of the design ranking forms are shown in Figure 7 - 3 and Figure 7 - 4.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Group No.	Drawing No.	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6	Rank7	Rank8	Rank9	Rank10	Average Ranking		Average Ranking
1	105	10	10	10	10	9	1	8	8	10	10	9.375	1	10
	1	2	5	4	3	10	10	9	5	1	8	5.75	5	6
	37	3	7	1	6	7	5	6	7	6	5	5.625	6	5
	104	7	8	9	9	2	6	10	1	2	6	6.125	4	7
	69	5	6	5	4	6	7	2	4	3	3	4.5	8	3
	121	8	3	2	8	8	8	4	9	7	9	6.875	2	9
	14	9	1	6	2	4	2	1	10	5	2	3.875	9	2
	42	6	2	8	7	5	4	7	2	4	4	4.875	7	4
	28	1	4	3	1	1	3	3	3	8	1	2.375	10	1
	23	4	9	7	5	3	9	5	6	9	7	6.5	3	8
2	9	10	10	10	10	8	10	8	10	5	8	9.25	1	10
	99	6	6	7	2	2	5	6	8	1	10	5.25	5	6
	64	3	5	3	3	7	2	1	7	3	1	3.375	10	1
	19	5	4	1	9	5	7	2	2	7	6	4.75	6	5
	138	9	7	2	5	10	8	7	6	2	5	6.125	3	8
	155	4	1	4	1	3	9	10	4	8	3	4.5	8	3
	156	2	3	8	8	4	1	5	5	9	9	5.5	4	7
	47	1	2	9	6	6	3	4	3	4	2	3.75	9	2
	101	7	8	6	7	9	4	9	9	10	7	7.75	2	9
	33	8	9	5	4	1	6	3	1	6	4	4.625	7	4
3	25	4	7	3	7	6	5	4	3	3	4	4.5	8	3
	22	3	4	5	6	4	4	5	1	4	5	4.25	9	2
	44	10	1	10	8	10	3	10	10	5	6	7.75	1	9

Figure 7 - 3 Example of design ranking form 1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Drawing No.	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9	Rank 10	Rank 11	Rank 12	Rank 13	Rank 14	Rank 15	Average Ranking		Final Ranking
50	8	5	4	6	7	8	5	16	6	1	17	11	15	9	12	8.615384615	13	7
159	4	12	11	2	3	4	4	18	5	10	5	19	8	3	19	8.153846154	15	5
95	3	9	12	4	12	3	8	15	2	19	19	9	3	10	16	9.461538462	10	10
17	1	1	13	3	13	5	9	9	19	17	15	4	7	1	13	8.083333333	16	3
147	9	16	15	9	2	9	1	3	18	2	3	6	12	11	18	8.083333333	16	3
78	17	8	14	5	1	1	3	1	4	8	4	13	19	12	11	7.5	19	1
89	5	11	2	19	5	10	11	10	3	12	18	7	14	14	1	9.384615385	11	9
61	7	3	16	1	10	6	6	17	7	6	7	14	16	5	6	8.384615385	14	6
73	11	14	18	13	11	11	10	4	14	13	14	3	2	4	17	10.69230769	7	13
68	10	4	19	12	14	13	19	14	11	16	11	17	11	13	2	12.69230769	2	18
130	13	19	3	18	16	14	7	19	8	9	8	10	17	15	14	12.92307692	1	19
163	2	13	5	11	4	18	2	13	9	11	1	5	13	18	8	8.769230769	12	8
107	19	10	8	17	9	15	12	2	15	3	2	16	18	19	9	11.76923077	5	15
51	16	17	9	7	6	2	15	5	16	18	10	15	10	16	5	11.30769231	6	14
36	12	2	10	10	8	19	16	8	17	7	12	8	9	17	3	10.53846154	8	12
142	14	18	17	8	15	12	14	11	13	15	6	18	4	7	4	11.84615385	3	16
28	6	7	6	14	19	16	13	7	10	14	16	2	6	6	15	10.46153846	9	11
64	15	6	7	15	18	7	18	6	1	5	9	1	5	2	7	7.923076923	18	2
82	18	15	1	16	17	17	17	12	12	4	13	12	1	8	10	11.84615385	3	16

Figure 7 - 4 Example of design tanking form 2

3. 2nd Generation Creation Task

Based on the evaluation results, the top three drawings were combined as Figure 7 - 5 illustrated below. For each combination, ten approved drawings were required from mTurk (i.e., for group one which were combined from No.78 and No. 64, ten approved drawings were required). In total, thirty new cat key fob sketches are collected as the 2nd generation (also called 1st generation combination designs – combination 1).

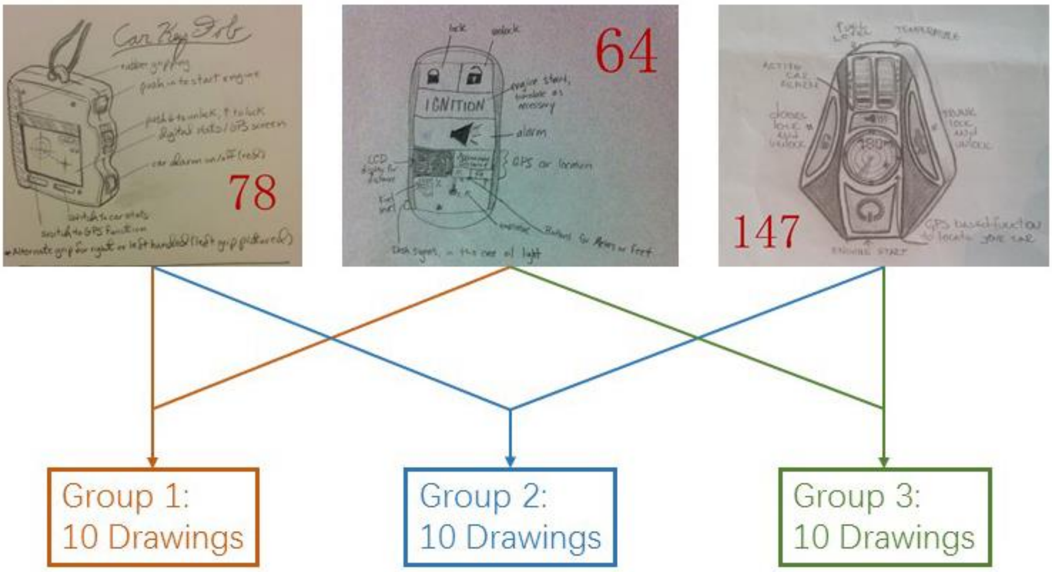


Figure 7 - 5 2nd generation creation task (the combination process using the best three designs from the 1st generation)

4. 2nd Generation Evaluation Task

After collected all approved 2nd generation sketches (in Figure 7 - 7, Figure 7 - 8 and Figure 7 - 9) (in the following figures, the number 'G2'or 'C1' of those designs are omitted during the evaluation process), designs were evaluated in each group individually by the same evaluation method applied in the 1st generation evaluation task. The best car key fob drawings will be combined to create the 3rd generation (2nd combination generation).

From the evaluation results (in Figure 7 - 6) it can be observed that the best drawings are G2.1.1 (C1.1.1), G2.2.8 (C1.2.8), G2.3.4 (C1.3.4) (in Figure 7 - 10). These best three drawings will be combined to generate the next generation.

Group No.	Drawing No.	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6	Rank7	Rank8	Rank9	Rank10	Average Ranking		Average Ranking
1	1.1	3	1	1	10	2	3	6	2	2	2	2.625	10	1
	1.2	1	5	10	6	8	10	8	4	5	5	6.375	3	8
	1.3	10	8	5	5	7	5	2	6	6	8	6.25	4	7
	1.4	4	7	7	3	3	8	7	5	4	9	5.625	5	6
	1.5	8	9	3	7	9	7	9	7	8	7	7.75	2	9
	1.6	2	4	8	1	5	2	5	8	1	1	3.5	8	3
	1.7	6	3	9	4	1	1	1	3	3	4	3.125	9	2
	1.8	5	10	6	9	10	9	10	9	9	10	9	1	10
	1.9	7	6	4	8	6	4	3	1	10	3	5.125	6	5
	1.01	9	2	2	2	4	6	4	10	7	6	5	7	4
2	2.1	7	7	8	10	2	6	3	4	5	7	5.875	7	4
	2.2	5	9	4	7	1	9	4	1	9	10	6	6	5
	2.3	8	4	10	5	8	4	6	6	6	6	6.125	5	6
	2.4	6	5	6	6	10	8	7	8	2	5	6.375	3	8
	2.5	4	1	5	1	6	3	5	7	1	2	3.375	9	2
	2.6	3	10	3	3	4	10	1	3	4	4	4.25	8	3
	2.7	10	2	9	9	7	5	9	10	8	3	7.5	1	10
	2.8	1	6	2	2	3	1	2	9	3	1	2.5	10	1
	2.9	2	3	7	4	9	7	8	5	7	9	6.25	4	7
	2.01	9	8	1	8	5	2	10	2	10	8	6.5	2	9
3	3.1	9	9	4	7	5	2	6	7	10	9	7	3	8
	3.2	5	8	6	10	10	8	8	9	8	10	8.375	2	9
	3.3	4	4	5	5	3	9	9	4	1	2	4.5	7	4
	3.4	1	2	2	1	1	7	2	3	2	1	1.75	10	1
	3.5	3	5	7	2	4	4	5	5	6	7	4.875	6	5
	3.6	6	7	8	6	8	6	7	8	4	5	6.625	4	7
	3.7	2	6	10	3	7	3	1	1	5	4	3.875	8	3
	3.8	8	3	3	9	6	5	3	2	7	6	5.125	5	6
	3.9	7	1	1	8	2	1	4	6	3	3	3.375	9	2
	3.01	10	10	9	4	9	10	10	10	9	8	9.375	1	10

Figure 7 - 6 Evaluation rankings results of the 2nd generation designs

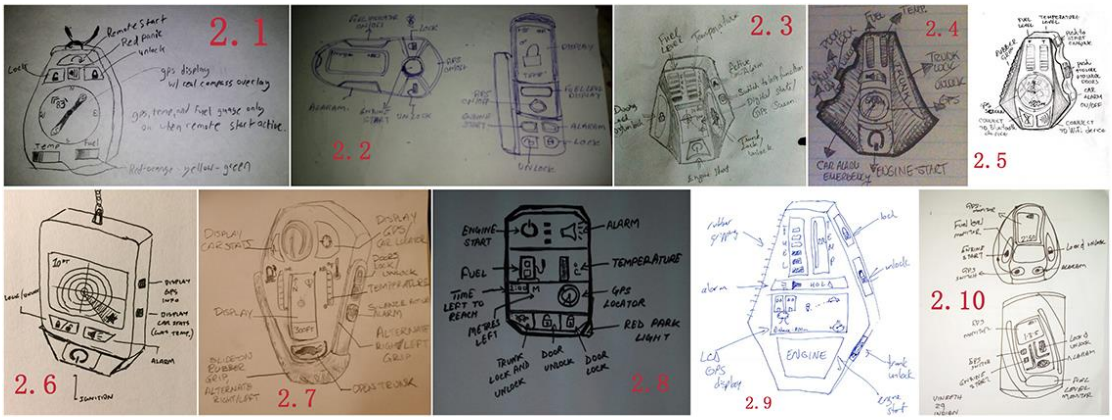


Figure 7 - 7 10 drawings of group 1 designs – 2nd generation, combination 1

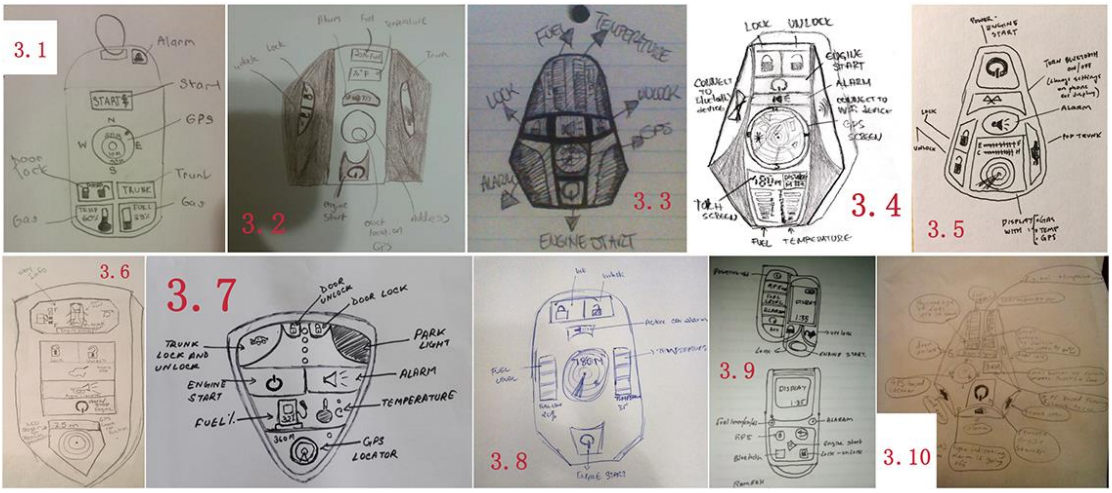


Figure 7 - 8 10 drawings of group 2 designs – 2nd generation, combination 1

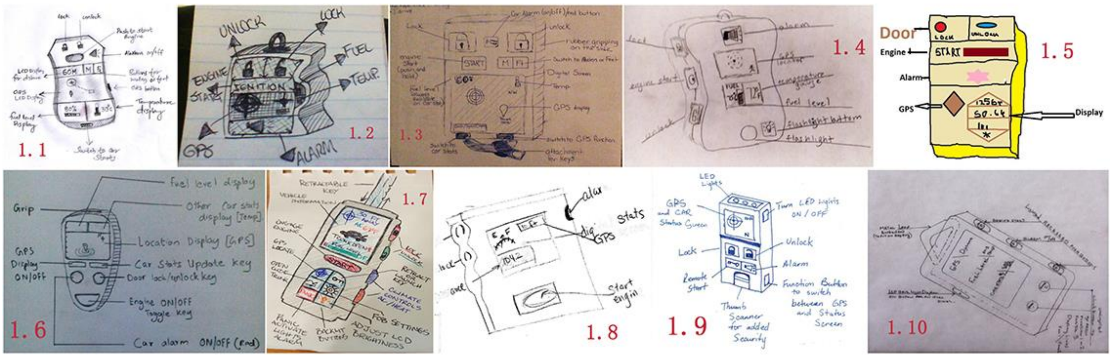


Figure 7 - 9 10 drawings of group 3 designs – 2nd generation, combination 1

Best 3 drawings in 2nd generation - combination 1

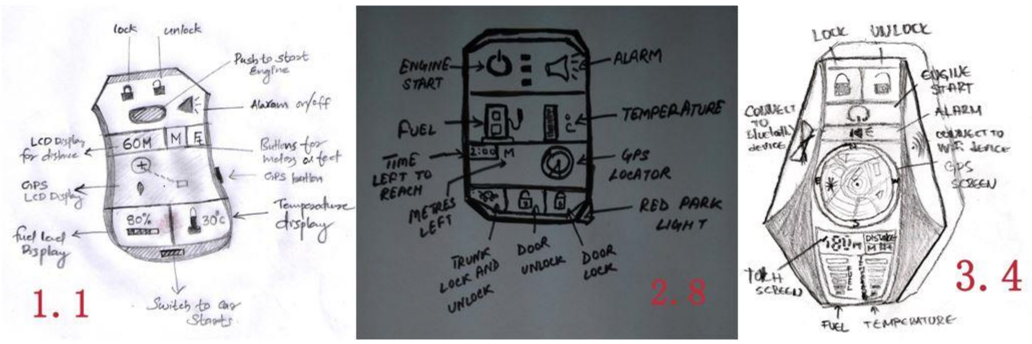


Figure 7 - 10 The best three designs from all 2nd generation car key fob designs

5. 3rd Generation Creation Task

As the same combination method for the 2nd generation drawings, the 3rd generation drawings were combined from the best three sketches evaluated from the 2nd generation (the combination process is shown in Figure 7 - 11). Also for each combination group, ten approved combined drawings were required. So in total, thirty 3rd generation drawings were created (in Figure 7 - 12, Figure 7 - 13 and Figure 7 - 14).

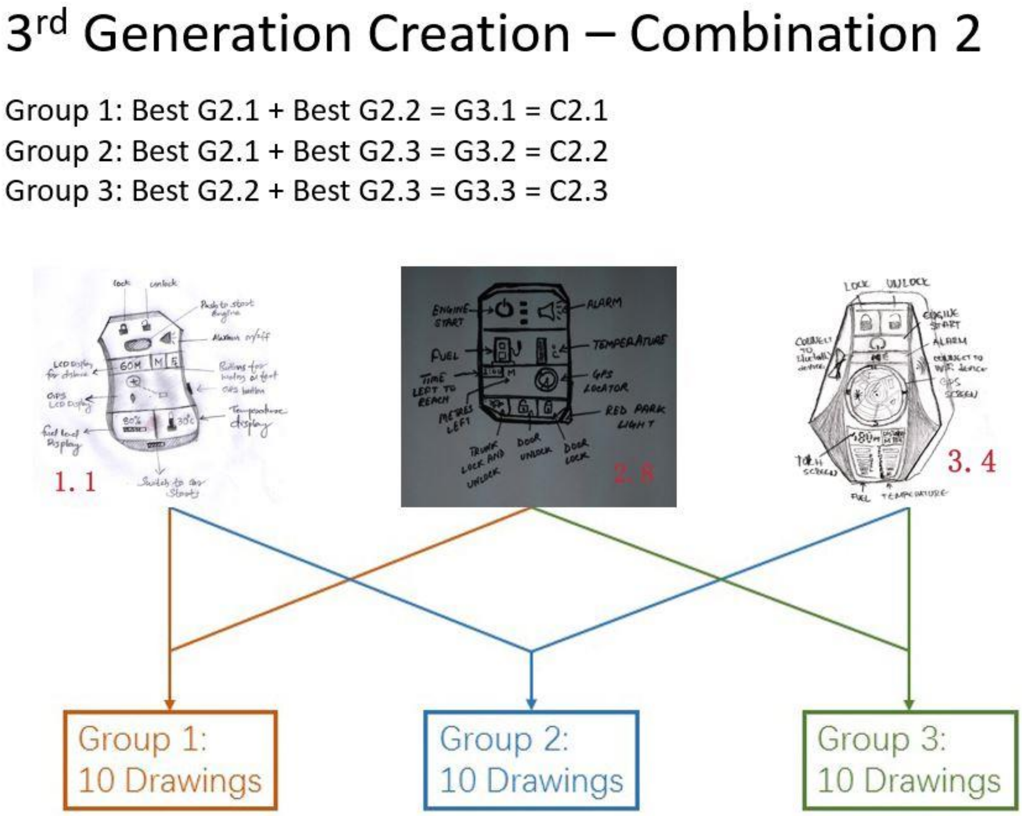


Figure 7 - 11 3rd generation creation task (the combination process using the best three designs from the 2nd generation)

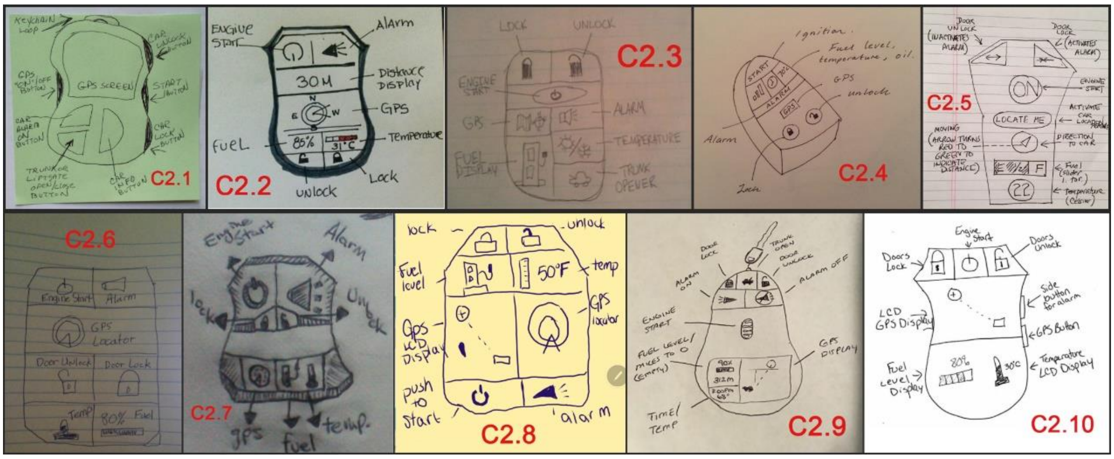


Figure 7 - 12 10 drawings of group 1 designs – 3rd generation, combination 2

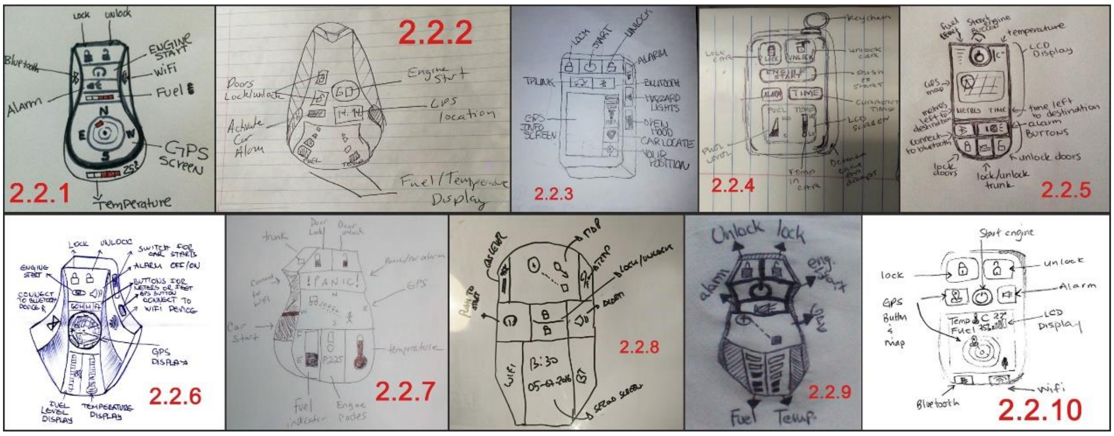


Figure 7 - 13 10 drawings of group 2 designs – 3rd generation, combination 2

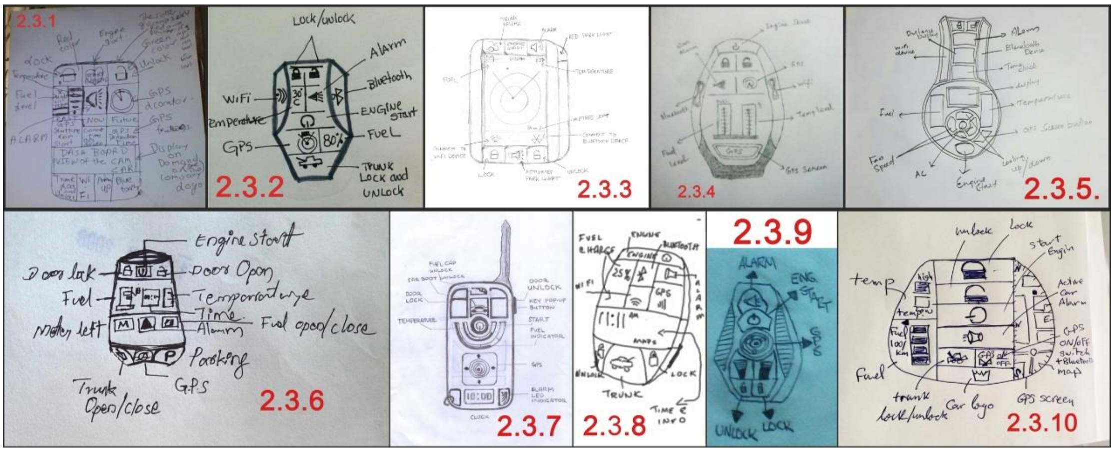


Figure 7 - 14 10 drawings of group 3 designs – 3rd generation, combination 2

6. 3rd Generation Evaluation Task

By the same evaluation process in the previous evaluation tasks, the thirty 3rd generation sketches were evaluated to create the fourth generation. The best three results are G3.1.2 (C2.1.2), G3.2.1 (C2.2.1) and G3.3.7 (C2.3.7) (in Figure 7 - 15).

Best 3 drawings in 3rd generation - combination 2

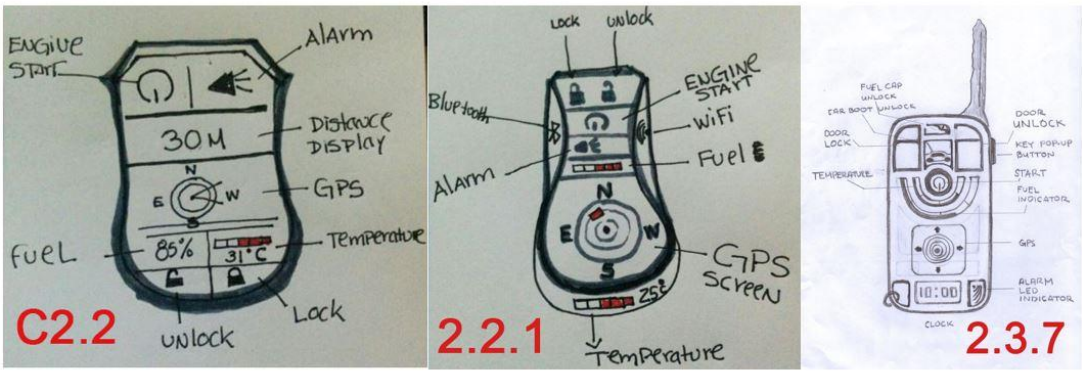


Figure 7 - 15 The best three drawings in the 3rd generation designs

7. 4th Generation Creation Task

The best three 3rd generation will be combined to create the 4th generation. As shown in the previous creation tasks, the same combination method was used again (illustrated in Figure 7 - 16). Three combination groups and ten approved combination drawings in each were required to collected. Figure 7 - 17, Figure 7 - 18 and Figure 7 - 19 show all approved 4th generation sketches submitted from crowd workers.

4th Generation Creation – Combination 3

Group 1: Best G3.1 + Best G3.2 = G4.1 = C3.1
Group 2: Best G3.1 + Best G3.3 = G4.2 = C3.2
Group 3: Best G3.2 + Best G3.3 = G4.3 = C3.3

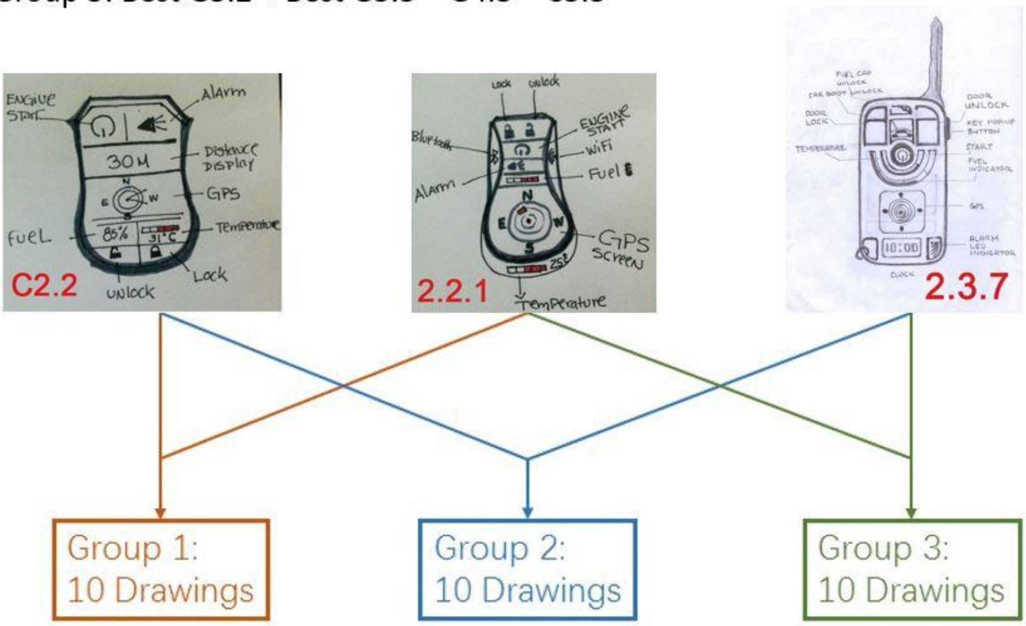


Figure 7 - 16 4th generation creation task (the combination process using the best three designs from the 3rd generation)

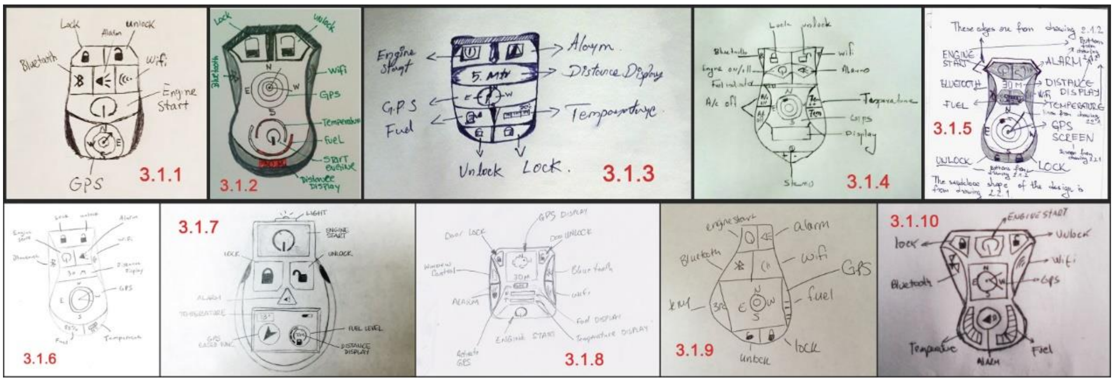


Figure 7 - 17 10 drawings of group 1 designs – 4th generation, combination 3

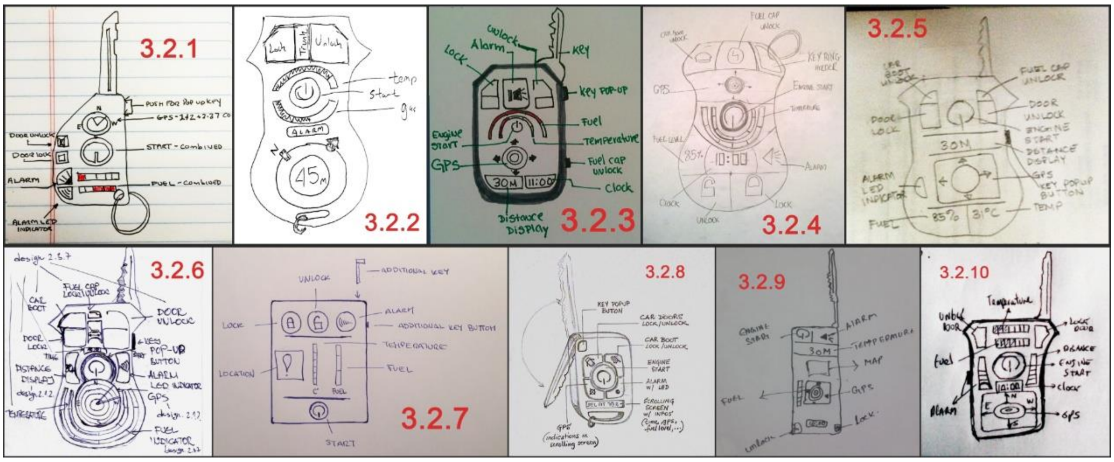


Figure 7 - 18 10 drawings of group 2 designs – 4th generation, combination 3

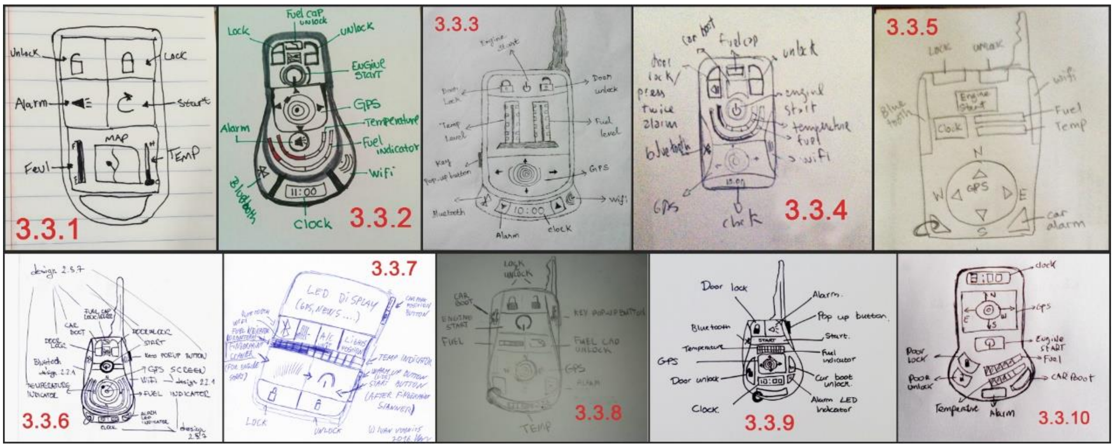


Figure 7 - 19 10 drawings of group 3 designs – 4th generation, combination 3

8. 4th Generation Evaluation Task

In this stage, all 4th generation drawings were evaluated in their respective group. So the experiment created the final outputs by this stage of the car key fob conceptual drawings. The best three results are shown below (in Figure 7 - 20): G4.1.2 (C3.1.2), G4.2.8 (C3.2.8) and G4.3.2 (3.3.2).

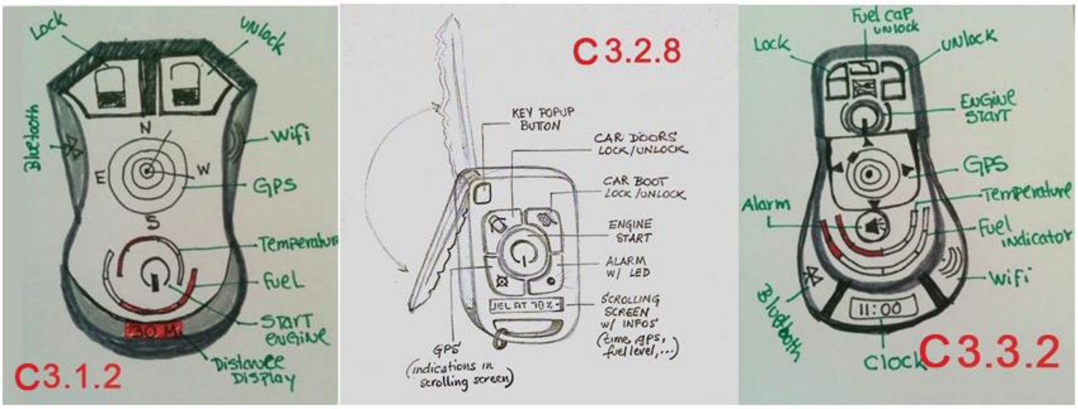


Figure 7 - 20 The best three 4th generation drawings

V. DISCUSSION & CHAPTER SUMMARY

In previously chapters of applications of the crowdsourced design process to the generation of living room floor plan and 3D kitchen plans open, cloud based design tools (i.e., Google Drawing, Autodesk Homestyler) had been used, in contrast the results here demonstrate improving design quality by using free-hand sketches and functional evaluation criteria. To verify the process had improved design quality after the iterative design and evaluation stages of the experiment, the best three drawings in the 4th generation and the best three drawings in the 1st generation were evaluated together using the same evaluation method employed throughout the experiment (e.g., crowd rank against five criteria). The results are as follows: the highest ranked drawing is C3.2.8, then is C3.1.2 followed by No. 64 (ranking results are shown in). This suggests that among the best three drawings, 2/3 comes from the last generation of car key fob conceptual drawings, which confirmed that after employing an HBGA process (that was structured using the cDesign Framework), the final output of design quality was improved.

Drawing No.	Rank1	Rank2	Rank3	Rank4	Rank5	Rank6	Rank7	Rank8	Rank9	Rank10	Average Ranking
3.3.2	3	2	2	5	4	2	6	6	5	5	5
147	2	6	3	4	1	5	1	4	4	6	4
64	4	1	1	3	6	6	3	3	2	4	3
3.2.8	6	3	4	2	2	4	4	1	1	2	1
78	1	4	6	6	5	3	2	5	6	3	6
3.1.2	5	5	5	1	3	1	5	2	3	1	2

Figure 7 - 21 Evaluation ranking results of the best designs from each generation

The validity of the cDesign framework can be judged in terms of:

Completeness

Effectiveness

Easy to use

Flexibility

Consequently, in this chapter, a car key fob design experiment on mTurk was described. This design experiment applied the cDesign Framework, and used the HBGA crowdsourced design method to systematically improve the design quality. In total, four generations of drawings were created by the crowd and evaluated. During the design creation stage, the best features from each pair of drawings were combined by the human workers to generate the new drawings. The process of evaluation and combination repeated to generate better quality of designs. The final evaluation shows that in this car key fob design task, the process resulted in improved conceptual design quality by a comparison between the last generation designs and the first generation designs. There were no steps in the creation or execution of the car key fob design task that were not covered by the cDesign framework.

In the next chapter, some more information from the crowd workers will be discussed, and the statistical analysis of the evaluations will be discussed.

CHAPTER 8

DISCUSSION

I. INTRODUCTION

In this chapter, the distributions of the evaluation results are discussed. It is suggested that a normal distribution of the evaluation results can be used to assess if the evaluation methods were appropriate to the experiments. Besides, in the 3D kitchen room layout design task a different evaluation method was used (different from the evaluation methods used in the living room layout design task and car key fob design experiment), so in this section the distribution will be discussed in two parts:

1) the distribution of the evaluation results for the living room layout design task and car key fob design task
2) the distribution for the 3D kitchen room layout design task

II. DISTRIBUTION TRENDLINES OF EACH EXPERIMENT

In this section, the frequency distribution of the crowd design evaluation data will be described in terms of statistical trendlines. Sternberg and Grigorenko (Sternberg & Grigorenko 2003) has discussed that the distribution of creativity in the population as follows:

"Francis Galton (1986) first established that human abilities tend to be distributed in the population according to the 'normal' or 'bell-shaped' curve. His demonstration was based partly on data – the fit of the normal curve to performance on examinations – and partly on analogy to the distribution of physical traits, such as height and weight. Since Galton, the normal distribution has become almost an article of dogma, firmly ingrained in the statistics psychologists use and in their conception of individual differences, including intelligence (Burt, 1963). Moreover, it is clear that this faith is not unfounded, for the bell curve provides a reasonable approximation to most empirically observed distributions. Not surprisingly,

creativity has often been perceived after the same fashion (Nicholls, 1972). Presumably, most human beings exhibit average levels of the capacity, the frequencies tapering off in either direction, with creative genius being about as rare as those who are virtually incapable of producing a creative idea."

He went on to discuss the exceptional individual who dominate some creative industries, however since the crowd is drawn from the general population his argument for a skewed distribution are not relevant.

Ultimately the results of a crowdsourced design methodology are critically dependent on the effectiveness of the evaluation process. Without an effective evaluation process the best designs cannot be reliably identified from the hundreds generated and consequently competition or HBGA approaches would perform poorly. This section investigates various forms of statistical analysis with the aim of:

1) Assessing the effectiveness of an evaluation process
2) Quantifying the process parameters of first, number of workers, and second, number of generations required to produce a certain level of performance.

To do this the following statistical tools were used:

Frequency Distribution

There is academic evidence that design performance like many other human activities is normally distributed over the population (Runco 2004). For example, Figure 8 - 1 below illustrates a normal distribution, and Figure 8 - 2 shows a random distribution in which the evaluation does not distinguish a standard performance.

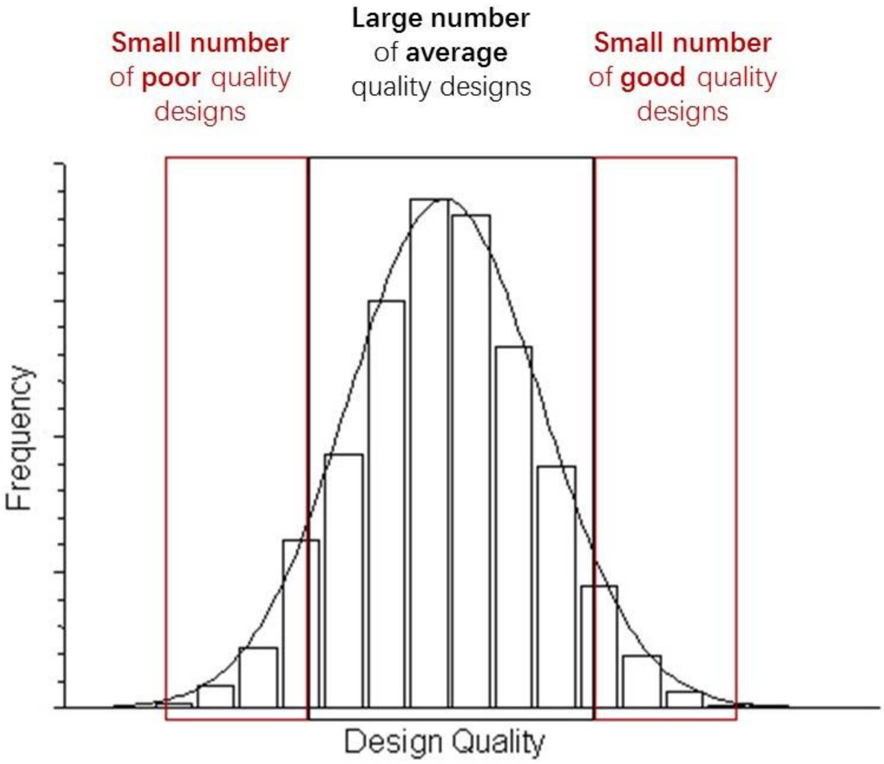


Figure 8 - 1 A normal distribution showing the design quality and the frequency

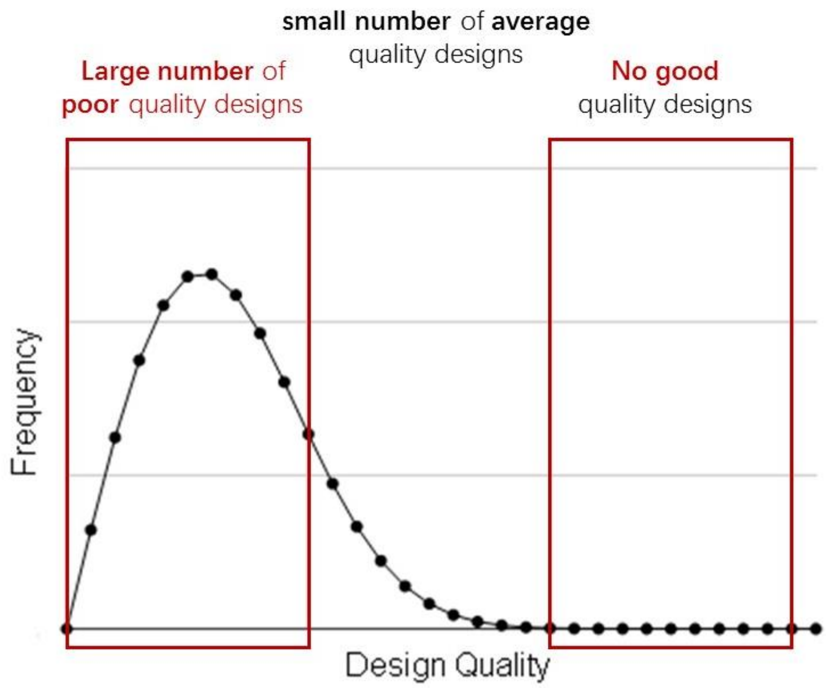


Figure 8 - 2 Evaluation does not distinguish a standard performance

Trendline Chart/Normal Probability Plot

The trendline chart is used to depict trends in the existing data or forecasts of future data. Figure 8 - 3 is an example of the trendline chart. In this chart, it can be observed that the trendline shows a normal distribution for the average ranking results, because $r^2 = 0.9699$ which is very close to 1.

Inverse Normal Distribution

Inverse normal distribution establishes a normal distribution can be used in reverse to answer question such as:

If the quality of designs is normally distributed with an average of μ and a standard deviation of σ, calculate the number of designs required to produce a, say, 80% probability that the results include designs ranked in the top, say, 10%.

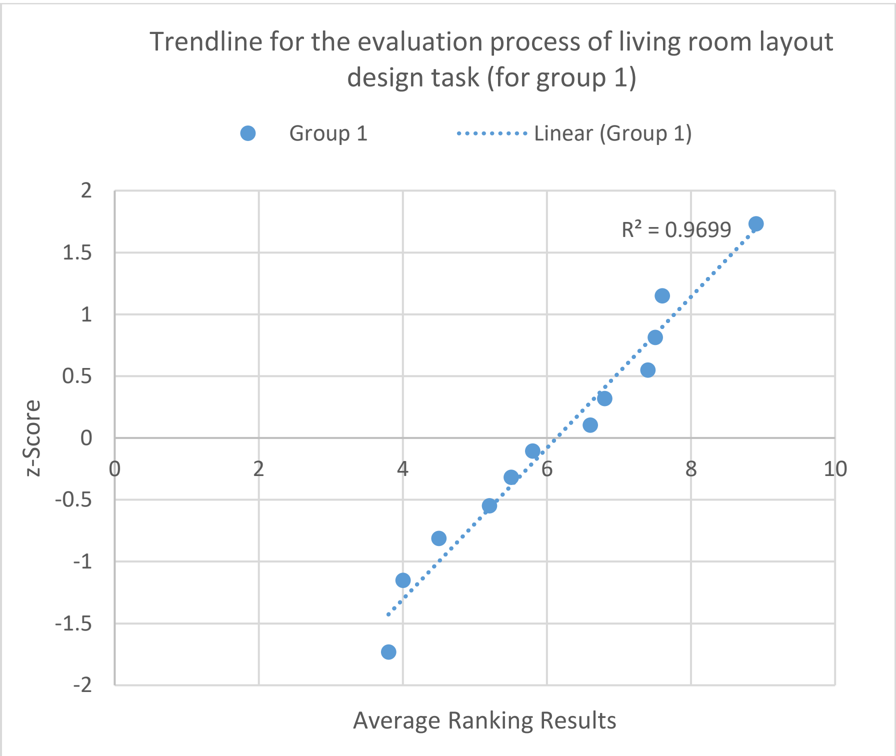


Figure 8 - 3 Trendline for the average ranking results

1. Distribution Trendlines in Living Room Layout Design Task & Car Key Fob Design Task

1.1 Living Room Layout Design Task

In the living room layout design task, when participants evaluated designs, it was required that all designs should be evaluated by marks: from 1 (the worst) to 100 (the best). However, because different people have different standards of judgements, the ranking of designs in groups was used to provide a relative ranking of design quality. In this section, the trendlines showing the distributions of these rankings will be described, and their significance discussed.

To assess the trendlines for the distribution of the living room layout design task, all the ranking results were integrated into one table (shown as Table 8 - 1). As described in Chapter 4, all living room layouts were randomly separated into seven groups (here were 12 layouts in group 1 to 6, and 11 layouts in group 7). As shown in the table below, the average rankings of each design layouts were listed, and the function of z-Score was calculated using the excel function NORMSINV[9].The results clearly shows linear trendlines of all rankings, the first step is to illustrate one group's trendline (in Figure 8 - 4).

To illustrate the distribution of the design evaluation results, the normal probability plot will be used. The normal probability plot is "a graphical technique to identify substantive departures from normality", and normal probability plots "are made of raw data" from experiments' results [10]. As shown in Figure 8 - 4, x-axis is the raw data of the average ranks, and y-axis is the z-Score. The dots in the figure are the ranking results of group 1 living room layout designs, and the line is the trendline of the distribution for the ranking results. The correlation coefficient with the straight line r^2 = 0.9699 suggests that the trendline of the

[9] =NORMSINV (returns the inverse of the standard normal cumulative distribution. The distribution has a mean of zero and a standard deviation of one) (for example in this ranking evaluation task, z- Score is =NORMSINV((An-0.5)/12), n = the relative vertical number). Because in group 12, there were 11 layouts in the group instead of 12 in other groups, so the function of group 7's z-Score is =NORMSINV((A2-0.5)/11).

[10] Normal probability plot, https://en.wikipedia.org/wiki/Normal_probability_plot, 10/2016

average ranking results is a good fit for a normal distribution. Furthermore, in Figure 8 - 5, all trendlines showing the average ranking data and r^2 information illustrate that all the evaluation ranking results in living room layout design task followed a “standard distribution”. And this standard distribution can approve that the evaluation ranks obey the statistical regulation. This result suggests that the judgements of design quality is in dependent and tend towards a normal distribution as the number increases.

Table 8 - 1 Average ranking data in each group (i.e., assessment of 10 MTurk workers) for the living room layout design task

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	z-Score 1 (for group 1 to 6)	z-Score 2 (for group 7)
1	3.8	3.7	3.6	4	3.9	2.8	3.2	-1.73166	-1.69062
2	4	3.8	4.4	4.4	4.7	4.6	3.3	-1.15035	-1.0968
3	4.5	4.6	5.1	5	4.7	5	4	-0.81222	-0.74786
4	5.2	4.7	5.2	5.2	5.2	5.1	4.2	-0.54852	-0.47279
5	5.5	5.8	6	5.6	5.3	5.3	6	-0.31864	-0.22988
6	5.8	6.5	6	5.7	5.7	5.3	6.1	-0.10463	0
7	6.6	6.7	6.5	6.4	6.4	5.8	6.7	0.104633	0.229884
8	6.8	6.8	6.9	6.4	6.8	6.1	6.8	0.318639	0.472789
9	7.4	7.1	7.1	7	6.9	7.2	7	0.548522	0.747859
10	7.5	7.4	7.6	7.2	7.7	7.8	7.6	0.812218	1.096804
11	7.6	7.8	7.8	7.5	8.1	8.2	8	1.150349	1.690622
12	8.9	8.7	8.1	9.1	8.4	9		1.731664	

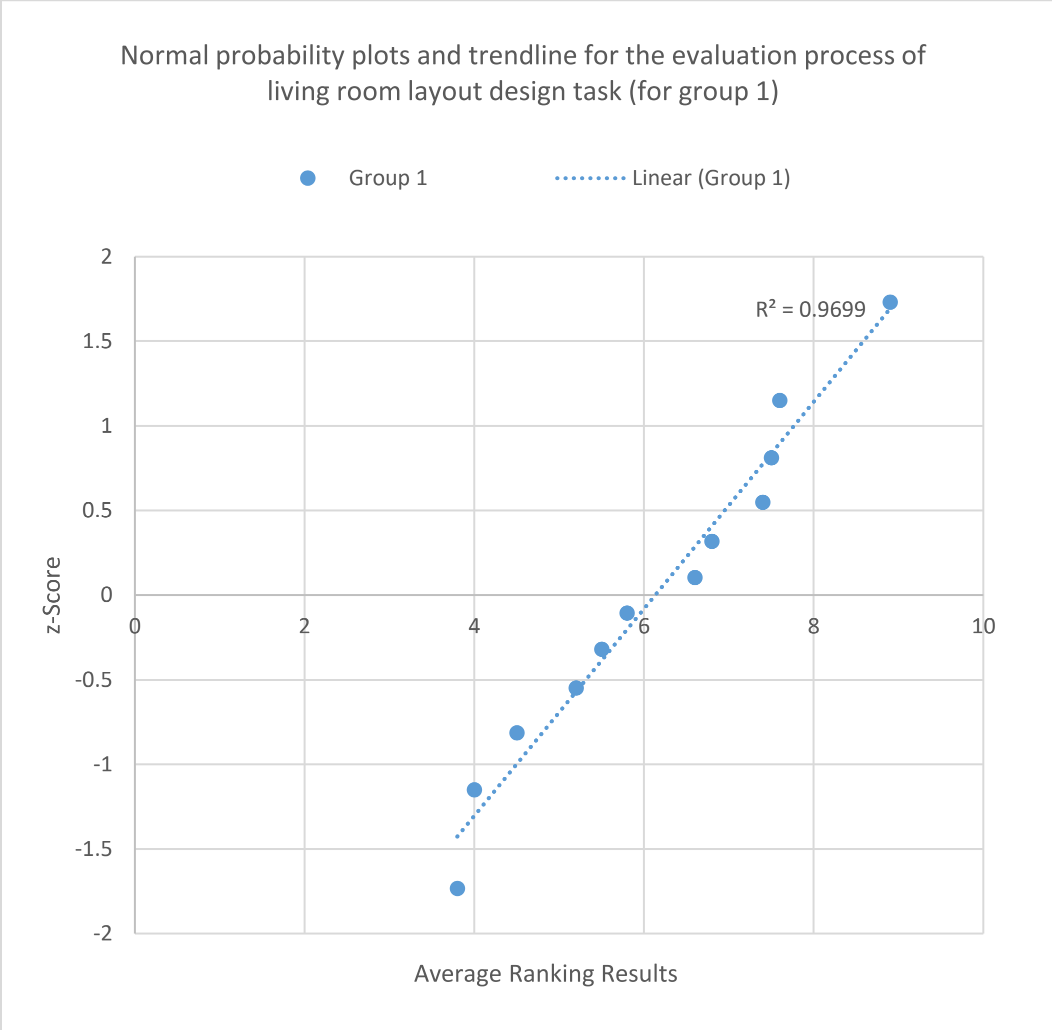


Figure 8 - 4 The normal probability plots, trendlines and r^2 information of group 1 average ranking results in living room layout design task

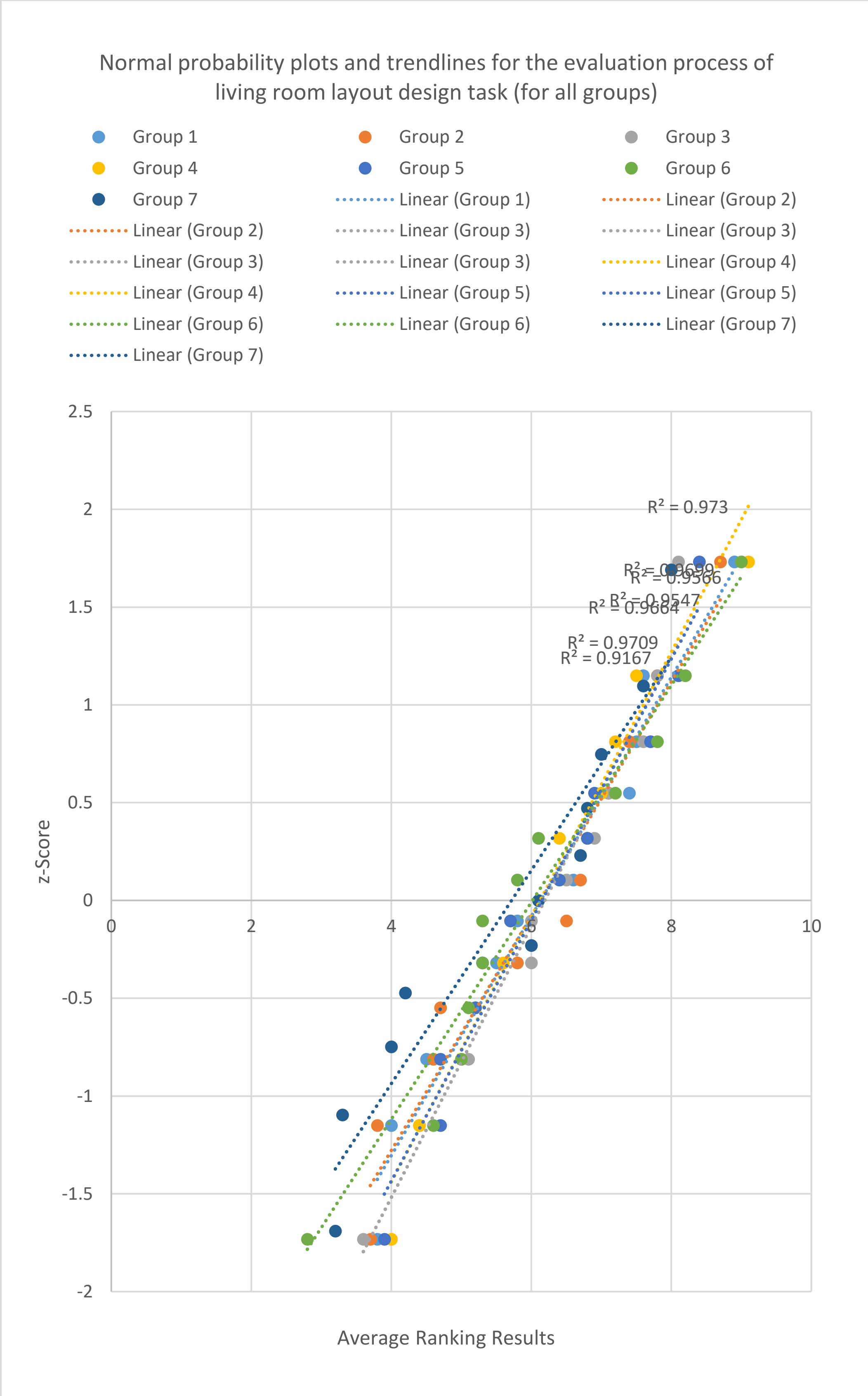


Figure 8 - 5 The normal probability plots, trendlines and r^2 information of group 1 average ranking results in living room layout design task

1.2 Car Key Fob Design Task

The car key fob design experiment employed similar evaluation methods, so it is assumed that the trendlines and distribution results will be similar as well. Table 8 - 2 shows the raw average data for all designs in the car key fob design tasks.

Table 8 - 2 Average ranking data for each generation in car key fob design task

	E4-1	E4-2	E4-3	E3-1	E3-2	E3-3	E2-1	E2-2	E2-3	E1-1
1	1.625	1.5	2.625	3.5	1.875	1.5	2.625	2.5	1.75	2.375
2	2.5	2.5	3.75	5	4	3.375	3.125	3.375	3.375	3.875
3	4.25	4.375	4.5	5.25	4.625	3.875	3.5	4.25	3.875	4.5
4	5.125	5.125	4.625	5.375	4.625	4.375	5	5.875	4.5	4.875
5	5.25	6.25	4.625	5.625	5	5.375	5.125	6	4.875	5.625
6	5.5	6.5	6	5.75	5.375	5.875	5.625	6.125	5.125	5.75
7	5.75	6.625	6.25	5.875	5.375	6.625	6.25	6.25	6.625	6.125
8	7.5	6.75	7.125	6.125	7	6.875	6.375	6.375	7	6.5
9	8.5	7.625	7.5	6.5	8	8	7.75	6.5	8.375	6.875
10	8.75	8.25	8.125	6.75	8.75	8.875	9	7.5	9.375	9.375

	E1-2	E1-3	E1-4	E1-5	E1-6	E1-7	E1-8	E1-9	E1-10	E1-11
1	3.375	3.125	2.625	2.25	2.5	2.75	1.625	1.125	2.25	2.5
2	3.75	4.25	3	2.75	3.75	4.625	2.5	4.25	2.25	2.875
3	4.5	4.5	4.25	2.875	4.25	4.75	4.75	4.375	3.5	3.75
4	4.625	4.75	4.375	4.875	4.75	4.875	5.5	4.5	4.875	4.5
5	4.75	5.5	4.75	5	5.5	5	6.125	4.75	5.75	5.125
6	5.25	5.75	5.125	5.75	5.75	5.5	6.125	5.25	6.25	5.75
7	5.5	5.875	6.375	6	6.125	5.5	6.25	7.25	6.375	7.375
8	6.125	6.75	7.875	7.125	6.375	6.125	6.875	7.375	7.5	7.375
9	7.75	7.75	8.125	8.875	7.5	6.25	7	7.5	7.625	7.625
10	9.25	7.75	8.75	9.125	8.5	9.25	8	8.375	8.375	8.375

	E1-12	E1-13	E1-14	E1-15	E1-16	E1-17	z-Score
1	2.25	3.375	2.875	2	3.375	2.375	-1.64485
2	3.25	3.375	3.25	2.625	4.375	2.875	-1.03643
3	4	3.625	3.75	3.625	4.375	3.875	-0.67449
4	4.75	4.125	4.75	4.625	4.625	4	-0.38532
5	5.125	5.375	4.875	4.625	5.25	5.625	-0.12566
6	5.5	5.625	5.625	6.125	5.625	6	0.125661
7	6	6.125	5.625	7	6	6.625	0.38532
8	6.375	7.375	6.75	7.375	6.375	6.75	0.67449
9	7.75	8.125	7.75	8.375	7.625	6.875	1.036433
10	9.75	8.375	9.25	9.375	7.875	9.625	1.644854

Clearly illustrated in Figure 8 - 6, (En-m in the figure means: the evaluation ranks for No.m design which is in the n^{th} generation) in terms of the normal distribution plots, all ranking results from the 1^{st} generation shows that the distribution is standard (in total, there were 17 groups of designs in the 1^{st} generation). The r^2 information are shown in Table 8 - 3. It can be found that only two groups' r^2 are less than 0.9 (group 7 and 8). As for the 2^{nd}, 3^{rd} and 4^{th} generation evaluation ranking data, the normal distribution plots are shown in Figure 8 - 7, and their r^2 are illustrated in Table 8 - 4. From the trendlines and r^2, it can be found that the distributions of the rank data are standard (only E2-2's r^2 = 0.8736 < 0.9).

Table 8 - 3 R^2 information of the 1^{st} generation evaluation ranks

Group Number	E1-1	E1-2	E1-3	E1-4	E1-5	E1-6	**E1-7**	**E1-8**	E1-9
R^2	0.9606	0.9031	0.966	0.9435	0.9481	0.9926	**0.8578**	**0.8912**	0.9108
Group Number	E1-10	E1-11	E1-12	E1-13	E1-14	E1-15	E1-16	E1-17	
R^2	0.9378	0.944	0.9740	0.9212	0.9609	0.9762	0.9624	0.9494	

Table 8 - 4 R^2 information of the 2^{nd}, 3^{rd} and 4^{th} generation evaluation ranks

Group Number	E2-1	**E2-2**	E2-3	E3-1	E3-2	E3-3	E4-1	E4-2	E4-3
R^2	0.9716	**0.8736**	0.9813	0.9051	0.9484	0.9899	0.9578	0.9202	0.9687

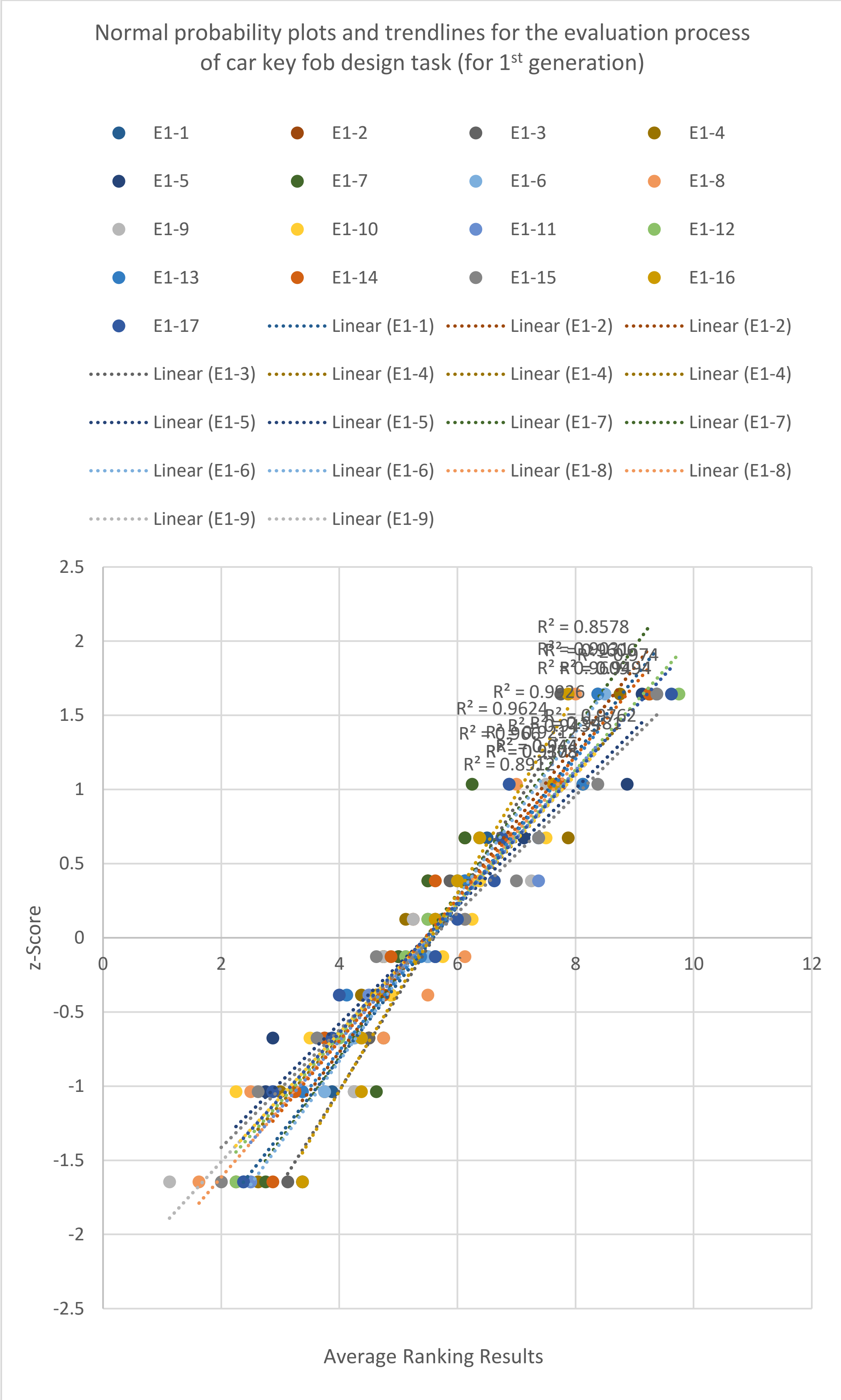


Figure 8 - 6 The normal probability plots, trendlines and r^2 information of the average ranking results in car key fob design task (1st generation)

Figure 8 - 7 The normal probability plots, trendlines and r^2 information of the average ranking results in car key fob design task (2nd generation)

2. Evaluation ranking distribution in 3D kitchen room layout design task

The 3D kitchen room layout design task applied a different 5-Likert Scale method to judge the design quality. As described in the chapter 5, to evaluate individual kitchen room layout design, participants were required to follow a ten-criteria questionnaire and give scores (from 1: the worst, to 5: the best). However, the difference between the method used in the 3D kitchen room layout design task and other two case studies was that crowd workers were not required to rank designs in an order. So this section investigates whether these two different methods can provide similar distribution results. To enable this comparison the scores provided by participants, all 3D kitchen room layout designs will be ranked. The method of assessment is:

1) Add all scores (each criterion has one score, and each design has ten criteria, so the score for one individual design is: the sum of ten criteria's scores)
2) Designs are ranked (i.e., ordered) by their total scores
3) Calculate the normal probability plots and r^2

As shown in Table 8 - 5, the scores for the 1st generation 3D kitchen room layout design are listed. So, the ranking results of each worker's evaluation scores can be illustrated as Table 8 - 6. To create the trendlines of the 1st generation designs, an average ranking results (i.e., the average ranking of each design is the sum of all ranking results (36 participants' scores) divide 36) was used. As a result, Table 8 - 7 shows the average ranking results of each designs. Based on the average rankings, the normal probability plot is illustrated as Figure 8 - 8. From the figure, it can be found that the trendline is different from the previous trendlines, and the r^2 = 0.0025 also suggests that the distribution of the average rankings is not a standard distribution. To investigate whether the same pattern occurs evaluation data of the other generation, the same method is used to generate the average rankings.

Table 8 - 5 Evaluation scores of the 1st generation 3D kitchen room layout designs

Worker No.	Worker 1	Worker 2	Worker 3	Worker 4	Worker 5	Worker 6	Worker 7
Design 1	32	34	30	36	32	42	30
Design 2	27	48	34	40	21	23	25
Design 3	29	36	20	31	30	33	29

Worker No.	Worker 1	Worker 2	Worker 3	Worker 4	Worker 5	Worker 6	Worker 7
Design 4	35	38	20	40	24	43	28
Design 5	37	39	27	46	32	32	27
Design 6	20	31	18	28	20	39	34
Design 7	28	50	29	37	37	43	33
Design 8	26	28	22	39	28	41	31
Design 9	25	33	20	37	19	44	31
Design 10	36	50	26	36	30	44	30
Worker No.	Worker 8	Worker 9	Worker 10	Worker 11	Worker 12	Worker 13	Worker 14
Design 1	39	38	36	29	36	21	30
Design 2	47	28	39	42	30	17	35
Design 3	35	40	34	41	25	26	29
Design 4	40	50	40	31	40	32	21
Design 5	35	32	42	41	28	40	35
Design 6	36	27	35	19	30	22	25
Design 7	41	44	37	47	26	38	38
Design 8	35	20	33	36	29	23	25
Design 9	36	21	35	30	26	21	35
Design 10	33	43	37	44	25	30	36
Worker No.	Worker 15	Worker 16	Worker 17	Worker 18	Worker 19	Worker 20	Worker 21
Design 1	39	35	33	23	32	38	49
Design 2	47	39	25	35	28	40	19
Design 3	42	31	32	18	28	39	42
Design 4	40	35	26	27	34	38	17
Design 5	45	28	24	28	38	40	40
Design 6	40	34	26	23	37	40	42
Design 7	41	38	40	17	37	38	43
Design 8	43	40	19	22	37	38	42
Design 9	36	31	25	19	44	34	45
Design 10	40	36	18	24	42	40	43
Worker No.	Worker 22	Worker 23	Worker 24	Worker 25	Worker 26	Worker 27	Worker 28
Design 1	22	27	30	38	38	35	17
Design 2	25	24	37	33	37	36	46
Design 3	25	24	43	32	36	38	22
Design 4	13	21	36	33	32	41	17
Design 5	34	31	45	30	45	38	49
Design 6	19	22	38	35	33	38	36
Design 7	35	37	50	34	39	36	45
Design 8	20	20	18	36	31	37	32
Design 9	19	27	21	30	32	38	19
Design 10	26	23	50	33	29	40	28
Worker No.	Worker 29	Worker 30	Worker 31	Worker 32	Worker 33	Worker 34	Worker 35
Design 1	50	37	35	44	37	44	18
Design 2	45	31	34	36	42	50	17

Worker No.	Worker 29	Worker 30	Worker 31	Worker 32	Worker 33	Worker 34	Worker 35
Design 3	45	36	24	42	38	42	11
Design 4	47	36	22	40	40	50	13
Design 5	45	43	29	47	40	50	22
Design 6	44	29	27	49	31	50	19
Design 7	46	45	37	47	28	31	25
Design 8	47	36	24	34	41	50	22
Design 9	45	31	27	45	38	50	13
Design 10	45	35	30	41	40	50	25

Table 8 - 6 Ranking results of the 1st generation 3D kitchen room layout designs

Rank No.	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7
Design 1	4	7	2	7	2	5	5
Design 2	7	3	1	2	8	10	10
Design 3	5	6	7	9	4	8	7
Design 4	3	5	7	2	7	3	8
Design 5	1	4	4	1	2	9	9
Design 6	10	9	10	10	9	7	1
Design 7	6	1	3	5	1	3	2
Design 8	8	10	6	4	6	6	3
Design 9	9	8	7	5	10	1	3
Design 10	2	1	5	7	4	1	5
Rank No.	Rank 8	Rank 9	Rank 10	Rank 11	Rank 12	Rank 13	Rank 14
Design 1	4	5	6	9	2	8	6
Design 2	1	7	3	3	3	10	3
Design 3	7	4	9	4	9	5	7
Design 4	3	1	2	7	1	3	10
Design 5	7	6	1	4	6	1	3
Design 6	5	8	7	10	3	7	8
Design 7	2	2	4	1	7	2	1
Design 8	7	10	10	6	5	6	8
Design 9	5	9	7	8	7	8	3
Design 10	10	3	4	2	9	4	2
Rank No.	Rank 15	Rank 16	Rank 17	Rank 18	Rank 19	Rank 20	Rank 21
Design 1	9	5	2	5	8	6	1
Design 2	1	2	6	1	9	1	9
Design 3	4	8	3	9	9	5	5
Design 4	6	5	4	3	7	6	10
Design 5	2	10	8	2	3	1	8
Design 6	6	7	4	5	4	1	5
Design 7	5	3	1	10	4	6	3
Design 8	3	1	9	7	4	6	5
Design 9	10	8	6	8	1	10	2
Design 10	6	4	10	4	2	1	3
Rank No.	Rank 22	Rank 23	Rank 24	Rank 25	Rank 26	Rank 27	Rank 28
Design 1	6	3	8	1	3	10	9
Design 2	4	5	6	5	4	8	2
Design 3	4	5	4	8	5	3	7
Design 4	10	9	7	5	7	1	9

Rank No.	Rank 22	Rank 23	Rank 24	Rank 25	Rank 26	Rank 27	Rank 28
Design 5	2	2	3	9	1	3	1
Design 6	8	8	5	3	6	3	4
Design 7	1	1	1	4	2	8	3
Design 8	7	10	10	2	9	7	5
Design 9	8	3	9	9	7	3	8
Design 10	3	7	1	5	10	2	6
Rank No.	**Rank 29**	**Rank 30**	**Rank 31**	**Rank 32**	**Rank 33**	**Rank 34**	**Rank 35**
Design 1	1	3	2	5	8	8	6
Design 2	5	8	3	9	1	1	7
Design 3	5	4	8	6	6	9	10
Design 4	2	4	10	8	3	1	8
Design 5	5	2	5	2	3	1	3
Design 6	10	10	6	1	9	1	5
Design 7	4	1	1	2	10	10	1
Design 8	2	4	8	10	2	1	3
Design 9	5	8	6	4	6	1	8
Design 10	5	7	4	7	3	1	1

Table 8 - 7 The average of the 1st generation

	Average Rankings		z-Score
Design 1	5.17	1	-1.64485
Design 2	4.80	2	-1.03643
Design 3	6.23	3	-0.67449
Design 4	5.34	4	-0.38532
Design 5	3.83	5	-0.12566
Design 6	6.14	6	0.125661
Design 7	3.46	7	0.38532
Design 8	6.00	8	0.67449
Design 9	6.29	9	1.036433
Design 10	4.31	10	1.644854

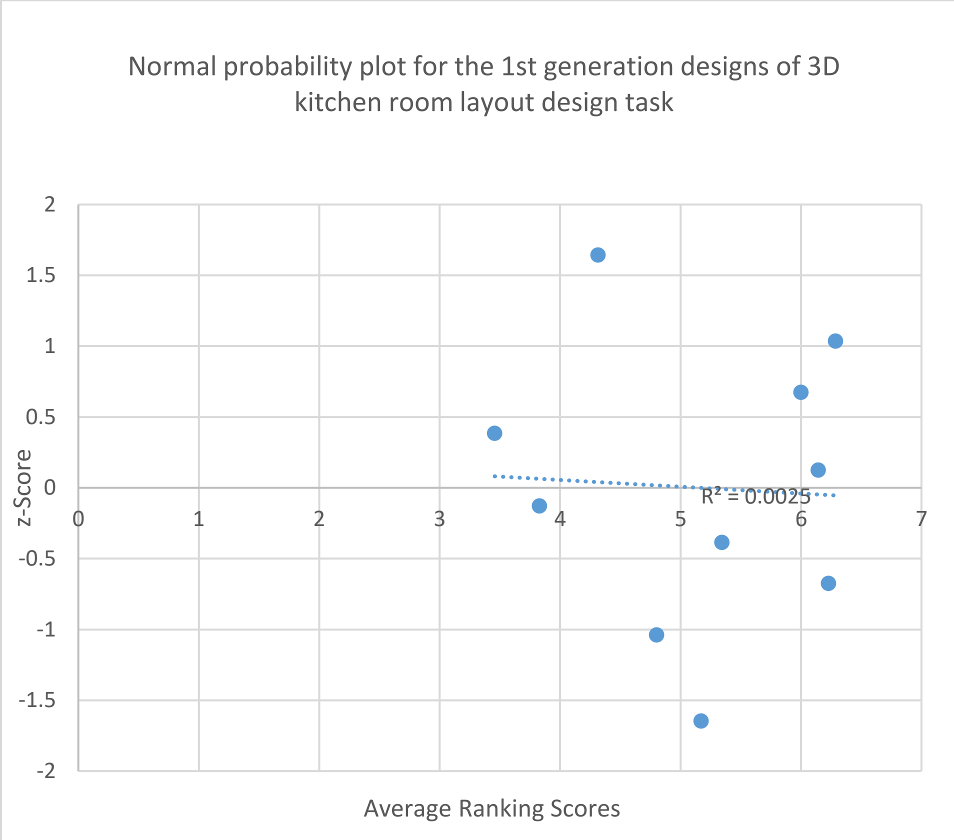


Figure 8 - 8 The normal probability plot of the 1st generation designs in 3D kitchen room layout design task

Using the same method, the illustrations of the average ranking results in groups 1, 2 and 3 designs of the 2nd generation are shown in Table 8 - 8 (in this case study, statistically three numbers of data are not convincing to present the distribution. The future work section in chapter will talk about this in detailed). Based on these average rankings, the plots of the 2nd generation designs can be illustrated by different groups in Figure 8 - 9, Figure 8 - 10 and Figure 8 - 11. Similarly, although the r^2 suggests that the distribution becomes better than the 1st generation, the distributions are not standard either (r^2 closer to 1, the more standard distribution).

Table 8 - 8 The average rankings of the 2nd generation designs in 3D kitchen room layout design task

	Average Rankings		z-Score
Group 1			
Design 1	1.80	1	-0.96742
Design 2	1.60	2	0
Design 3	2.27	3	0.967422
Group 2			
Design 1	1.80	1	-0.96742
Design 2	1.60	2	0

	Average Rankings		z-Score
Group 2			
Design 3	2.27	3	0.967422
Group 3			
Design 1	2.07	1	-0.96742
Design 2	1.80	2	0
Design 3	1.80	3	0.967422

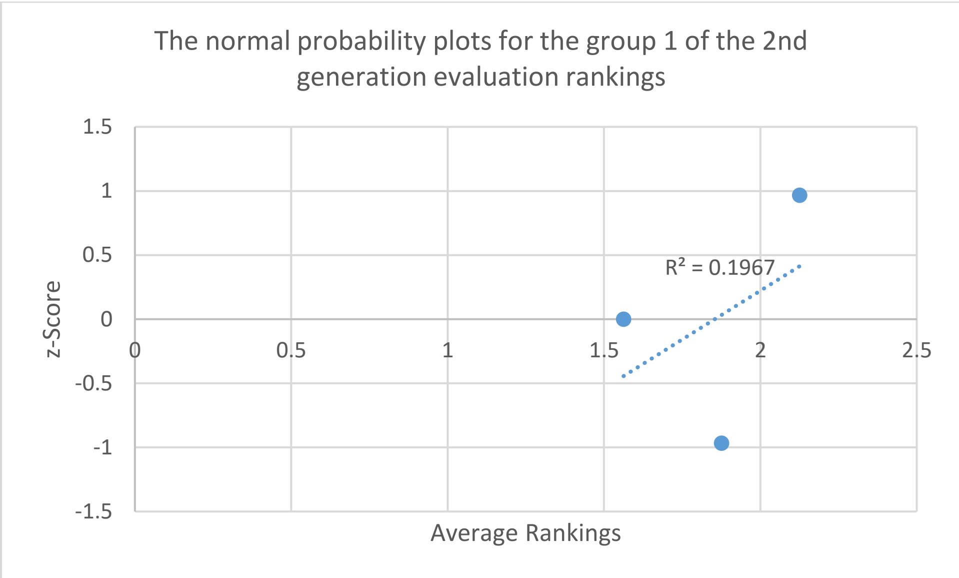


Figure 8 - 9 The normal probability plots for the group 1 of the 2nd generation evaluation rankings

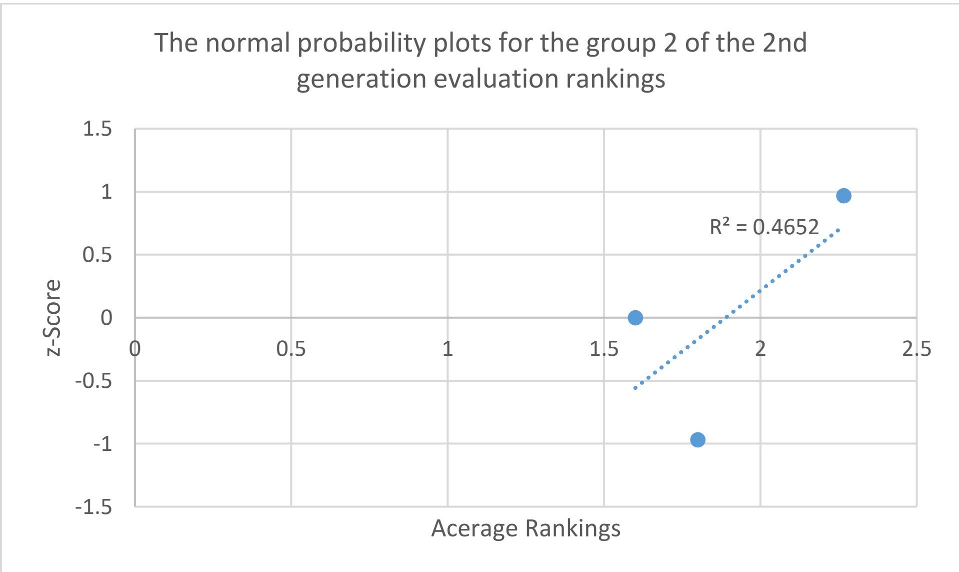


Figure 8 - 10 The normal probability plots for the group 2 of the 2nd generation evaluation rankings

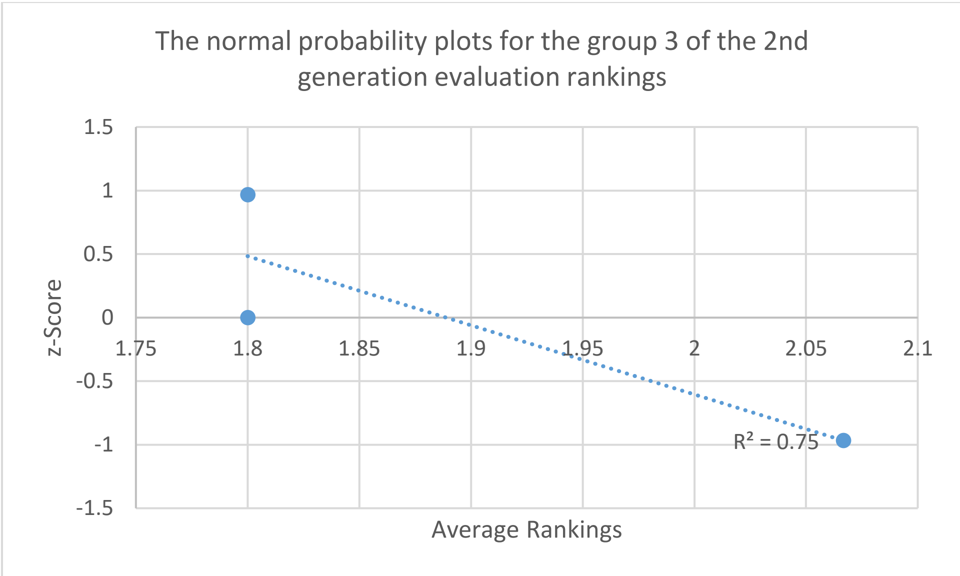


Figure 8 - 11 The normal probability plots for the group 3f the 2nd generation evaluation rankings

What is more, in the next evaluation stage, six best designs (three from the 1st generation, and three from the 2nd generation) among all generations were evaluated together. Besides, the best two designs among these six were evaluated with two more layouts generated by design experts in the final stage. Based on the average rankings in Table 8 - 9, Figure 8 - 12 and Figure 8 - 13 illustrate the trendlines and r^2s. The similar results on trendlines and r^2s determine that the distribution of the average rankings are not as standard as the distribution in other two case studies.

Table 8 - 9 Average rankings in the top 6 design evaluation stage and the final design evaluation stage

	Average Rankings		z-Score
Top 6 design rankings			
Design 1	2.57	1	-1.38299
Design 2	3.00	2	-0.67449
Design 3	3.48	3	-0.21043
Design 4	2.95	4	0.210428
Design 5	3.00	5	0.67449
Design 6	3.67	6	1.382994
Final Evaluation rankings			
Design 1	2.00	1	-1.15035
Design 2	2.30	2	-0.31864

Design 3	2.75	3	0.318639
Design 4	2.55	4	1.150349

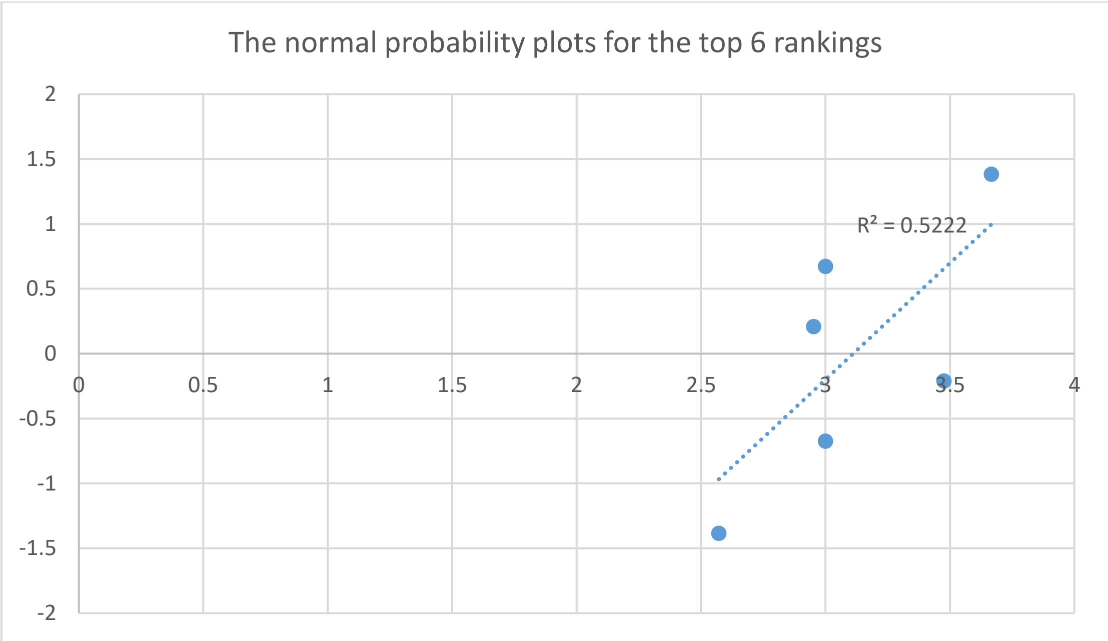


Figure 8 - 12 The normal probability plots for the top 6 rankings

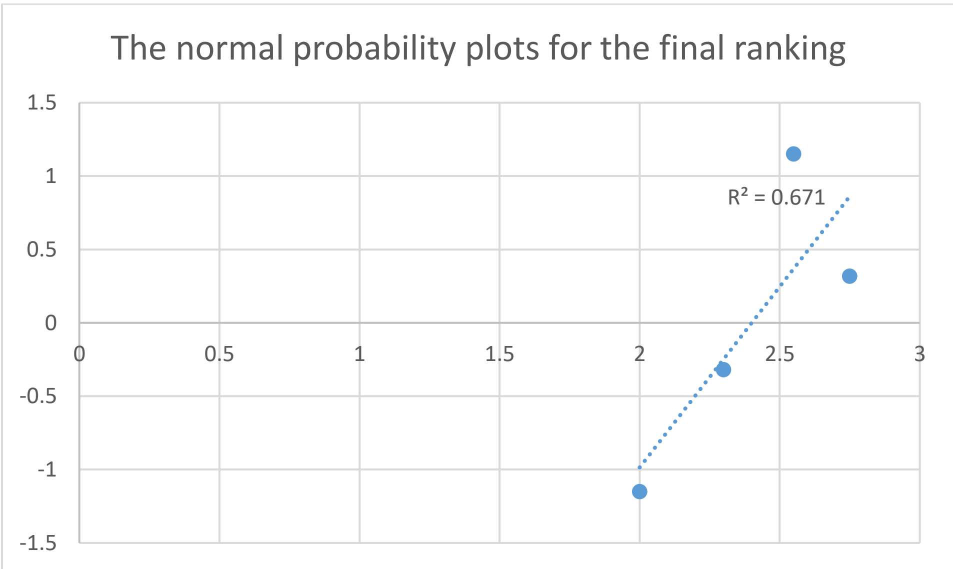


Figure 8 - 13 The normal probability plots for the final ranking

III. CHAPTER SUMMARY

In this chapter, the distribution of the evaluation ranking results was investigated as well. The discussions about the ranking distributions, results suggested that in the living room layout designs task and car key fob design task, standard distribution trendlines and r^2 results proved the effects of the ranking evaluation method used in these two case studies. In terms of the 3D kitchen room layout design case study, although the 5-Likert Scale evaluation method was applied in the experiment, and the experiment results suggested that this method carried out the appropriate process to the experiment, the distribution analysis determines that this method are required to be optimized in the feature cDesign tasks. However, the r^2 results in the 2nd generation did increase dramatically and were much higher than that of the 1st generation, a hypothesis can be planed as the feature work that following the cDesign framework, in the new generations, the distribution will be more standard.

The next chapter will conclude the Ph.D. research work, the contribution of knowledge, the limitations and the future work.

References

Eickhoff, C. & Vries, A. de, 2012. Increasing cheat robustness of crowdsourcing tasks. Information Retrieval, 16(2), pp.121–137. Available at: http://link.springer.com/10.1007/s10791-011-9181-9 [Accessed June 22, 2013].

Levay, K.E., Freese, J. & Druckman, J.N., 2016. The Demographic and Political Composition of Mechanical Turk Samples. SAGE Open.

Runco, M.A., 2004. Everyone Has Creative Potential, Washington, DC: American Psychological Association.

Sternberg, R. & Grigorenko, E., 2003. The Psychology of Abilities, Competencies, and Expertise, Cambridge University Press. Available at: https://books.google.co.uk/books?id=9LEthsz9oCoC&dq=normal+distribution+of+creativity+ability&hl=zh-CN&source=gbs_navlinks_s.

Yu, L. & Nickerson, J., 2011. Cooks or cobblers?: crowd creativity through combination. CHI 2011, May 7–12, Vancouver, BC, Canada. Available at: http://dl.acm.org/citation.cfm?id=1979147 [Accessed June 25, 2013].

CHAPTER 9

CONCLUSION, LIMITATION & FUTURE WORK

I. INTRODUCTION

In this chapter, the results of the research work are summarized and conclusions made. The research started by understanding the process of crowdsourcing and the contributions of other researchers. Then three case studies were undertaken to:

1) Understand the crowdsourced design methods, platforms and crowdsourced design process
2) The relationship between payment, quality of design, quantity of design and evaluation method in crowdsourced design
3) The limits of the design representation that can be supported by investigation of 3D layout designs can be crowdsourced and the different design evaluation method.

The experience of these case studies was integrated to generate the cDesign framework for crowdsourced design, which is one of the contributions to knowledge of this work. Finally, to validate the cDesign Framework, the last case study was carried out to test each stage in the framework and whether the whole crowdsourced design process can be used in new type of designs (i.e., free-hand sketch).

The structure of this chapter is as follows: in the first section, each chapter will be summarized, then the objectives and aims of the studies will be discussed in the light of the case studies' findings. Also the research questions listed in methodology chapter will be discussed in the context of the reported investigations. Importantly, the main research contribution to knowledge in cDesign research area will be illustrated, and the limitations of this research work will be discussed (e.g., the design tools limitation, the platform limitation, etc.). Lastly the opportunities for the future work in the area of crowdsourced generative design.

II. RESEARCH SUMMARY

Chapters 1 – 'Introduction of crowdsourcing and crowdsourced design' provides an overview introduction of crowdsourcing and crowdsourced design including their background and a brief description on the benefits of their application. Then the aims and objectives of this research are identified. In chapter 2, the literature review chapter, the history of crowdsourcing is described, as well as the improvement of crowdsourcing in different research area. Moreover, to understand the crowdsourced design process, the crowdsourcing platforms and crowdsourcing process were investigated from the reported research work. In the last section of chapter 2, the research gap is illustrated: although the existed crowdsourced design process (i.e., Yu's combination process(Lixiu Yu & Nickerson, 2011)) can provide a workflow, but the process lacks a systematic process and does not detail each crowdsourced design stage (i.e., crowdsourcing platform selection, crowdsourced design prototype task, design evaluation criteria, etc.). So, the establishment of a framework for crowdsourced design was targeted in this research.

Based on chapter 2: the literature review in crowdsourced design and related research areas, the first case study (named the 'desk lamp design task') was undertaken to test the functionality of the crowdsourcing platform, crowdsourced design tool and the reported design process (Nickerson, Sakamoto, & Yu, n.d.)(L Yu & Nickerson, n.d.)(Zhao & Zhu, 2012)(Lixiu Yu, 2011)(Little, Chilton, Miller, & Goldman, 2009)(Heer & Bostock, 2010). By undertaking the desk lamp design task in chapter 4, the basic crowdsourcing process on the commercial crowdsourcing platforms were established, and the experience gained of using the Human-based Genetic Algorithms (that supports the final crowdsourced design model) – this case study was to understand the mechanism of crowdsourcing, the crowdsourcing platform and the basic processes of crowdsourced design. In the second case study – living room layout design task in chapter 4 (in this experiment, participants were required to draw a living room layout plan with the given requirements, i.e., including TV, sofa, tea table, etc.), to investigate one basic question on crowdsourcing platform in design tasks that what is the relationship between payment to crowd workers and the design quality, the payment to online workers were set into different levels (the design task was the same). The aim of the experiment was to test whether a higher payment can bring higher quality of design results. So, by judging the experiment results (design quality) in the different payment levels, the statistical results suggested that a higher payment does not bring a higher average design quality. However, because workers were paid more, an increasing number of participants involved in the experiment, which means that the increase of the payment resulted in a larger number of designs being submitted. As a result, requesters would be more likely to receive higher quantity of designs, raising the payment to participants could results in higher quality if the design evaluation process was effective in identifying the best designs.

After two case studies, in chapter 5, the third experiment – 3D kitchen room layout design experiment was undertaken to investigated if 3D designs be crowdsourced. In this crowdsourced 3D design task, participants were required to design a 3D kitchen using a public online 3D interior design tool. During the experiment, the Human-based Genetic Algorithms was applied to generate new designs by 'combination'. In the design process, participants were required to design a 3D kitchen room which should meet the given requirements (i.e., the design should contain: a microwave oven, oven, dinner table/chairs, etc.). Then the collected approved designs were shown to participants to evaluate. Importantly, the evaluation criteria were collected from the participants using the qualitative research method called coding to identify the criteria. The resulting integrated evaluation criteria were called Crowdsourced Design Evaluation Criteria (cDEC). Workers evaluated design by cDEC and ranked them, the first three designs were combined with each other's best features to generate the following generation. As a conclusion, this 3D kitchen room layout design case study suggested the following results:

1) the 3D design can be crowdsourced as well 2D design
2) after a systematic crowdsourced design process, the final design quality will be better than the first generation
3) although it has been reported that crowd can design, and crowd can evaluate designs (Bao, Sakamoto, & Nickerson, 2011), crowd can evaluate by using evaluation criteria created by themselves instead of the evaluation criteria provided by requesters or experts.

Based on these three case studies – to conduct a number of case studies to identify the choice and parameters inherent in using crowdsourcing for design (Research Aim 1), in chapter 6, a novel crowdsourced design framework (cDesign framework) were established (details will be concluded in the next Contribution of Knowledge section, also the Research Aim 2 and 3). So, to validate this framework, the last framework validation case study was designed.

In the case study – car key fob design task (in chapter 7, for Research Aim 3), crowd workers were required to draw a car key fob conceptual sketch by the free-hand drawing, instead of any computer-based design tool (e.g., Google Drawing ™, Autodesk Homestyler ™). The purposes of this design experiment were: firstly, to validate the cDesign framework. Each step in this experiment followed the process detailed in the cDesign framework. For example, before the experiment, the platform and the crowd workers had to be selected, the cDEC were collected from crowd, the Human-based Genetic Algorithms was applied to generate new designs, etc. In total, four generations of designs were generated. The statistical results illustrated that guided by the cDesign framework, the final design quality was increased.

Secondly, because in the previous investigations, the free-hand sketch has not been approved that can be used as a design tool in the crowdsourced design task, and this design tool can provide an increased design quality followed by the crowdsourced design methods. In this experiment, participants were asked to submit their designs as free-hand sketches. The use of free-hand sketches placed less constraints on the participants but created more work for the requester. This was different from the digital designs tools, such as the Google Draw ™, which had been applied in the previous case studies. This was because in contrast to the online digital designs or drawings can be shared to requesters directly. The scans of the free-hand drawings had to be formatted and arranged so they could be easily viewed online by the requester. However the use of used the free-hand sketches allowed participants who had never used an online digital design tool or any other computer-based design tools to take part without spending a long time learning that how to use an online system. In the case studies 1, 2 and 3, the first section of the task description was a short tutorial illustrating how to use the required design tool (i.e., by a tutorial video webpage link). Consequently, it was considered that this method may influence and limit the creative ability or capacity to illustrate ideas for online crowd workers. The free-hand sketch does not have this limitation. The work reported in chapter 7 shows that the free-hand sketch and scan (i.e., take photo of drawings or scan drawings) method was successfully applied as a tool design representation in a crowdsourced design task.

Consequently, the results of all case studies have addressed the research hypothesis. The crowdsourcing technology can be used to create the complete crowd-based design process, and the cDesign Framework is a general used crowdsourced design model.

III. CONTRIBUTION TO KNOWLEDGE

The research makes four contributions within the overall structure work of the cDesign framework (shown in Figure 9 - 1). The novelty of the framework is that the framework provides a systematic process and detailed stages for crowdsourced design, which can be applied in general crowdsourced design task (not for specific design tasks). Although before this research work some crowdsourced design methods and process were reported, for example, Yu's crowdsourced design method in children chair design task (Lixiu Yu & Nickerson, 2011) which applied the Human-based Genetic Algorithms research method in the first time (in crowdsourced design area), a general framework analogous to Pugh's Total Design model had not been reported. The framework will enable crowdsourcing to be more quickly applied in design applications. Based on the experiments' results in this research, the cDesign framework systematically defines each stage in the creation of crowdsourced design tasks on commercial platforms. This general cDesign framework includes four main following stages:

1) Crowdsourced design task specification
2) Crowdsourced design task validation (the task prototype)
3) Crowdsourced design task execution (NIDT: Non-iterative Design Task & IDT: Iterative Design Task – using Human-based Genetic Algorithms)
4) Crowdsourced design results evaluation

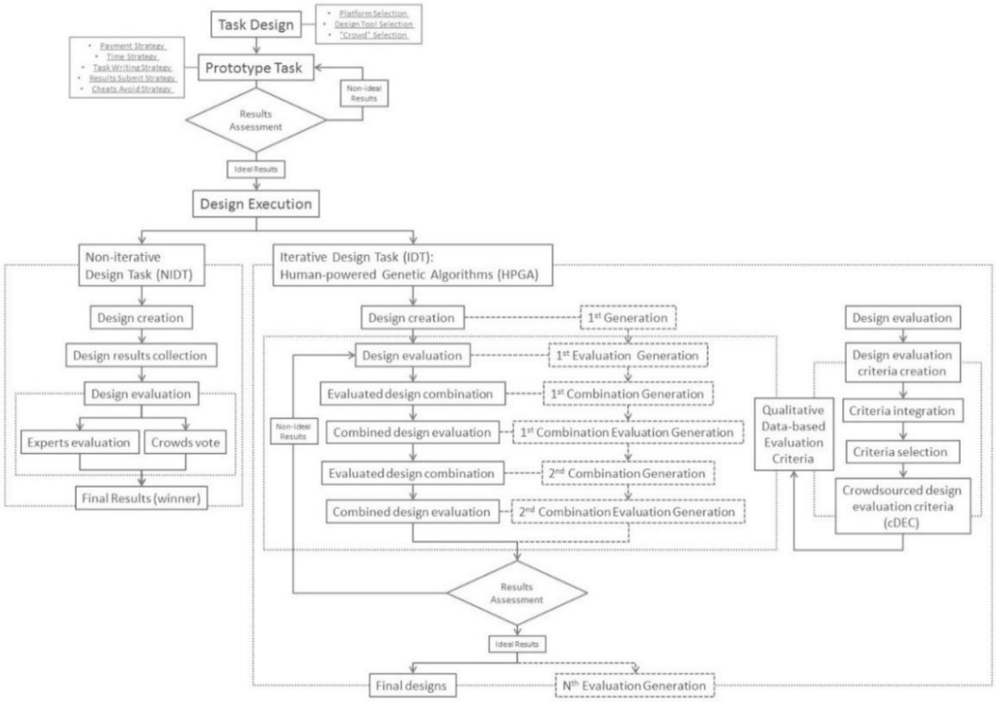


Figure 9 - 1 The cDesign framework

These four main stages in cDesign framework integrate the general solutions, methods and tools that together can enable effective crowdsourced design tasks.

For example, in the first stage: crowdsourced design task specification, if requesters already have the design problem, the first issue to address is the selection of an appropriate crowdsourcing platform to host their jobs. Because different platforms have different characteristics, for example, Crowdflower is fast to collect information, which can be used for questionnaire tasks (i.e., collect design specifications and evaluation criteria).

In the second stage: the design task prototyping stage, the framework listed the following possible elements which could influence the design results, such as the payment to crowd workers, the submission sharing and transferring with participants and requesters, methods to avoid cheats, etc. Because the cDesign framework is a general framework that is based on several different case studies, the model can be applied to different type of design tasks to create good quality of designs.

The second contribution to knowledge of this research work are the insights reported about the relationship between payment to crowd workers and the design quality. As described in chapter 4, the results suggest a higher payment, or reward, to participants does not increase the quality of designs. However, if requesters would like to receive large numbers of results, or submissions with a wide range of design quality, one possible way is to increase the payment (Hao Wu, Corney, & Grant, 2014b).

Thirdly, the 3D kitchen room layout design task (the 3rd case study) demonstrated how the HBGA approach can be applied to 3D design using online, cloud based environments (i.e., Autodesk Homestyler ™). Although for companies like GE, 3D modelling of engineering components are said to have been 'crowdsourced' in an online community where gathering the engineering design experts (Hao Wu, Corney, & Grant, 2014a), this community was not open to crowd and adopted a competition. In this research work, the 3D layout design firstly totally open to online crowd workers. The results validated that 3D interior layout designs can be effectively crowdsourced using the cDesign framework the final design quality was increased (Hao; Wu, Corney, & Grant, 2015).

Fourthly, the use of scanned (or photographed) free-hand sketches was shown to be an effective design tool when applied in the cDesign framework. As described in the last section, although from the reported research work, a kind of free-hand sketch has been used to generate ideas for product design in crowdsourced design task, these reported free-hand sketches were based on computer and cloud technology based (i.e., Google Drawing ™) (Lixiu Yu, 2011)(Lixiu Yu & Nickerson, 2011). The importance of 'pure' free-hand sketch (i.e., non-computer based, for example, conceptual ideas were created by only a pen and a piece of paper physically) has been reported in different research areas, such as conceptual design and architectural design (Puttre, n.d.) (Juchmes, Leclercq, & Azar, n.d.), it was assumed that if free-hand sketch can be used to represent the design in crowdsourcing, participants could save time of learning new design tools. But the sharing and transferring of participants drawing results with requester's receiving was one of the most difficult problem. As described in chapter 7 (car key fob design task), participants drew conceptual designs each followed by the design specifications in the first step, and then transferred the physical drawings into digital format (i.e., took pictures, scanned drawing by scanner). Those digital format drawings were finally submitted to requesters as their task submissions. Results indicated that the non-computer based free-hand sketch can be used as a design tool in crowdsourced designs. Additionally, in chapter 8, the results also demonstrated how statistical methods can be used to validate methods of assessing design quality.

To sum up, in this section the contribution to knowledge of this research work was described. The crowdsourced process was firstly integrated into a general framework which can guide the crowdsourced design process on commercial crowdsourcing sites. In the next section, limitations of the research work will be described.

IV. LIMITATIONS

Although this research work has produced the mentioned contributions, limitations always emerged during the experiments. So, in this section, the following limitations will be described:

1) Design tools limitations
2) Limited types of crowdsourced design tasks
3) cDesign framework in real creative design project (instead of the theoretical application)

In this research work, three different design tools were applied in the crowdsourced design tasks, which were Google Drawing ™ and free-hand sketch for 2D design tasks, and Autodesk Homestyler ™ for 3D design tasks. Although it has been validated by the research results that these kinds of design tools can provide solutions for crowdsourced design tasks, more design tools or creative applications have already existed. With the development of the worldwide Internet environment (especially the booming improvement for mobile Internet environment) and the smart devices (i.e., drawing or design applications for mobile phone and other smart devices), more design tools started to appear in human's daily life. For example, even some product design specifications can be directly created drawings on mobile phone's screen (i.e., Samsung's Note series mobile phone can create 2D drawings fast), which is convenient to share designs as submissions to requesters. This participation in crowdsourced by applying different mobile design tools may influence the design quality. So, the limitation of the application of various design tools was the first limitation of the research.

The second limitation caused by the limited types of crowdsourced design tasks in this research. There were four case studies which was discussed in this thesis which are as follows: 1), desk lamp design task: product design task; 2), living room layout design task: interior design task; 3), kitchen room layout design task: interior design task; and 4), car key fob design task: product design task. However, design includes various basic categories which is shown in Figure 9 - 2. This Walker's Design Family Tree (1989) shows a useful view for design category (the tree does not specify all design categories). This research investigated two types of designs (which are in the design category: product design and interior design) and the application of cDesign framework in these designs, it is assumed that different types of designs may require different types of cDesign frameworks (the main structure and fundamental framework may the same, or similar, with cDesign framework illustrated in this thesis).

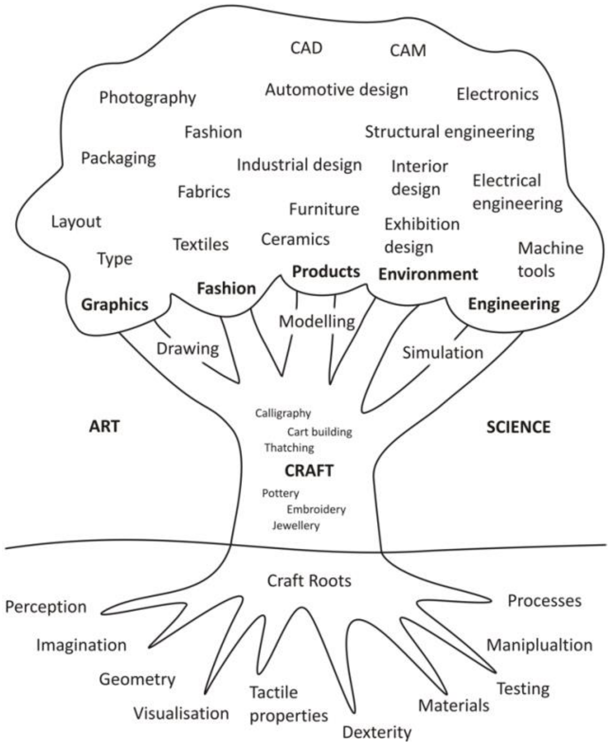


Figure 9 - 2 Walker's Design Family Tree

As for the last limitation: cDesign framework applied in real design projects (i.e., product design project for mobile phones), because all case studies in this research work were theoretically applied the crowdsourced design as a tool to generate designs, if use the framework in the physical design projects, influenced by the different surroundings and environment, the framework may need more optimization.

After the conclusion about the limitations of this PhD research, this thesis will be end up with the future work and a final conclusion.

V. FUTURE WORK & CONCLUSION

Based on the discussions in the last section, the research presented here demonstrated that crowdsourcing can be applied as an effective tool for the generation of designs. There are several obvious directions for future research, first, more online, or cloud, design creation tools can be investigated for use in the crowdsourced design process. Especially, the mobile design tools or applications with smart devices. In contrast to the current view that people (i.e., crowd workers) search design tasks on commercial crowdsourcing sites sitting in front of a PC or laptop, and use computer-based design tools to draw their designs or express ideas simply by a pen and a piece of paper, and then share the solutions with requesters, it can be imagined that in the future workers can participate in tasks on a bus or taxi, use their mobile devices to undertake different kinds of tasks and share results with requesters. What is more, the cDesign framework can be optimized in various design tasks. For example, can cDesign framework be used in a 3D modelling design task (design category: computer-aided design), or a tea table design task (design category: furniture design)? Furthermore, cDesign framework can be introduced to real design companies to investigate its application in real the market. Additionally, based on the statistical results in chapter 8, if more data can be collected from the future crowdsourced design tasks, the quality of designs (or concepts) could be correlated with the numbers of participants and thus the budget required? Additionally, because the cDesign framework is not the only design framework providing a crowdsourced design process to improve design quality, it is also important to investigate the differences between the cDesign framework and other design frameworks, or design models in the future. Last but not the least, as described in the Appendix 3, the background of crowd work may influence the design results as well. For example, whether the design/art educated workers can provide more creations, whether people living in different countries can show different abilities in innovation of crowdsourced design. These influence will be investigated in the future as well.

In conclusion, the research questions, aims and objectives were answered and investigated. A novel cDesign methodology and framework were established successfully. It is believed that crowdsourcing can raise the quality designs by the appropriate methods, and bring a better future for human life.

References

Bao, J., Sakamoto, Y., & Nickerson, J. (2011). Evaluating Design Solutions Using Crowds. In *Seventeenth Americas Conference on Information Systems, August 4th-7th, Detroit, Michigan*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2201651

Heer, J., & Bostock, M. (2010). Crowdsourcing graphical perception: using mechanical turk to assess visualization design. *Proceedings of the 28th International Conference on …, 152*(4), 203–212. Retrieved from http://portal.acm.org/citation.cfm?id=1753357

Juchmes, R., Leclercq, P., & Azar, S. (n.d.). A freehand-sketch environment for architectural design supported by a multi-agent system. *Computers & Graphics*, *Volume 29*(Issue 6), Pages 905–915.

Little, G., Chilton, L. B., Miller, R., & Goldman, M. (2009). TurKit : Tools for Iterative Tasks on Mechanical Turk. *IEEE Symposium on Visual Languages and HumanCentric Computing VLHCC*, (Figure 1), 252–253.

Nickerson, J., Sakamoto, Y., & Yu, L. (n.d.). Structures for creativity: The crowdsourcing of design. In *CHI 2011 Workshop on Crowdsourcing and Human Computation: Systems, Studies, and Platforms, May 8, 2011, Vancouver, BC, Canada.* Retrieved from http://crowdresearch.org/chi2011-workshop/papers/nickerson.pdf

Puttre, M. (n.d.). Gearing up for conceptual design. *Mechanical Engineering*, *115.3*(46.).

Wu, H., Corney, J., & Grant, M. (2014a). Crowdsourcing Measures Of Design Quality. In *34th Computers and Information in Engineering Conference (CIE)*.

Wu, H., Corney, J., & Grant, M. (2014b). Relationship between quality and payment in crowdsourced design. In *Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 499–504). Ieee. http://doi.org/10.1109/CSCWD.2014.6846895

Wu, H., Corney, J., & Grant, M. (2015). THE APPLICATION OF CROWDSOURCING FOR 3D INTERIOR LAYOUT DESIGN. In C. Weber, S. Husung, M. Cantamessa, G. Cascini, D. Marjanovic, & S. Graziosi (Eds.), *Proceedings of the 20th International Conference on Engineering Design (ICED 15)* (Vol. Vol 4, pp. 123–134). Retrieved from https://www.designsociety.org/publication/37769/the_application_of_crowdsourcing_for_3d_interior_layout_design

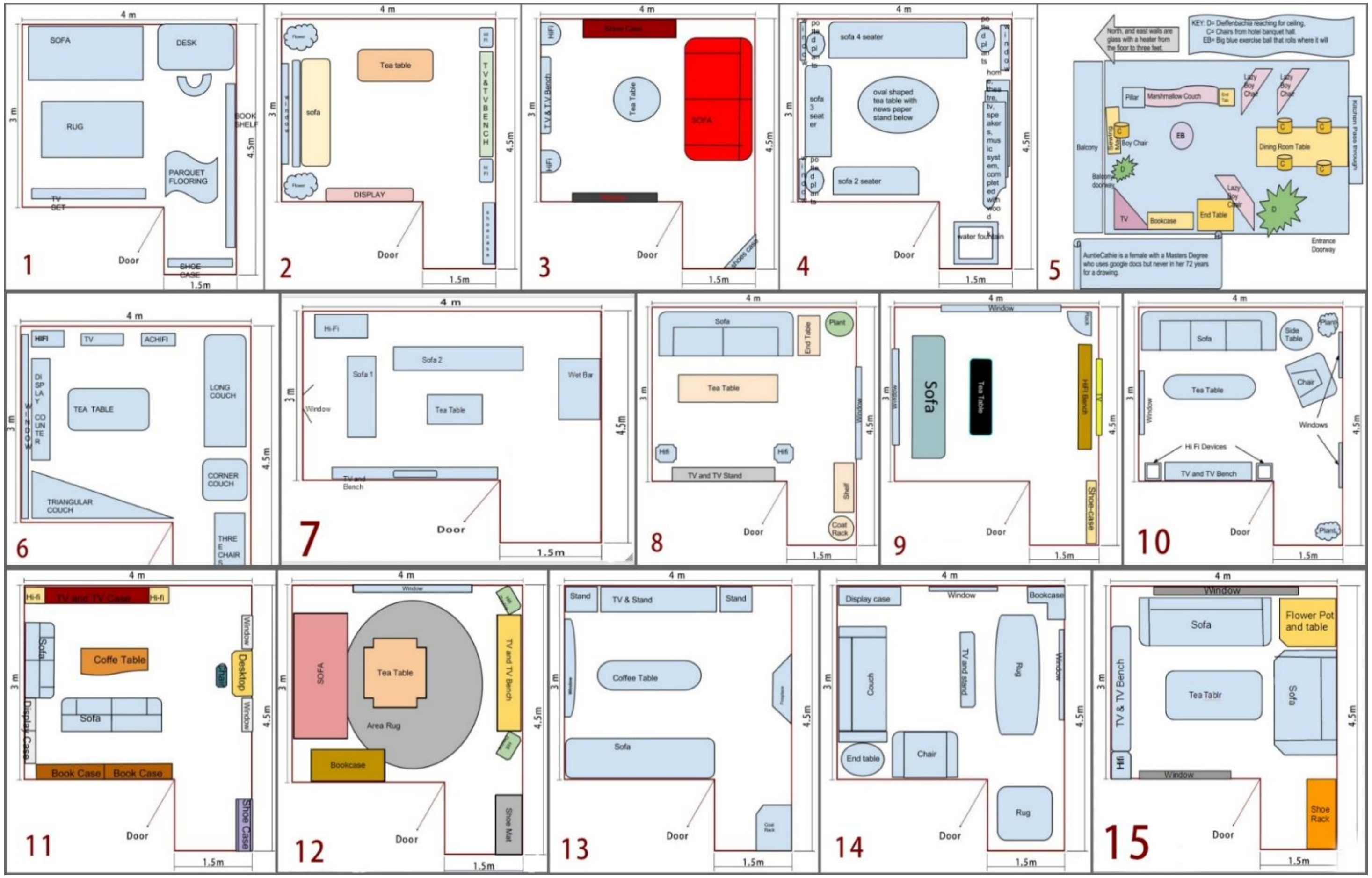
Yu, L. (2011). Crowd creativity through combination. *Proceedings of the 8th ACM Conference on Creativity and Cognition - C&C '11*, 471. http://doi.org/10.1145/2069618.2069756

Yu, L., & Nickerson, J. (n.d.). Generating creative ideas through crowds: An experimental study of combination. In *Thirty Second International Conference on Information Systems, Shanghai 2011*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2039547

Yu, L., & Nickerson, J. (2011). Cooks or cobblers?: crowd creativity through combination. *CHI 2011, May 7–12, Vancouver, BC, Canada*. Retrieved from http://dl.acm.org/citation.cfm?id=1979147

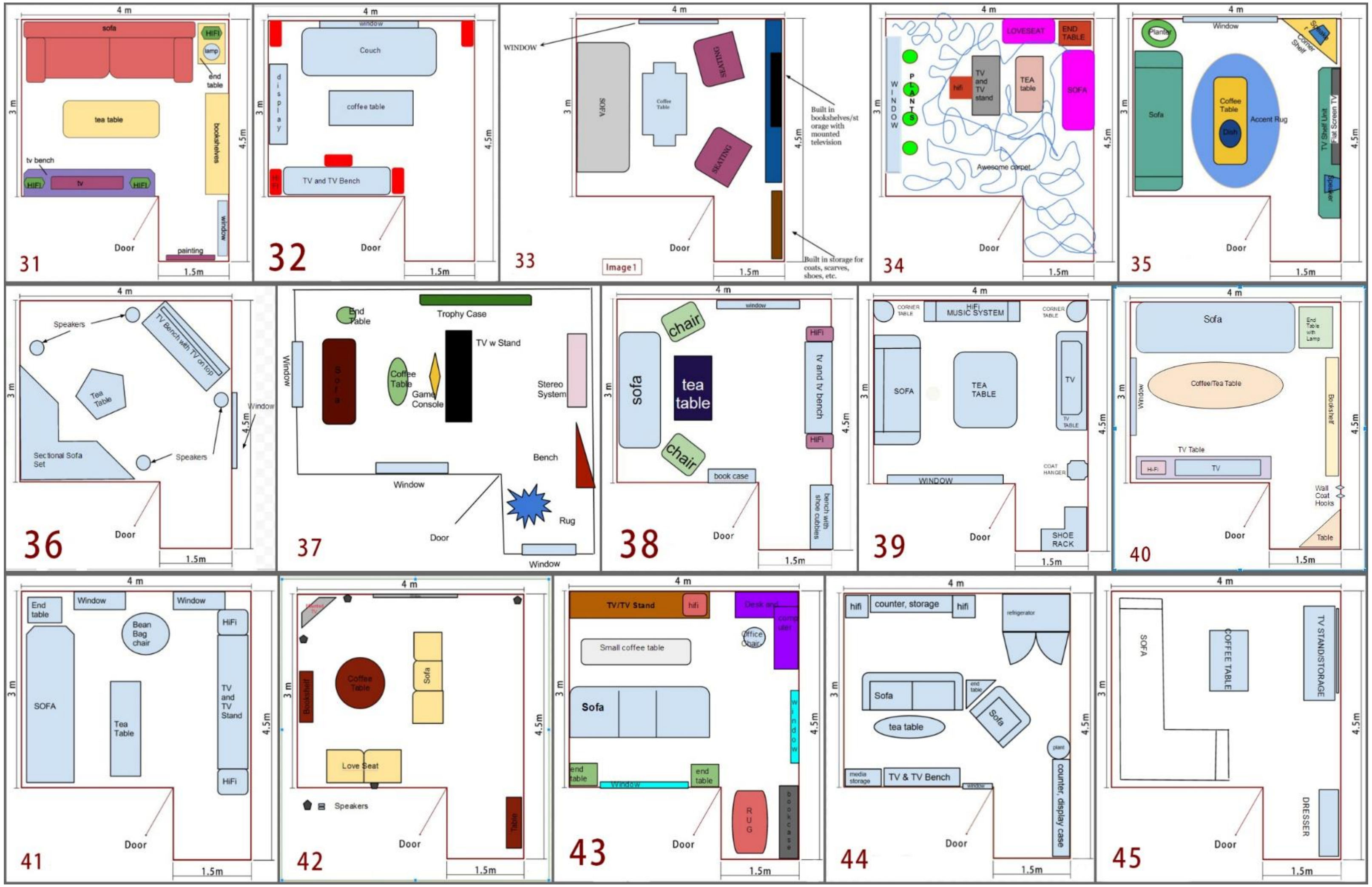
Zhao, Y., & Zhu, Q. (2012). Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers*. http://doi.org/10.1007/s10796-012-9350-4

APPENDIX 1: All living room layout designs generated by MTurk workers in case study A

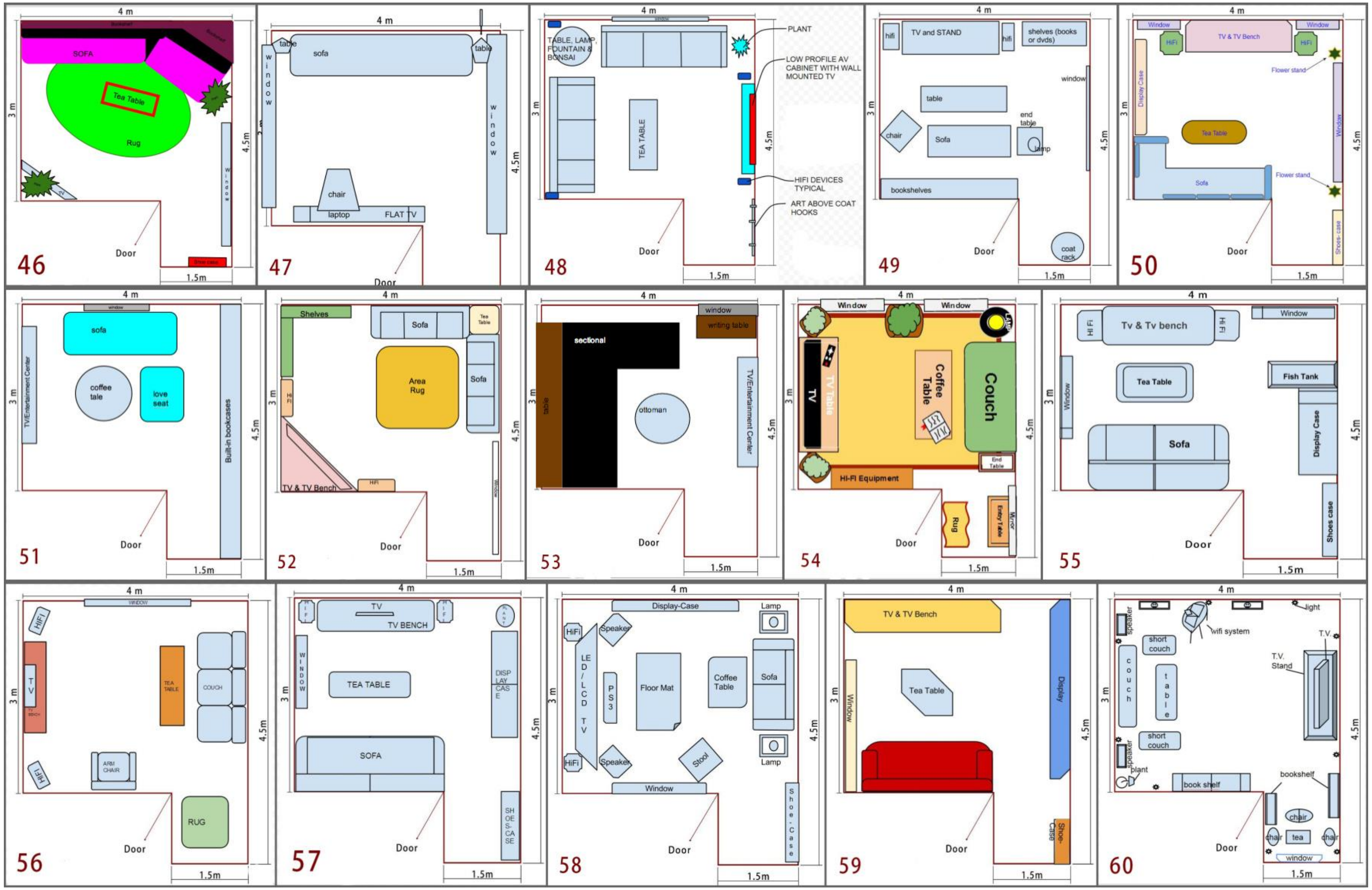


Appendix 1 - 1 Living room layout designs generated by MTurk workers in case study A

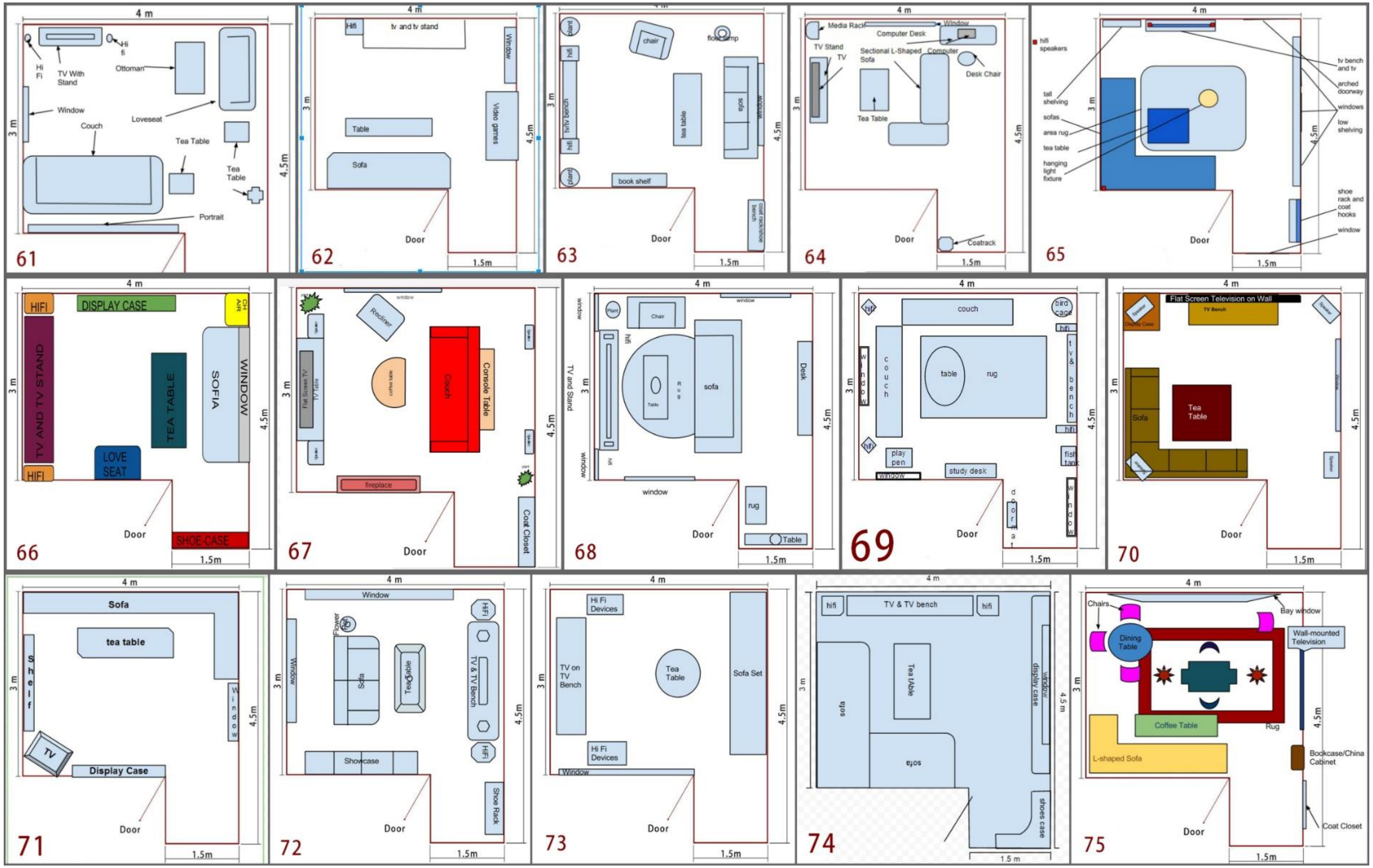
Appendix 1 - 2 Living room layout designs generated by MTurk workers in case study A



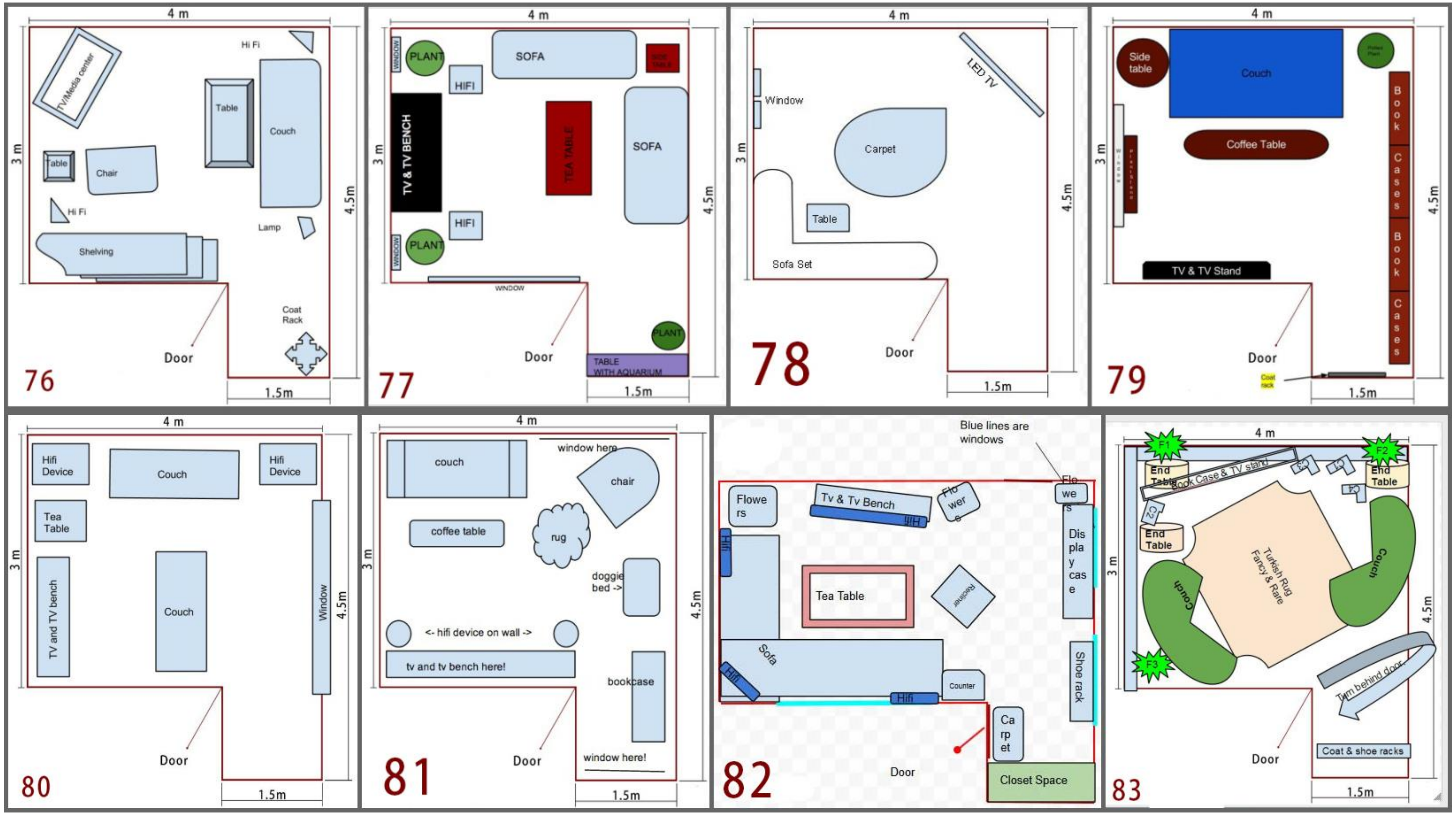
Appendix 1 - 3 Living room layout designs generated by MTurk workers in case study A



Appendix 1 - 4 Living room layout designs generated by MTurk workers in case study A

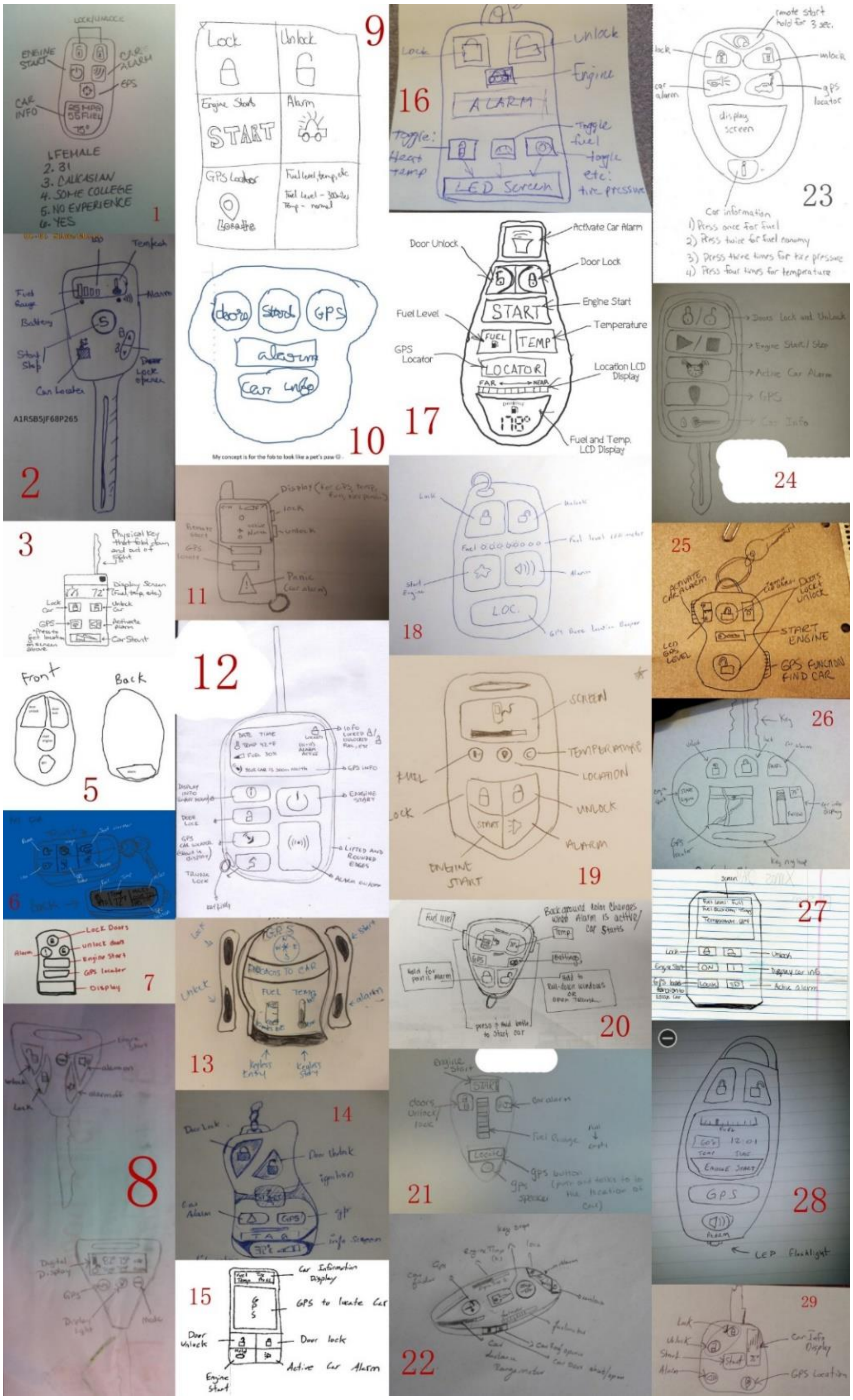


Appendix 1 - 5 Living room layout designs generated by MTurk workers in case study A

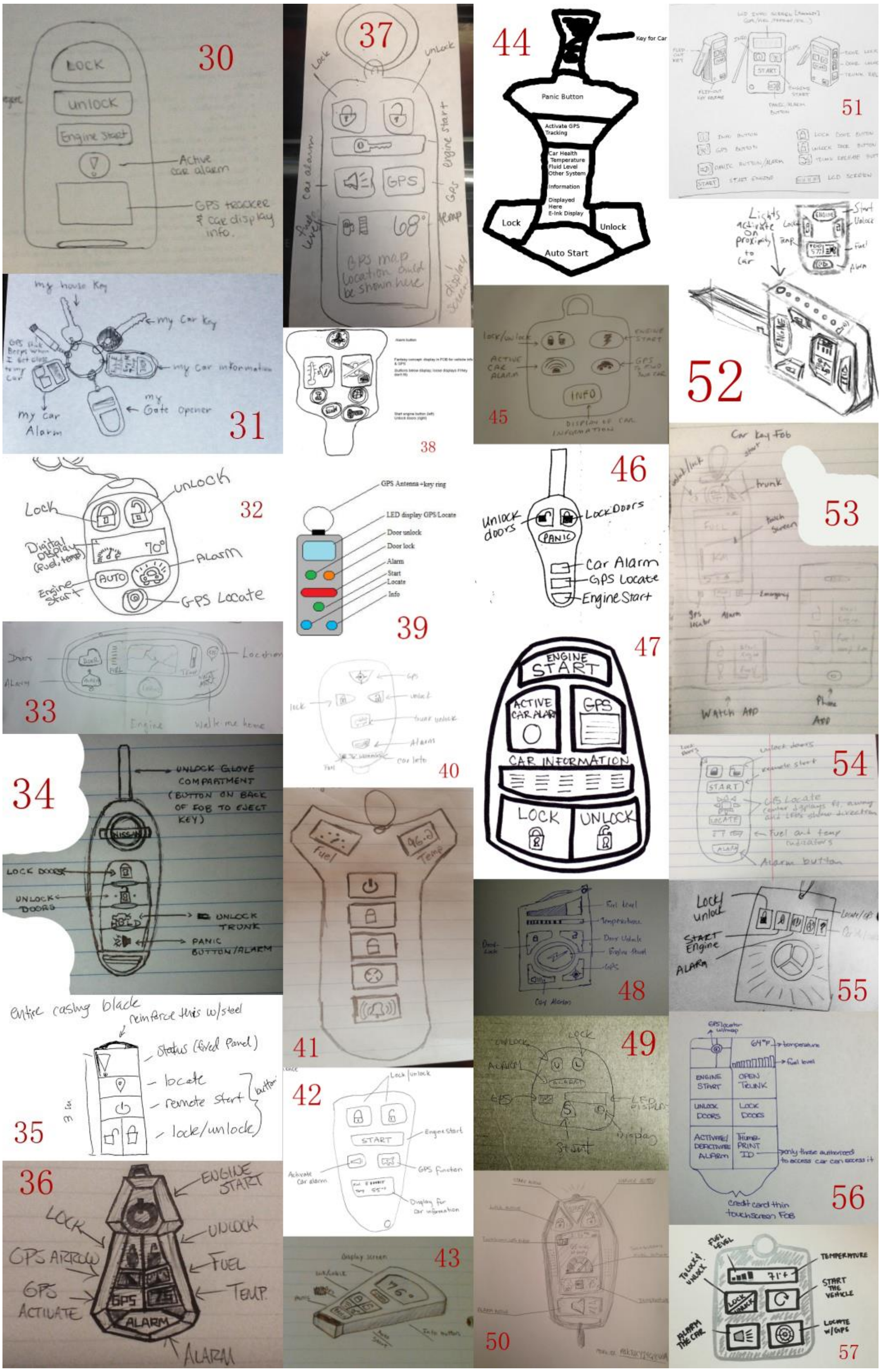


Appendix 1 - 6 Living room layout designs generated by MTurk workers in case study A

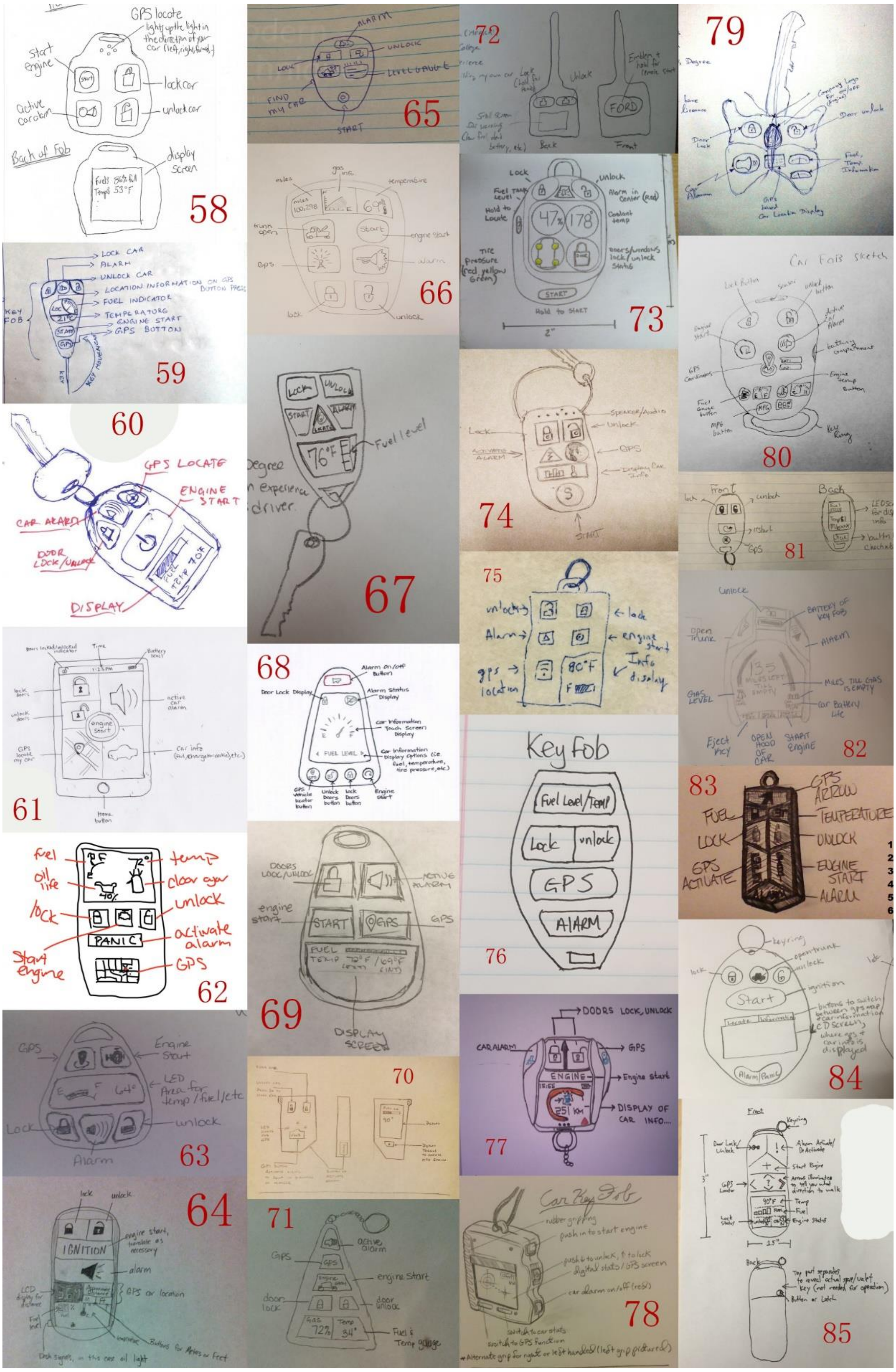
APPENDIX 2: All 1st generation car key fob designs generated by MTurk workers in case study D



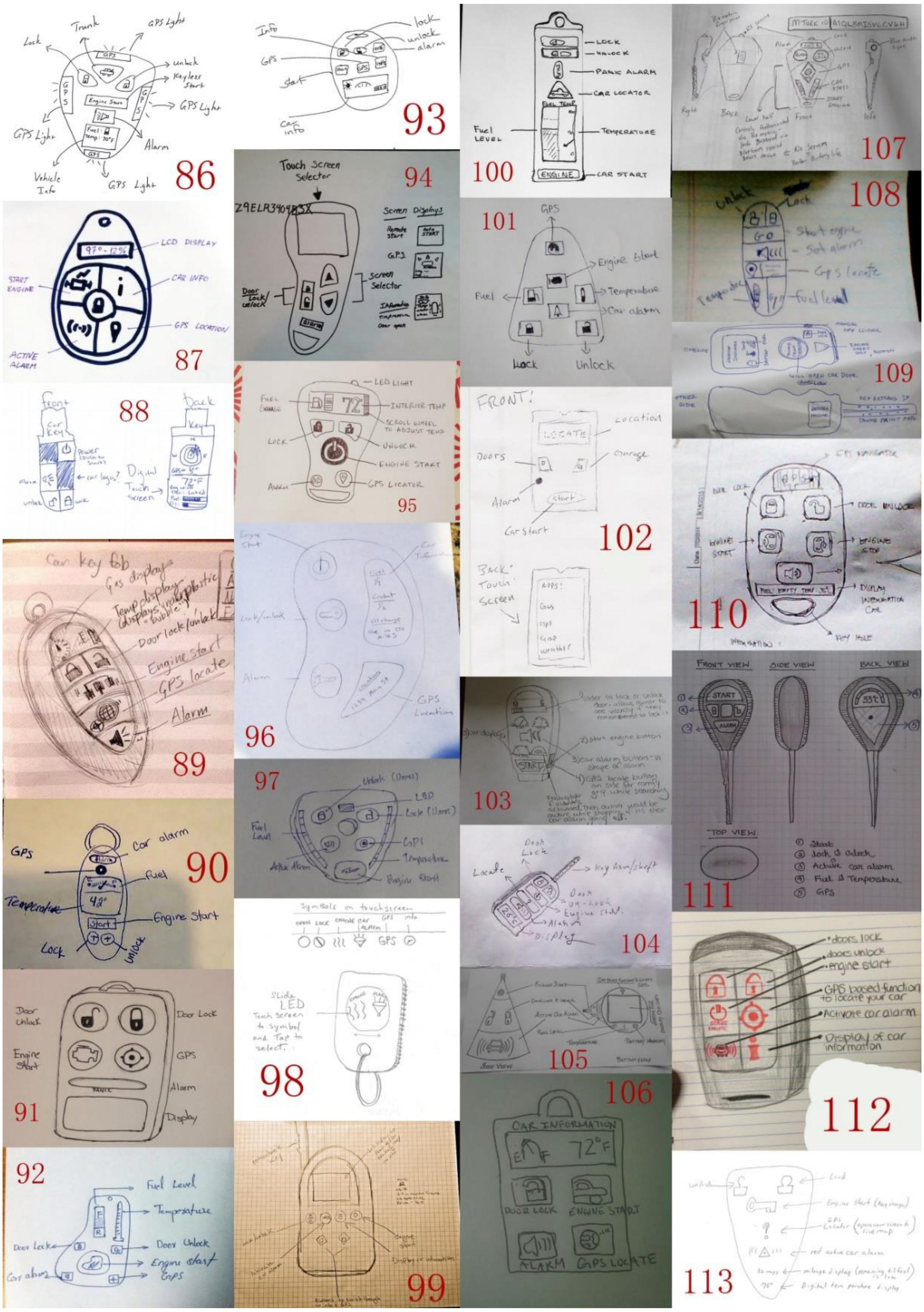
Appendix 2 - 1 All 1st generation car key fob designs generated by MTurk crowd workers in case study D



Appendix 2 - 2 All 1st generation car key fob designs generated by MTurk crowd workers in case study D



Appendix 2 - 3 All 1st generation car key fob designs generated by MTurk crowd workers in case study D



Appendix 2 - 4 All 1st generation car key fob designs generated by MTurk crowd workers in case study D

Appendix 2 - 5 All 1st generation car key fob designs generated by MTurk crowd workers in case study D

Appendix 2 - 6 All 1st generation car key fob designs generated by MTurk crowd workers in case study D

APPENDIX 3: A report of demographic information of crowd workers in this research

In crowdsourced design tasks, the human workers are critical to the outcome, so consequently their background (i.e., gender, location behaviour, etc.) needs to be understood. In this chapter, the results of the cDesign experiments presented in the previous chapters are discussed as a whole. Although the demographic information associated with individual experiments has been discussed in earlier chapters, the **integration of the demographic information** of all participants involved in the experiments is presented in the following sections. In total, over one thousand participants have been involved in the experiments, these workers can be characterized by the following properties:

1) Gender

2) Education Level (i.e., high-school diploma, Bachelor degree, etc.)

3) Nationality

4) Design Experience (i.e., whether they have design experience before doing the task, and how many years' experience they have)

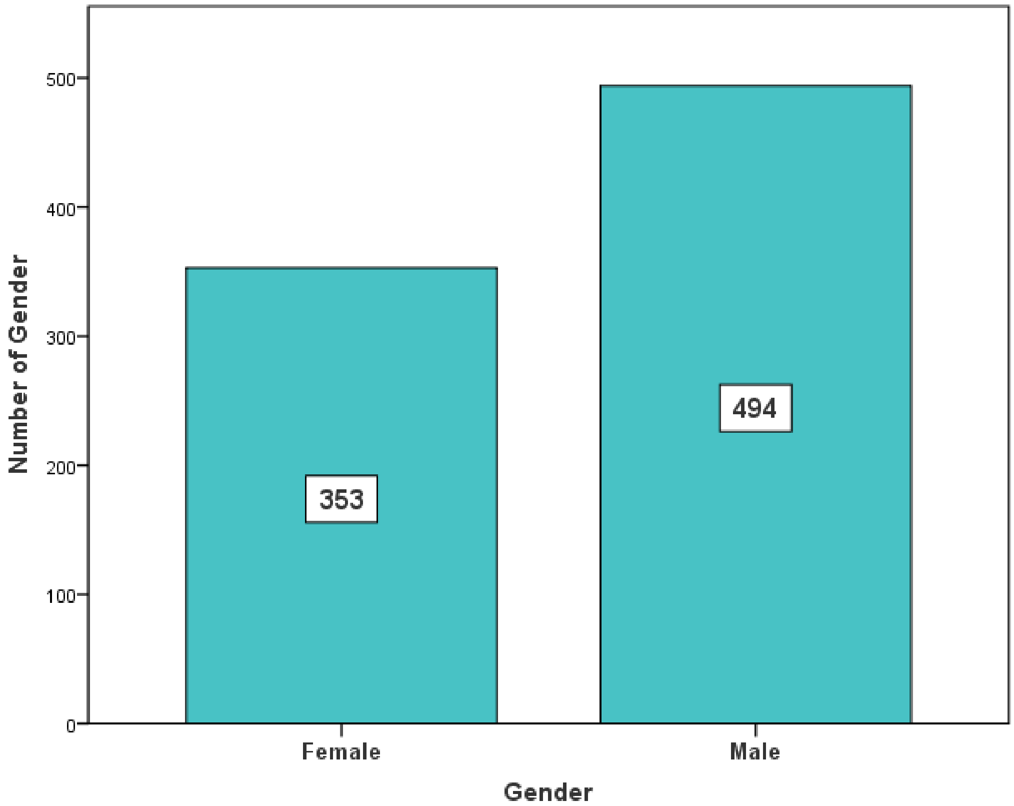
5) Response Time (i.e., the result was received in 24 hours, 48 hours, etc.)

6) Submission approved or rejected

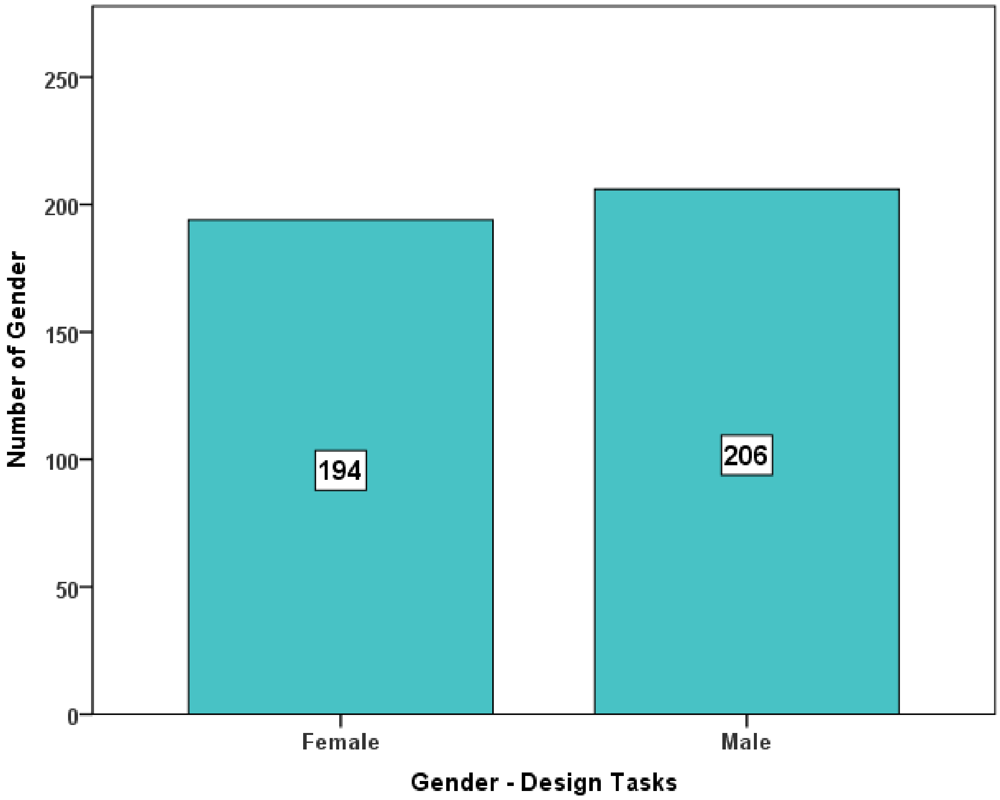
Understanding the nature of the crowd could provide insights that allow specific crowds to be assembled for specific jobs. It is also important to verify the nature of the crowd to enable comparisons to be made with other researchers' results (i.e., confirm approximate parity of age, gender and education, etc.).

1. Gender

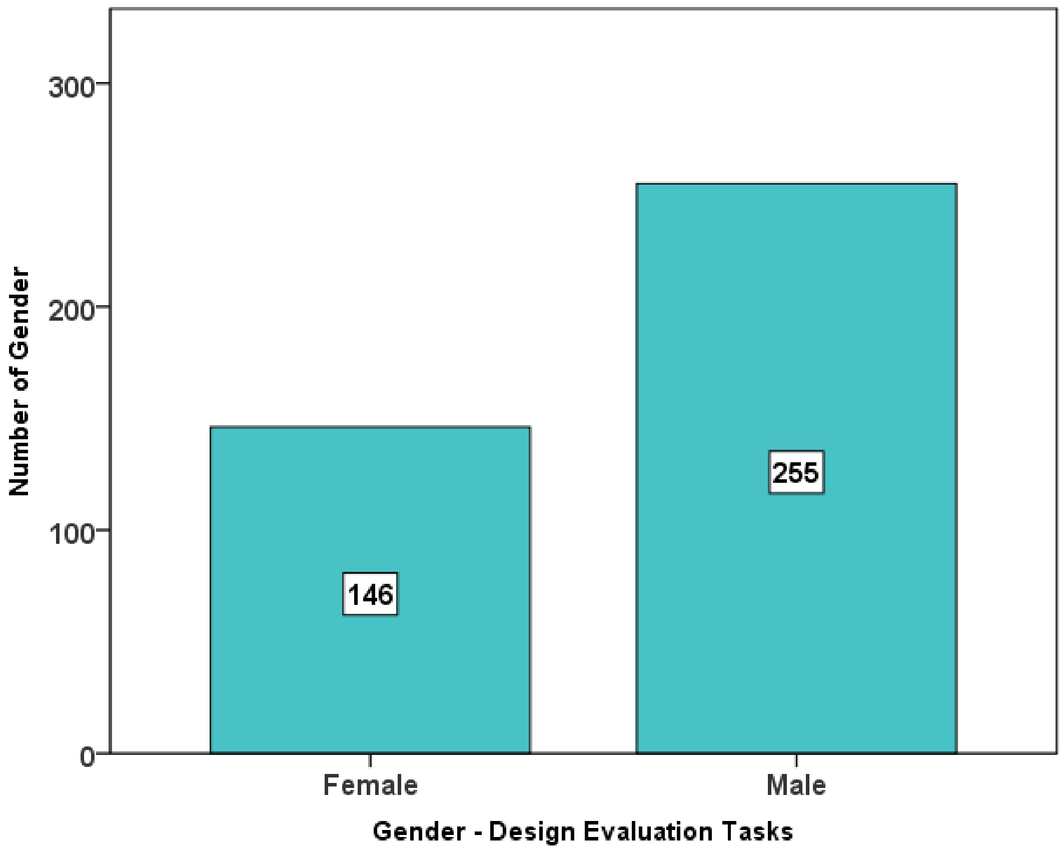
The gender of the crowd workers was supplied for all experiments with the exception of the 3D kitchen layout case study and is shown in Appendix 3- 1. In total, there were 847 participants who provided their gender information (41.7% were female and 58.8% were male). What is more, compared with the socio-demographic characteristics of MTurk from the Levay's research work (Levay et al. 2016): 46.1% were female and 53.9% were male, the presentation are close. Because broadly speaking the experiments can be divided into design tasks and design evaluation tasks, it is interesting to find the two tasks appeal equally to both sexes (i.e., overall demographic information, in design tasks, in design evaluation tasks). Appendix 3- 2 shows that female took 48.5%, and male took 51.5%. In design tasks, in other words the number of female participants are similar with the number of male participants. However, in design evaluation tasks, in Appendix 3- 3 it can be seen that males were 63.6%, and female were 36.4%. These number determines that on MTurk and ShortTask crowdsourcing platforms, more males are willing to evaluate designs than females.



Appendix 3- 1 Gender Proportion in all tasks



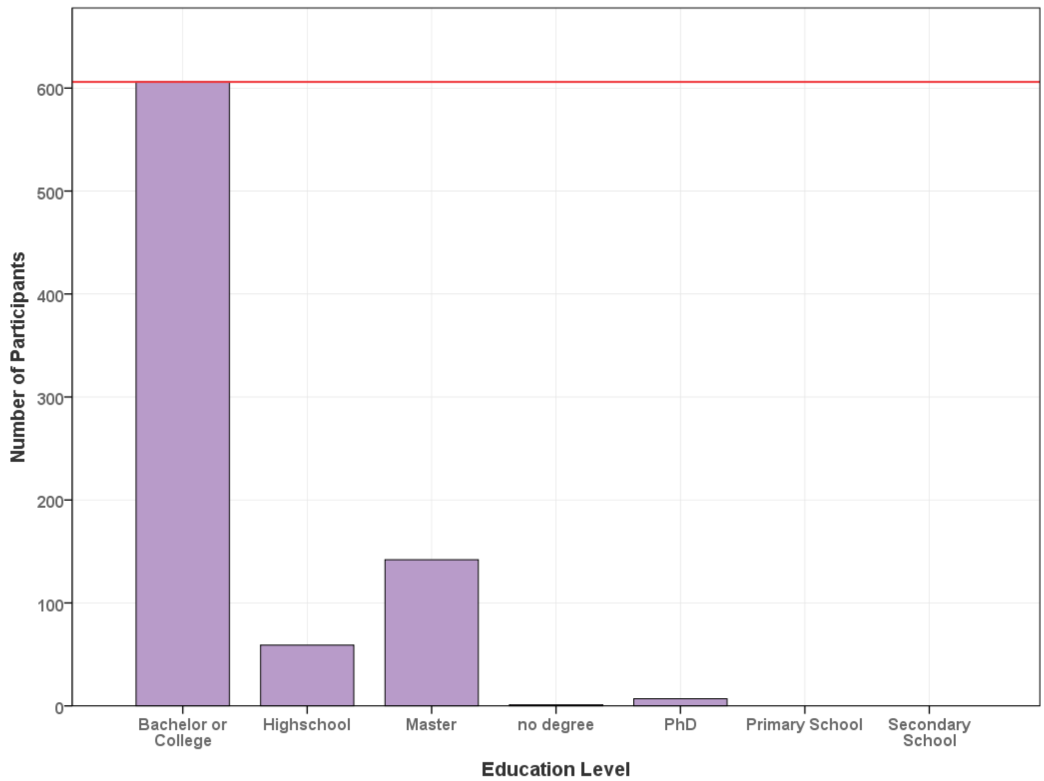
Appendix 3- 2 Gender proportion of design tasks



Appendix 3- 3 Gender proportion of design evaluation tasks

2. Education Level

The education level of all the workers involved in the experimental tasks (i.e., including design tasks, evaluation tasks and also some other tasks (i.e., specification collection tasks)) is shown in Appendix 3- 4: the number of participants with Bachelor or College degree – 606, Master degree – 142, High-school – 59, PhD – 7, no degree – 1, Secondary School – 0, Primary School – 0. This information shows that on the commercial crowdsourcing platform, 74.4% of the available crowd workers were educated to at least college level (indeed almost 1% had PhD degrees). Compared with Levay's work, 45% MTurk workers possess at least a bachelor's degree (but does not explain about the college degree or associate degree) (Levay et al. 2016). However, in Yu's children chair design task, 79% of participants had earned college or graduated degree which is similar with this research (Yu & Nickerson 2011). Furthermore, except for only one participant who have no higher education, the lowest education level was High school.



Appendix 3- 4 Education level in all experiments.

3. Nationality

The nationality of the 743 participants was also investigated. However, during the course of the research there was some variation in how this factor was assessed. In Case Study A and Case Study B, the question to participants in the task description was "Are you a native English Speaker?" instead of "Your Nationality". So, the collected answer from participants were "Yes" or "No" which is shown in the second and third row (in grey colour in **Error! Reference source not found.**). As a result, for the answer of "Yes" or "No", the number of the native English speakers were over two times the number of participants who were not the native English speakers.

Appendix 3- 5 All collected nationality information from crowd workers

Nationality	Number of Participants	Continent

<table>
<tr><td>Yes</td><td colspan="2">133</td><td></td></tr>
<tr><td>No</td><td colspan="2">65</td><td></td></tr>
<tr><td>African-American</td><td>2</td><td rowspan="2">337</td><td rowspan="2">North America</td></tr>
<tr><td>American</td><td>335</td></tr>
<tr><td>Argentinean</td><td colspan="2">3</td><td>South America</td></tr>
<tr><td>Asian</td><td colspan="2">2</td><td></td></tr>
<tr><td>Bahamian</td><td colspan="2">1</td><td>North America</td></tr>
<tr><td>Belgian</td><td colspan="2">10</td><td>Europe</td></tr>
<tr><td>Bosnia and Herzegovina</td><td colspan="2">2</td><td>Europe</td></tr>
<tr><td>Brazilian</td><td colspan="2">5</td><td>South America</td></tr>
<tr><td>British</td><td colspan="2">18</td><td>Europe</td></tr>
<tr><td>Canadian</td><td colspan="2">15</td><td>North America</td></tr>
<tr><td>Caucasian</td><td colspan="2">9</td><td></td></tr>
<tr><td>Croatian</td><td colspan="2">1</td><td>Europe</td></tr>
<tr><td>Dominican</td><td colspan="2">10</td><td>North America</td></tr>
<tr><td>Egyptian</td><td colspan="2">1</td><td>Africa</td></tr>
<tr><td>Filipino</td><td colspan="2">3</td><td>Asia</td></tr>
<tr><td>German</td><td colspan="2">2</td><td>Europe</td></tr>
<tr><td>Greek</td><td colspan="2">4</td><td>Europe</td></tr>
<tr><td>Hispanic</td><td colspan="2">7</td><td></td></tr>
<tr><td>Hungarian</td><td colspan="2">2</td><td>Europe</td></tr>
<tr><td>Indian</td><td colspan="2">127</td><td>Asia</td></tr>
<tr><td>Indonesia</td><td colspan="2">2</td><td>Asia</td></tr>
<tr><td>Irish</td><td colspan="2">5</td><td>Europe</td></tr>
<tr><td>Italian</td><td colspan="2">7</td><td>Europe</td></tr>
<tr><td>Latino</td><td colspan="2">1</td><td></td></tr>
<tr><td>Macedonian</td><td colspan="2">8</td><td>Europe</td></tr>
<tr><td>Mixed Race</td><td colspan="2">1</td><td></td></tr>
<tr><td>Moroccan</td><td colspan="2">1</td><td>Africa</td></tr>
<tr><td>Pakistan</td><td colspan="2">1</td><td>Asia</td></tr>
<tr><td>Persian</td><td colspan="2">1</td><td></td></tr>
<tr><td>Peruvian</td><td colspan="2">2</td><td>South America</td></tr>
<tr><td>Polish</td><td colspan="2">1</td><td>Europe</td></tr>
<tr><td>Portuguese</td><td colspan="2">1</td><td>Europe</td></tr>
<tr><td>Puerto Rican</td><td colspan="2">1</td><td>North America</td></tr>
<tr><td>Romanian</td><td colspan="2">3</td><td>Europe</td></tr>
<tr><td>Nationality</td><td colspan="2">Number of Participants</td><td>Continent</td></tr>
<tr><td>Russian</td><td colspan="2">2</td><td>Europe</td></tr>
<tr><td>Serbian</td><td colspan="2">1</td><td>Europe</td></tr>
</table>

South African	1	Africa
Spanish	7	Europe
Swedish	1	Europe
Turkish	1	Asia
Ukrainian	2	Europe
Venezuelan	10	South America
Total	**817**	

Some of the nationalities provided by the workers had to be excluded (i.e., Appendix 3- 5: Asian, Caucasian, Hispanic, Latino, Mixed Race and Persian), because they did not map exactly a name of country or state. In total, crowd workers participated in the experiments were from 37 different countries (Appendix 3- 5). Appendix 3- 8 shows Americans were the most frequent participants (337) with Indians workers (127) in second place, and British people (18) in the thirdly place. This is not surprising given that MTurk is an America based crowdsourcing platform, and most of the experiments were done on that platform. However, it is interesting to note that in the validation experiment – the evaluation task of creating a car key fob – the number of European crowd workers started to increase. Appendix 3- 9 shows the proportion of the continents with European countries being the most frequent (18 different European countries), which is followed by Asia and North America (5 countries for both). But from the view of the number of participants in different continents, in Appendix 3- 10, the number of participants in North America was the most – 364; the second place was the number of crowd workers in Asia – 134, and then in Europe – 77. As a conclusion of the nationality information, it is clear that most participants came from America in North America, but there was no participant in Oceania (i.e., Australia) and Antarctica. Additionally, in Levay's report, only the percentage of Whites (71.8%), Blacks (7.1%) and Hispanics or Latino/as (5.6%) were collected. What is more, Yu pointed that 59% participants were native English speakers (Yu & Nickerson 2011), but did not collected the nationality information either. After the discussion about the nationality of participants, the next section will describe the design experience information.

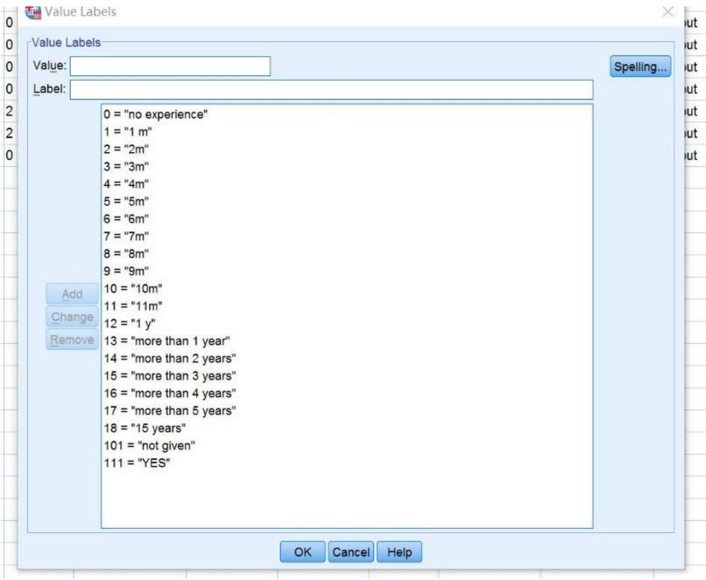
4. Design Experience

In the questionnaire associated with the tasks, crowd workers were asked:

Do you have any design experience before?

i.e., no experience, one-year design experience

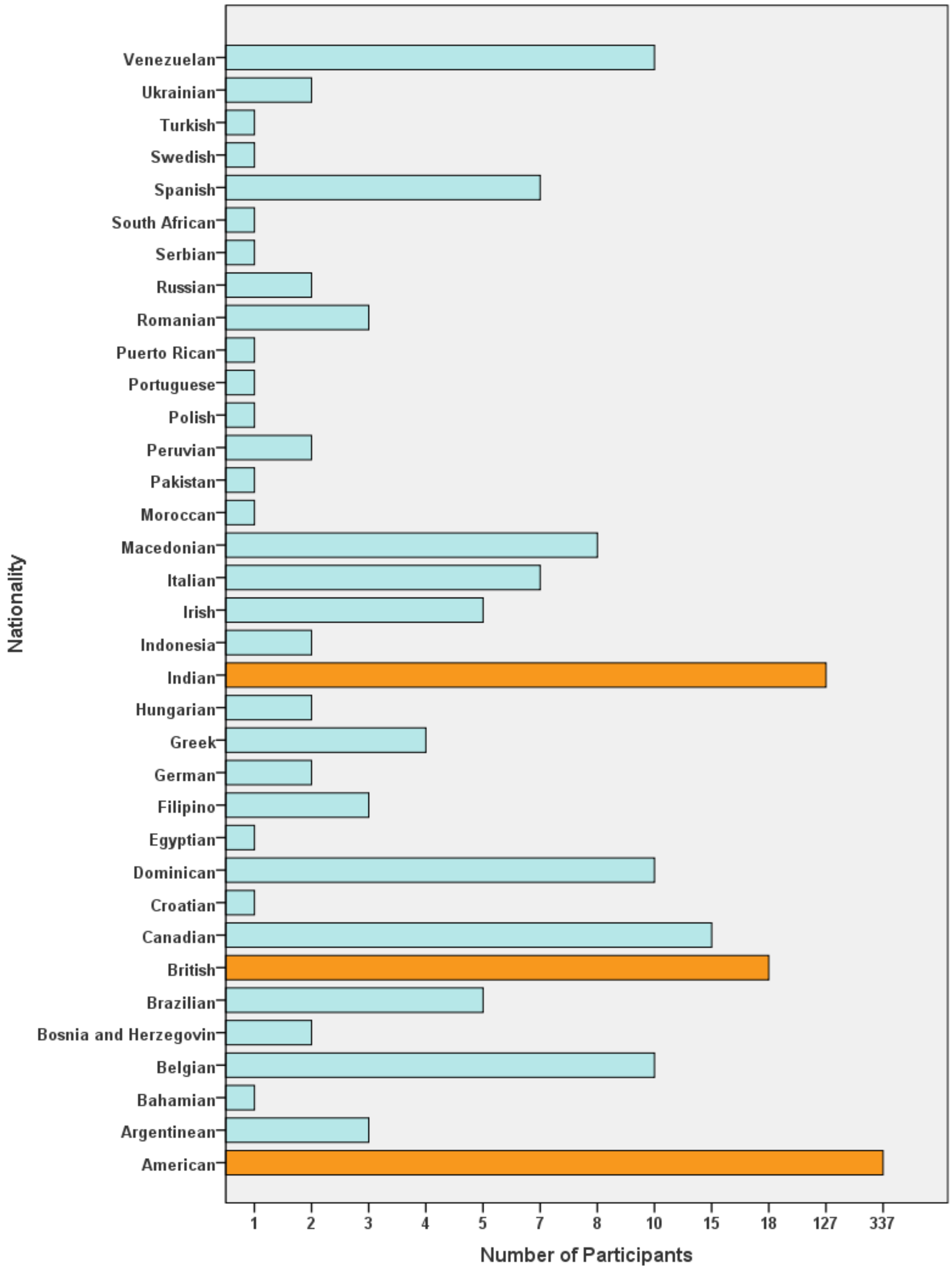
Based on the participants' responses, the answers were as follows: no experience (complete have no design experience before), have experience (have experience, but did not provide the exact time, i.e., one-year), one month, two months, three months, six months, eleven months, one-year, more than one-year, more than two years, more than three years, more than four years, more than five years and fifteen years (an example of the format of collection form is shown in, and an example of the collection form is shown in). In total, 786 participants provided their design experience information. The results (illustrated in Appendix 3- 12) suggested that 65.1% (512/786 * 100%) of crowd workers did not have any design experience before participating the tasks. Similarly, Yu's reported research work also suggested that only 20% of participants had 1-3 years of design experience (Yu & Nickerson 2011). This result supports the assertion that many people without formed training can contribute to creative design tasks. In terms of the participants who had design experience, one-year design experience was the most collected. More detailed information is shown in Appendix 3- 11. In other words, if the appropriate design method can be applied in the cDesign process, the potential of a crowd with mixed abilities can collaborate effectively in design tasks. The next section will discuss about the submissions time – when participants submitted their results (i.e., within 24 hours, within 48 hours).



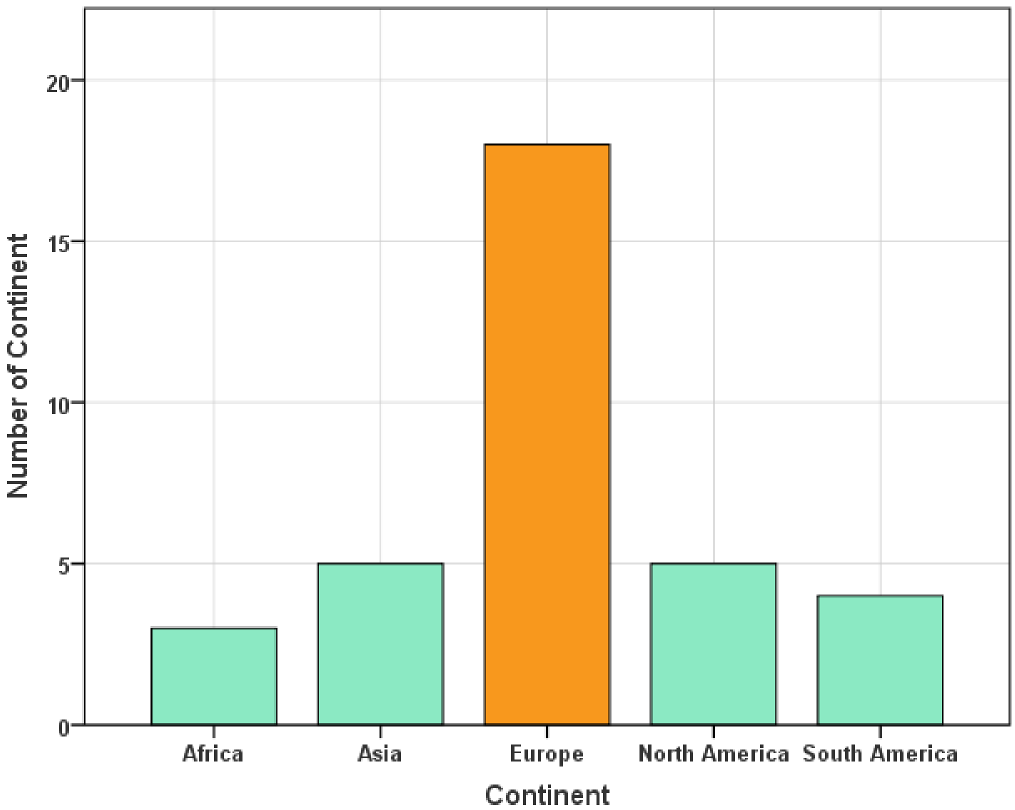
Appendix 3- 6 An example of the design experience options of collection form

	WorkerNumber	Gender	Age	NativeEnglishSpeaker	Education	DesignExperience	Drive	Remark
1	A2ABSQBRUAUS4C	Male	27	American	Bachelor or College	YES	yes	Approved
2	A1Z2F9AYM401YC	Male	28	American	Bachelor or College	no experience	yes	Approved
3	AZTWHK37CN5RB	Male	29	Venezuelan	Bachelor or College	more than 5 years	yes	Approved
4	ARLGZWN6W91WD	Female	28	American	Bachelor or College	no experience	yes	Approved
5	A30BLOFPF8HRU9	Male	23	American	Bachelor or College	no experience	yes	Approved
6	A2VJAAYE13IXNQ	Female	29	Indian	Bachelor or College	YES	yes	Approved
7	AZTWHK37CN5RB	Male	29	Venezuelan	Bachelor or College	more than 5 years	yes	Approved
8	AZTWHK37CN5RB	Male	29	Venezuelan	Bachelor or College	more than 5 years	yes	Approved
9	AEMAUIVLD8GLY	Male	41	American	Bachelor or College	15 years	yes	Approved
10	A12X43VFAKMR89	Male	30	Macedonian	Bachelor or College	YES	yes	Approved
11	A1U7W3010U1T49	Male	23	American	Bachelor or College	no experience	yes	Approved
12	A142196DZICV3F		.			.		
13	A2P44QU8NYNKGR	Male	24	American	Bachelor or College	6m	yes	Rejected
14	A2BCAXZH5XS39E	Female	28	Dominican	Bachelor or College	YES	yes	Approved

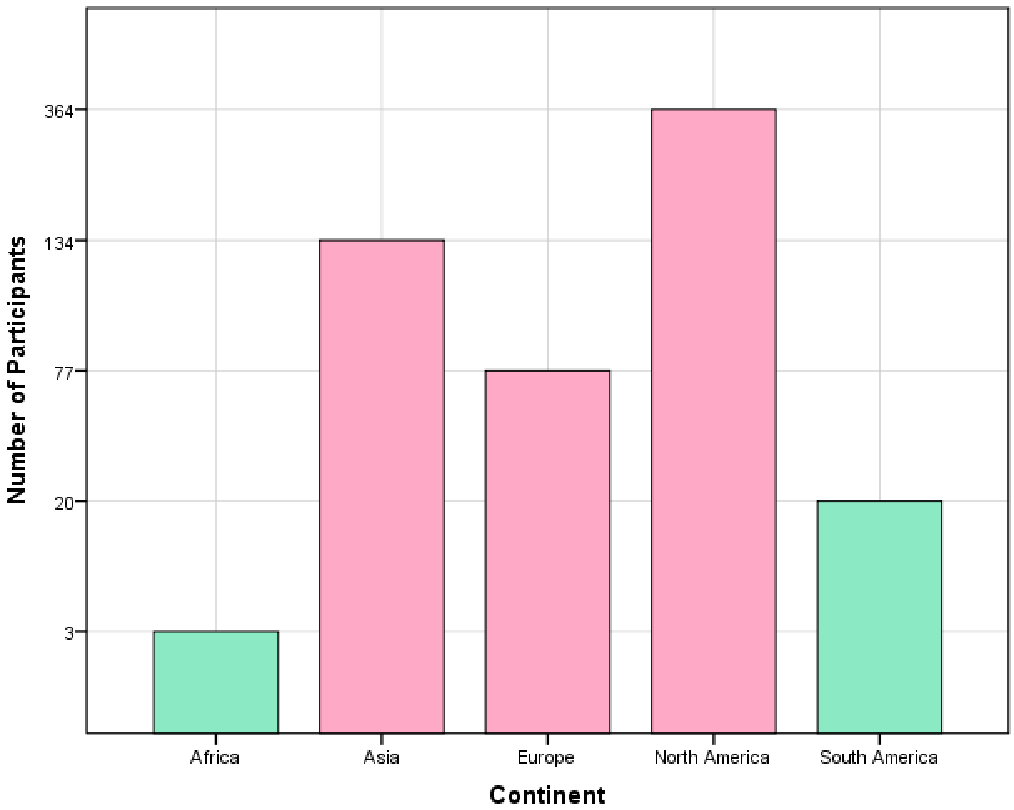
Appendix 3- 7 An example of the collection results form



Appendix 3- 8 Exact nationality of crowd worker in all experiments



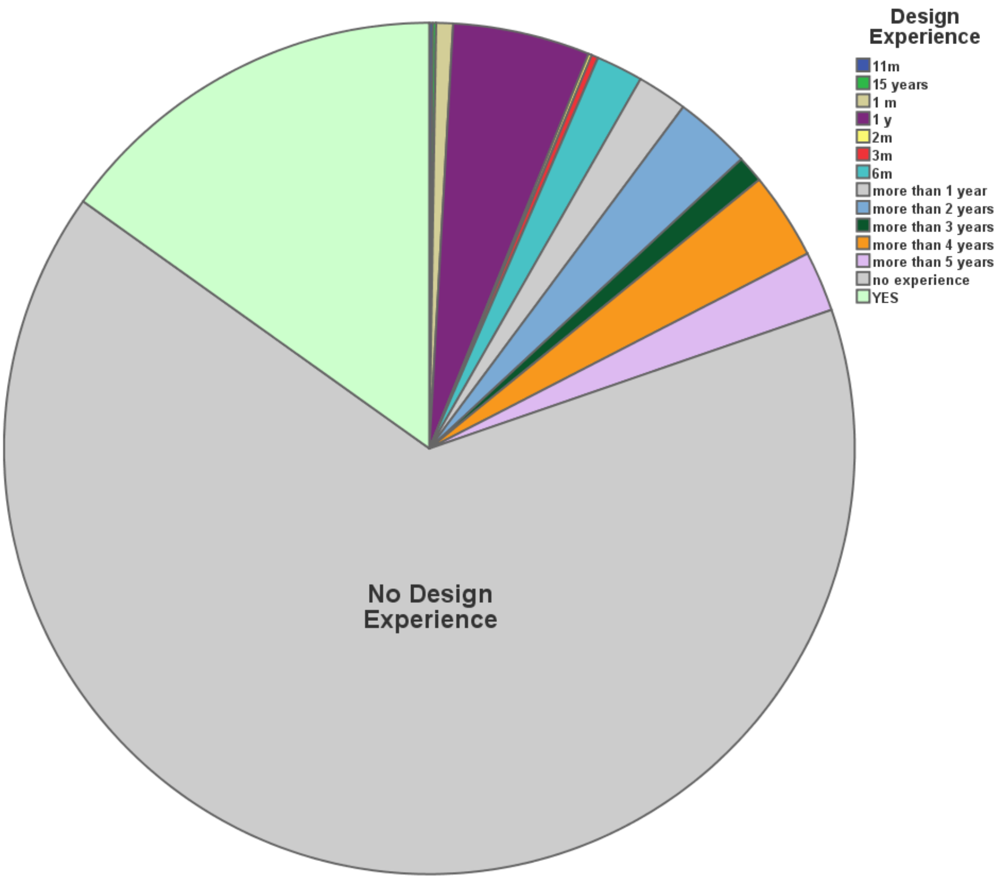
Appendix 3- 9 Continent of crowd workers



Appendix 3- 10 Number of participants in different continents

Appendix 3- 11 Design experience information of participants

	No. of Participants	Percent (%)
1 month	5	.6
2 months	1	.1
3 months	2	.3
6 months	14	1.8
11 months	1	.1
1 year	41	5.2
more than 1 year	15	1.9
more than 2 years	23	2.9
more than 3 years	8	1.0
more than 4 years	26	3.3
more than 5 years	18	2.3
15 years	1	.1
YES	119	15.1
No experience	512	65.1
Total	786	100.0



Appendix 3- 12 All design experience proportion

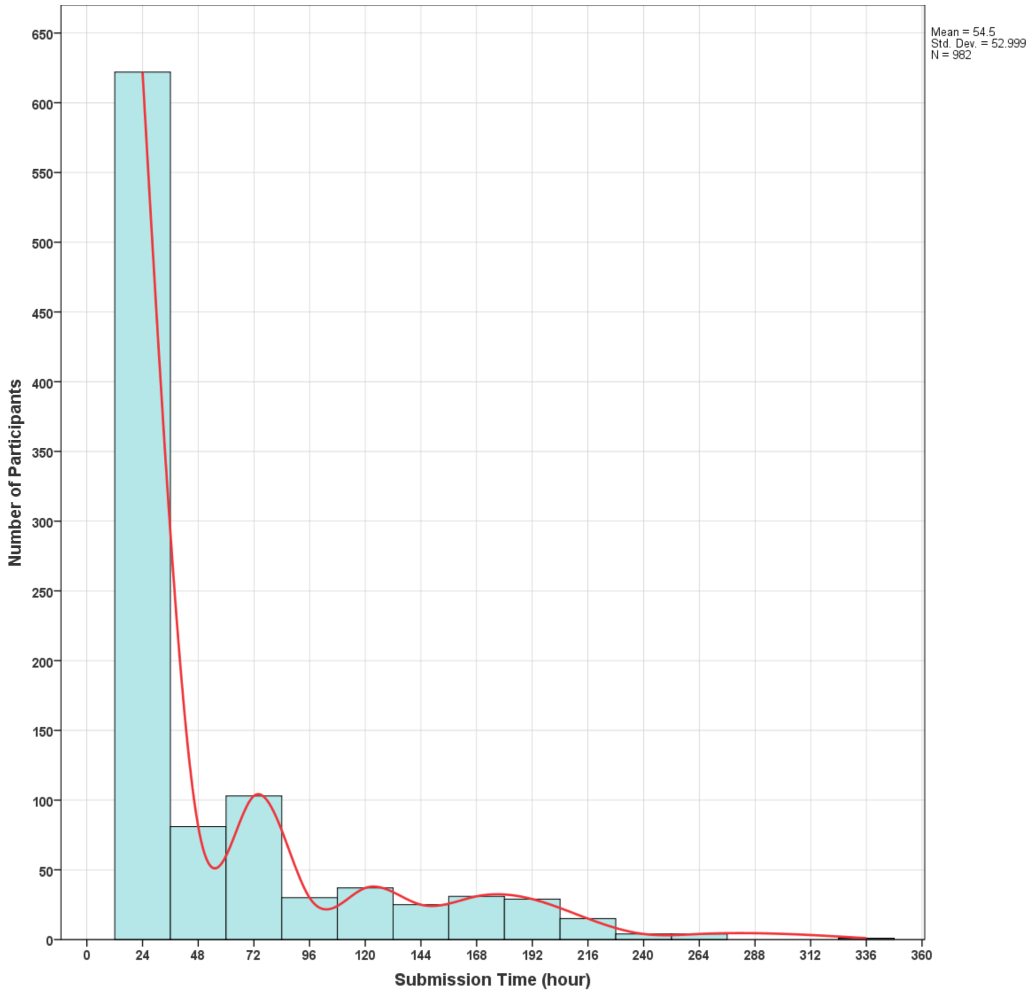
5. Submission Time

In this section, the submission time (i.e., the time taken by an online worker to complete a task) will be discussed. As described in the living room layout design task, the level of the submission time was set as 24 x n (hour) (i.e., n = 1, 2, 3, …). Understanding the issues around setting the time of submission is an important part of task design and can help requesters to manage their tasks (i.e., determines how long time they can receive the results after posting tasks on the crowdsourcing platforms). Because the research experiments include questionnaire tasks, design tasks, design evaluation tasks, vote and rank tasks, specification collection tasks (design evaluation criteria collection tasks), these various categories of tasks can provide a broad view of response times (i.e., rather just focusing on design tasks of questionnaire tasks). Appendix 3- 13 shows that the submission time with 24 hours (i.e., results submitted during the first day after posting the tasks) are the most common (63.3%), and only one submission was received by 336 hours (two weeks) (although over 1000 worker participated in the tasks, it was not possible to collect all the submission times. But 982 instances collected still provide a useful insight).

The histograms and the distribution curve of the submission time information is illustrated in Appendix 3- 14. It clearly shows that in the first 24 hours, a requester can except to receive the most results, and with time going, the trend of submission decreased. Indeed, after two weeks, there were no further submission received. This distribution determines that for requesters posting tasks on crowdsourcing platforms, the most likely time to collect enough solutions of their problems is in the first 24 hours. If requesters increase the tasks' time (i.e., the task available time to participants) over 24 hours, they could stop the tasks by the third day, because after 72 hours the submission will decrease dramatically. In the next section, the last information will be discussed which is the "acceptance" of the submitted results by the requester. In this research, the response time in design tasks on MTurk was first collected in this research area.

Appendix 3- 13 Submission time information

Submission Time (hour)	Frequency	Percent (%)
24	622	63.3
48	81	8.2
72	103	10.5
96	30	3.1
120	37	3.8
144	25	2.5
168	31	3.2
192	29	3.0
216	15	1.5
240	4	.4
264	4	.4
336	1	.1
Total	982	100.0

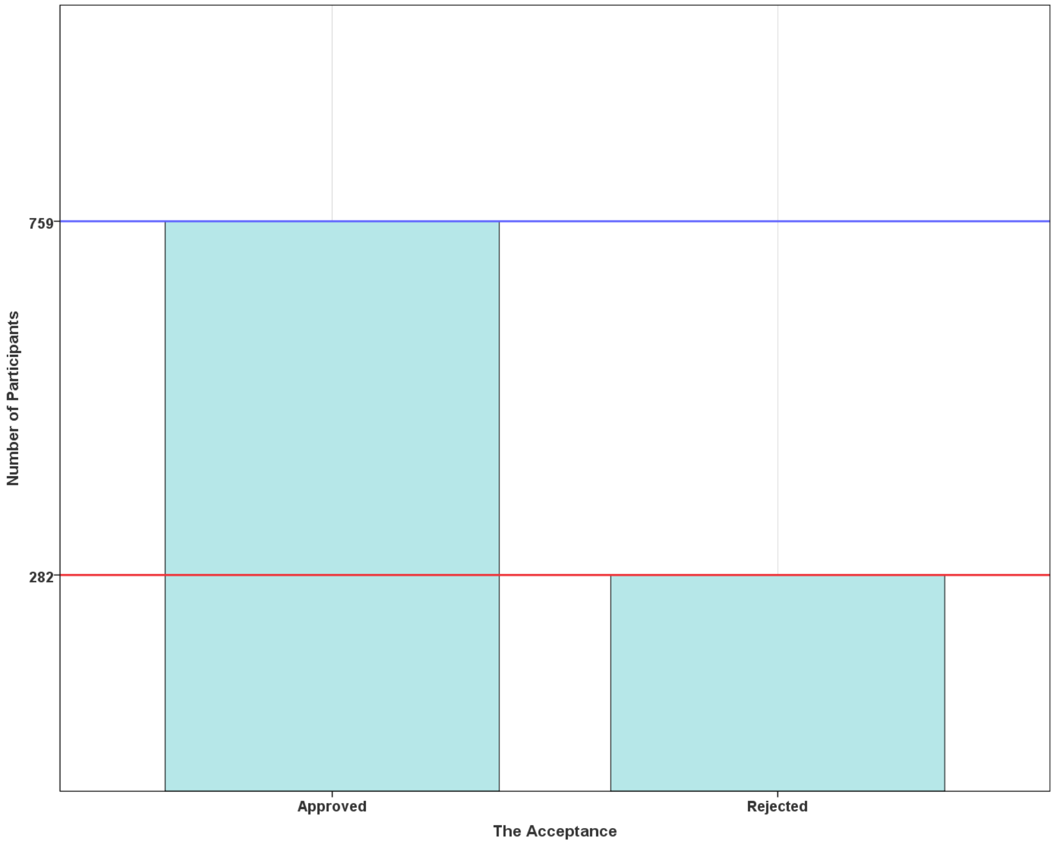


Appendix 3- 14 Submission time information and the distribution curve

6. The Acceptance for Submissions

Although the tasks posted during the course of this research generated large numbers of approved design (i.e., drawings or layouts, design evaluation results and other information), there was also amount of work rejected for payment created in the same time. Consequently, it is valuable to investigate the rate on the accepted and rejected submissions for insights in to provide strategy to requesters to avoid the cheats or decrease the rejected submissions. In total, the approved results were 759 and the rejected submissions were 282 (in Appendix 3- 15). This result suggests that over 25% of the submissions were rejected by the requesters. Additionally, in Eickhoff's experiment, 37.3% of submissions were rejected before using a method (i.e., gold standard data) to avoid cheats (Eickhoff & Vries 2012). After the analysis of the rejected results, the following categories of the rejected results can be observed:

1. Cheats – completely wrong answers for the tasks (submit only to gain payment)
2. Misunderstand the task meaning – wrong answers (some parts of the answer were right)
3. The result does not meet the task requirements (some parts of the answer were right)



Appendix 3- 15 The acceptance information

The cheating submissions are usually rejected by the requester immediately, and the ID of the crowd workers is flagged to avoid them to participate in the tasks again. Indeed, the task description always includes a sentence to warn people who are willing to cheat that if the submissions are identified as cheats, workers will not be allowed to participate in the tasks. But because of the nature of the crowd and the crowdsourcing platforms, requesters can never completely avoid the cheats. In contrast, for participants who misunderstood the meaning of the task, or made small mistakes of their results (for example, the evaluation method should be 7-point Likert Scale, but workers use 5 or 10-pioit Likert Scale method), requesters can send messages to these workers and communicate with them to change the wrong parts of their results and submit them again by emails (because once a result is submitted, the participate cannot change or resubmit new result on MTurk platform). Based on the experience during the experiments, the following suggestions can be given to try to avoid cheats during design tasks:

1. Write the task description as clearly and understandable as possible, because a readable and understandable task description will be easy for participants to carry out the results.
2. Warn cheats in the task description. For example, if participants cheat the requester, their ID will be tagged and noticed to the crowdsourcing platform managers.
3. Communicate with participants. Communication can provide chances for participants who make mistakes in tasks, and these mistakes can be corrected.

Report Summary

In this report, the demographic information collected from all case studies has been discussed. Firstly, the gender of the crowd workers was described. Results show that both in design and design evaluation experiments, the proportion of males are more than females. One possible reason is that MTurk (and the similar crowdsourcing platforms) the male crowd workers who are interested in design and other jobs related with design are more than female crowd workers. In the future, if the number of females willing to designs increases, the gender proportion might be changed. Perhaps it is a question of perception (i.e., how women perceive themselves as designers).

Secondly, the education background of participants was discussed (i.e., high-school, bachelor degree, etc.). Collected results suggest that almost 75% of participants (who provided their individual information have college or higher degrees). And almost all participants were educated over high-school degree (only one participant did not have any degree). This suggests that the quality of the crowd is high and not represented of the population as a whole.

Thirdly, in terms of the nationality of crowd workers, the experiments involved a various range of participants who came from 37 different countries. This number guaranteed the diversity of crowd workers. However, the majority of participants were from one or two countries. However, although relatively small, this diversity could provide alternative views and inspirations for innovations in design tasks. What is more, the nationality information also presents the power of crowdsourcing that all internet users can work together and exchange knowledge with each other.

Fourthly, the design experience information was described after the nationality information. For one hand, results determined that although over half of the crowd workers having no design experience before participating the experiments, they can make an effective contribution to the design and evaluation of designs. For another, over 1/3 participants had design experience, and could use their knowledge and design skills in the experimental case studies. These give two pieces of advices that although participants have no design experience, if the appropriate method are used, designs can be created; and the potential power of the crowd is unlimited, because people will never completely investigate the limits of crowd workers' ability.

Fifthly, a useful information was discussed regarding the time for submissions received. This information provides the following two suggestions: in the first place, during the first 24 hours after posted the tasks on the crowd sourcing platforms, requesters receive the largest number of submissions; furthermore, with time going, the general trend of the number of submission will decrease.

Finally, the author integrated the acceptance information of submissions. This information can help requesters to plan strategies to avoid submissions of cheats. Results show that the rejected submissions were less than 50% of approved submissions. Additionally, three suggestions were given to avoid cheats.

Reference

Eickhoff, C. & Vries, A. de, 2012. Increasing cheat robustness of crowdsourcing tasks. *Information Retrieval*, 16(2), pp.121–137. Available at: http://link.springer.com/10.1007/s10791-011-9181-9 [Accessed June 22, 2013].

Levay, K.E., Freese, J. & Druckman, J.N., 2016. The Demographic and Political Composition of Mechanical Turk Samples. *SAGE Open*.

Yu, L. & Nickerson, J., 2011. Cooks or cobblers?: crowd creativity through combination. *CHI 2011, May 7–12, Vancouver, BC, Canada*. Available at: http://dl.acm.org/citation.cfm?id=1979147 [Accessed June 25, 2013].