

8-2016

Understanding the Role and Importance of Design Problems in Creativity Research

Varun Kumar

Clemson University, kumar3@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Recommended Citation

Kumar, Varun, "Understanding the Role and Importance of Design Problems in Creativity Research" (2016). *All Theses*. 2460.
https://tigerprints.clemson.edu/all_theses/2460

This Thesis is brought to you for free and open access by the Theses at TigerPrints. It has been accepted for inclusion in All Theses by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

UNDERSTANDING THE ROLE AND IMPORTANCE OF DESIGN PROBLEMS IN
CREATIVITY RESEARCH

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Mechanical Engineering

by
Varun Kumar
August 2016

Accepted by:
Dr. Gregory Mocko, Committee Chair
Dr. Joshua Summers
Dr. Cameron Turner

ABSTRACT

The overall objective of this research is to address the need for using similar conceptual design problems in experiments in engineering design creativity. This is accomplished by addressing three sub-objectives i) to identify the pattern of design problem usage, ii) to enable comparison between two conceptual design problems based on their natural language representations and iii) to analyze the impact of design problems on effectiveness of example interventions used in user studies in engineering design creativity.

Design problems are an essential component of experiments in creativity research. The requirements of experiment's design sometimes limit problem sharing between researchers or studies conducted by them. For understanding and identifying the design problem usage pattern, two network representations of design problems, connected to each other by authors and papers using them has been used. Both networks indicate that several problems have been used for creativity experiments and suggest the need for using same or 'similar' design problems to reduce between-study differences in design problem usage. This addresses the first objective of identifying pattern of design problem usage in creativity research.

Problem similarity is assessed using two methods. The first method is based on identification of five structural elements of a design problem namely goals of a problem, functional requirements, non – functional requirements, reference to an existing product and information about end user. The protocol for identifying these elements in problem

statement and then comparing design problems is illustrated through two examples. The second method for similarity assessment is based on Latent Semantic Analysis (LSA) of problem statements. LSA provides an objective method to compare semantic similarity of problem statements. Both methods help address the research objective of comparing problems based on their representation but fail to evaluate problem solvability.

For understanding whether design problems influence the effectiveness of examples used as interventions, a meta-regression model between effect size and problem size has been used. Regression models suggest that problem size might have a linear relationship with effectiveness of examples for quantity of ideas produced by treatment group participants but enough evidence did not exist to suggest similar relationship for metrics quality and novelty. This addresses the sub-objective of design problems affecting the effectiveness of methods tested in experiments and overall objective of the need for using similar problems in creativity research.

DEDICATION

I dedicate this to my parents, grandparents and my best friend who have motivated and supported me in all my endeavors.

ACKNOWLEDGMENTS

I wish to thank my advisor, Dr. Gregory Mocko for his support and advice throughout my master's research. His enthusiasm and generous guidance have motivated me to strive for the best. Under him, I have grown as a student, engineer and researcher.

I would also like to thank Dr. Joshua Summers and Dr. Cameron Turner for providing valuable feedback on my research and serving on my committee. Their critical questions helped me to discover gaps and limitations in my work and has developed an open mind to accepting critical remarks as opportunities for improvement.

I thank all my fellow companions in CEDAR for their friendship and collaboration. Their friendship has been a memorable and joyous experience, something I will cherish all my life. Lastly, I would like to thank the all mighty for blessing me with strength and dedication through all these years that I have spent away from my family and siblings.

TABLE OF CONTENTS

	Page
TITLE PAGE	i
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	xii
 CHAPTER ONE: INTRODUCTION	 1
1.1. Design problems in creativity research.....	1
1.2. Research motivation	4
1.3. Research objectives	6
1.4. Key research tasks	7
1.5. Overview of thesis	7
 CHAPTER TWO: LITERATURE REVIEW.....	 9
2.1. Characterizing design problems: a historic overview	9
2.2. Methods for design problem evaluation	21
2.3. Differences between user studies in design creativity	27
2.4. Research questions	37
 CHAPTER THREE: RESEARCH APPROACH.....	 40
3.1. Graph based analysis of design problem usage	41
3.2. Comparison of design problems based on structural elements.....	41
3.3. Comparison of design problems based on Latent Semantic Analysis.....	42
3.4. Meta –regression analysis with problem size as potential moderator	43
 CHAPTER FOUR: DESIGN PROBLEM USAGE GRAPHS.....	 45
4.1. Graph based representation of design problem usage	45
4.2. Bipartite graph between different researchers and design problems	47
4.3. One mode representation of design problem graph connected by researchers	50
4.4. Bipartite graph between design problems and papers using them.....	54
4.5. Discussion	54
4.6. Comments and recommendations	57

CHAPTER FIVE: ASSESSMENT OF DESIGN PROBLEM SIMILARITY	59
5.1. Approach 1: Estimating size of a design problem by identifying structural elements of design problem	60
5.1.1. Examples showing identification of elements in design problems	64
5.1.2. Protocol for identification of elements in design problems	65
5.1.3. Inter-rater reliability test for approach one	67
5.1.4. Discussion for approach 1: similarity assessment by identification of structural elements in design problems	71
5.2. Approach 2: Similarity assessment through semantic analysis of design problems	73
5.2.1. Some measures for sentence level semantic similarity assessment	74
5.2.2. The LSA approach to problem similarity evaluation	77
5.2.3. Discussion for LSA results (approach 2)	78
5.2.4. Graphical representation of semantically similar problems	79
5.2.5. Limitations of approach 2 (LSA)	84
5.3. Possible application of the two similarity assessment approaches in creativity experiments	85
5.4. Comments and recommendations	90
CHAPTER SIX: META-ANALYTICAL REVIEW OF CREATIVITY STUDIES	93
6.1. Meta-regression analysis in creativity studies	93
6.2. Basics of meta-analysis	94
6.3. Models for effect size estimation	96
6.3.1. Fixed effect model	97
6.3.2. Random effect model	98
6.4. Heterogeneity in effect size	99
6.5. The meta-regression approach	101
6.6. Steps for meta-regression analysis of user studies using examples as intervention	103
6.6.1. Step 1: Collecting studies for meta-regression	103
6.6.2. Step 2: Extracting information	106
6.6.3. Step 3: Effect size estimation	108
6.6.4. Step 4: Estimation of variance in effect size	109
6.7. Meta-regression analysis with problem size as moderator	110
6.7.1. The meta-regression model	112
6.7.2. Quantity	113
6.7.3. Quality	116
6.7.4. Novelty	119
6.7.5. Verifying assumptions in regression models	120

Table of contents (continued)

	Page
6.7.6. Discussion of results	121
6.8. Influence of requirement type on example interventions	122
6.8.1. Quantity	123
6.8.2. Quality	124
6.8.3. Novelty.....	126
6.8.4. Discussion of results	128
6.9. Comments and recommendations	128
 CHAPTER SEVEN: CONCLUSIONS AND RECOMMENDATIONS	 130
7.1. Pattern of design problem usage (RQ1)	130
7.2. Design problem similarity assessment (RQ2 and RQ3).....	132
7.3. Influence of design problems on effectiveness of example intervention (RQ4).....	134
7.3.1. Influence of functional and non-functional requirements on effectiveness of examples	135
7.3.2. Contribution to overall research objective	136
7.4. Overall research conclusion	137
7.5. Limitations of methods for comparing design problems.....	138
7.6. Limitations of meta-regression results	139
7.6.1. Publication bias.....	139
7.6.2. False positive conclusions.....	141
7.6.3. Generalizability of meta-regression results.....	142
 CHAPTER EIGHT: FUTURE WORK.....	 143
8.1. Assessing similarity of design problems based on their representation	143
8.2. Understanding the role of design problems in ideation experiments	145
8.3. Understanding the impact of other sources of difference between studies	146
8.4. Opportunities in design tool validation	147
 APPENDICES	 154
Appendix A: List of design problems and authors	155
Appendix B: Design problem network analysis and representation.....	160
Appendix C: Design problem similarity assessment.....	165
Appendix D: Data for meta-regression analysis.....	185
 REFERENCES.....	 195

LIST OF TABLES

	Page
Table 1-1: Examples of design problem used in experimental studies in design creativity.....	4
Table 1-2: Overview of objectives and research tasks accomplished in thesis	8
Table 2-1: Comparison between wicked problems and conceptual problems	12
Table 2-2: List of characteristics identified by different researchers	20
Table 2-3: Summary of literature review on design problems and open research opportunities.....	25
Table 2-4: Example of filtering process chosen for study collection	32
Table 2-5: List of studies and elements of experiments used in them.....	33
Table 3-1: Overview of research tasks and corresponding research questions answered.....	40
Table 4-1: Sample coding procedure for researcher - design problem graph.....	46
Table 5-1: Example of scoring system and problem comparison based on approach 1	67
Table 5-2: Summary of alpha values for five characteristics identified by four evaluators	69
Table 5-3: A snippet of similarity scores. Top row and first column represent examples of problem statements used for analysis	78
Table 5-4: Structural and LSA similarity for alarm clock and corn shucking device	86
Table 5-5: Structural and LSA similarity for peanut shelling machine and blind cup device.....	88
Table 5-6: Structural and LSA similarity for automatic ironing device and recycling device.....	89
Table 6-1: Studies included for meta-regression and creativity metrics reported by them.....	106
Table 6-2: Design problems used in seventeen studies used for analysis and problem size	111
Table 6-3: Coefficients of regression model for metric quantity	113
Table 6-4: Summary of heterogeneity metrics for SMD distribution without moderator.....	115
Table 6-5: Summary of heterogeneity metrics for SMD distribution with moderator.....	115

List of tables (continued)

	Page
Table 6-6: Coefficients of regression model for metric quality	117
Table 6-7: Summary of heterogeneity metrics for SMD distribution without moderator.....	118
Table 6-8: Summary of heterogeneity metrics for SMD distribution with moderator.....	118
Table 6-9: Coefficients of regression model for metric novelty.....	119
Table 6-10: Summary of heterogeneity metrics for SMD distribution without moderator.....	120
Table 6-11: Summary of heterogeneity metrics for SMD distribution with moderator.....	120
Table 6-12: Summary of results for regression for metric quantity	124
Table 6-13: Summary of results for regression for metric quality	126
Table 6-14: Summary of regression results for metric novelty	127
Table A-1: List of design problems and their statements collected from 34 studies	155
Table B-1: List of authors and design problems used by them for 34 studies collected.....	160
Table B-2: List of node numbers and type of nodes for researcher-problem bipartite graph.....	163
Table C-1: Protocol used for inter-rater reliability test	165
Table C-2: Scores for element 'No. of goal' for all problems	178
Table C-3: Scores for element 'No. of functional requirements' for all problems	178
Table C-4: Scores for element 'No. of non - functional requirements' for all problems	178
Table C-5: Scores for element 'Information about end user' for all problems.....	178
Table C-6: Scores for element 'Reference to existing product' for all problems	179
Table C-7: LSA scores for design problem statements. Top row and first column represent the different problem statements.....	180
Table D-1: Standardized mean difference and within - study errors for metric 'quantity'	185
Table D-2: Standardized mean difference and within - study errors for metric 'quality'	186

List of tables (continued)

	Page
Table D-3: Standardized mean difference and within - study errors for metric ‘novelty’	187
Table D-4: Problem size as used in meta-regression analysis. NFR represents a non – functional requirement and FR represents a functional requirement.....	188
Table D-5: Results of asymmetry test of funnel plots	194

LIST OF FIGURES

	Page
Figure 1-1: Overall objective and supporting objectives for research.....	7
Figure 2-1: Ullman's taxonomy for design problems and research opportunity arising from it.	15
Figure 2-2: Summary of Summers and Shah's model and research opportunity arising from it	22
Figure 2-3: Summary of Thoe's model for exam problem evaluation and research opportunities arising from it.	24
Figure 2-4: Summary of Durand's model and research opportunities arising from it.	25
Figure 2-5: Elements of an experiment for creativity studies in engineering design	28
Figure 4-1: Bipartite graph showing design problems used by different researchers. Square vertices represent different researchers while circles represent design problems used by them.....	48
Figure 4-2: Node degree for different design problems.	49
Figure 4-3: One - mode graph projection showing connection between different design problems.	50
Figure 4-4: Bar graph of node degrees for all vertices of one-mode design problem network	52
Figure 4-5: Communities identified in design problem network using edge betweenness algorithm.....	53
Figure 4-6: Bipartite network showing design problems used in different papers published. Square vertices represent different papers while circles represent design problems used by them.	55
Figure 4-7: Design problems and their frequency of use in user studies published	56
Figure 5-1: Structural elements of a design problem	61
Figure 5-2: Example 1 showing characteristic identification in a design problem	64
Figure 5-3: Example 2 showing characteristic identification in a design problem	65
Figure 5-4: Variation of network density with choice of similarity cut off score	81
Figure 5-5: Design problems connected at cut off score of 0.35.....	82
Figure 5-6: Design problems connected at cut off score of 0.40.....	82
Figure 5-7: Design problems connected at cut off score of 0.45.....	83
Figure 5-8: Design problems connected at cut off score of 0.50.....	83
Figure 5-9: Alarm clock and corn shucking device problems used by Durand and coauthors.....	86

List of figures (continued)

	Page
Figure 5-10: Peanut shelling device and blind cup problems used by Durand and coauthors	87
Figure 5-11: Automatic clothes ironing device and recycling device used by Patel and coauthors	89
Figure 5-12: Design problems connected by size and LSA scores (>0.35).....	91
Figure 6-1: Steps for meta-regression analysis.....	104
Figure 6-2: Standard mean difference for treatments versus design problem size for metric quantity. Plotted line shows the regression line while the size of each circle represents the weight assigned to effect size during regression analysis.....	114
Figure 6-3: Standard mean difference for treatments versus design problem size for metric quality	117
Figure 6-4: Standard mean difference for treatments versus design problem size for metric novelty	119
Figure 6-5: Standard mean difference for treatments versus number of FRs for metric quantity	123
Figure 6-6: Standard mean difference for treatments versus number of NFRs for metric quantity	124
Figure 6-7: Standard mean difference for treatments versus number of FRs for metric quality	125
Figure 6-8: Standard mean difference for treatments versus number of NFRs for metric quality	125
Figure 6-9: Standard mean difference for treatments versus number of FRs for metric novelty	127
Figure 6-10: Standard mean difference for treatments versus number of NFRs for metric novelty	127
Figure 8-1: Four level hierarchical approach proposed for problem similarity assessment	144
Figure 8-2: Proposed validation framework for design tools and methods.....	149
Figure C-1: Scores for evaluator 1 for problem 1	166
Figure C-2: Scores for evaluator 1 for problem 2	167
Figure C-3: Scores for evaluator 1 for problem 3	168
Figure C-4: Scores for evaluator 1 for problem 4	168
Figure C-5: Scores for evaluator 2 for problem 1	169

List of figures (continued)

	Page
Figure C-6: Scores for evaluator 2 for problem 2	170
Figure C-7: Scores for evaluator 2 for problem 3	171
Figure C-8: Scores for evaluator 2 for problem 4	171
Figure C-9: Scores for evaluator 3 for problem 1	172
Figure C-10: Scores for evaluator 3 for problem 2	173
Figure C-11: Scores for evaluator 3 for problem 3	174
Figure C-12: Scores for evaluator 3 for problem 4	174
Figure C-13: Scores for evaluator 4 for problem 1	175
Figure C-14: Scores for evaluator 4 for problem 2	176
Figure C-15: Scores for evaluator 4 for problem 3	177
Figure C-16: Scores for evaluator 4 for problem 4	177
Figure D-1: Standardized residual vs predicted value of effect size for metric 'quantity' for design problem regression model	191
Figure D-2: Standardized residual vs predicted value of effect size for metric 'quality' for design problem regression model	191
Figure D-3: Standardized residual vs predicted value of effect size for metric 'novelty' for design problem regression model	191
Figure D-4: Normal Q-Q plot for metric 'quantity' for design problem regression model	192
Figure D-5: Normal Q-Q plot for metric 'quality' for design problem regression model	192
Figure D-6: Normal Q-Q plot for metric 'novelty' for design problem regression model	192
Figure D-7: Funnel plot for metric 'quantity'	193
Figure D-8: Funnel plot for metric 'quality'	193
Figure D-9: Funnel plot for metric 'novelty'	194

CHAPTER ONE

INTRODUCTION

1.1. Design problems in creativity research

A design problem (or design task) is a statement of requirement, need or function desired in the product [1]. The role of any engineering activity is to solve existing or anticipated problems of society, either through a new, innovative solution or improvements to the existing solutions. Identification of the problem is essential, in order to solve it. Often, design problems are presented to engineers and practitioners in the form of a requirement or need statement [2]. It is essential that such ‘statement of requirements’ are represented accurately to capture the essence of the original problem. Design problems are typically expressed using natural language, so that they are well understood by everyone [3][4]. Using natural language representation, however may lead to ambiguity in semantic interpretation by different designers (or problem solvers) [5,6]. At times, restating problem statements in a different way alters the requirements and objectives of the original problem as well [7].

An important characteristic of problems in engineering design, which differentiates it from problems encountered in other fields such as mathematics is the absence of unique solution state for a given design problem. Also, the steps and transformations needed to traverse from problem state to solution state is unknown, and is a function of the designer’s background, experience, creativity and other factors[8]. The progression from initial state to final is believed to occur in a sequence of events which involve co-evolution of the

design problem and solution space simultaneously [9–12]. Designers (or participants) formulate a partial representation of design problem as perceived by them, then work towards a solution to satisfy their problem representation. In the next progressive sequence, the problem is restructured based on initial solution generated, and the process is repeated until a final solution is arrived at, which satisfies the task requirements, as deciphered by the designer [13]. However, this process may also be solution driven, where the final solution state is chosen first, only to be re-modified and re-configured to suite the requirements mentioned in the problem task [14].

Research in engineering creativity often requires the use of predetermined design problems which is assigned to the study participants who generate solutions for the problem. These solutions are then compared across a variety of factors [15]. For example, a design problem used in an experiment conducted by Johnson and coauthors [16] was to *‘Design a litter control device for volunteers’*. However, design problems used in creativity research differ from problems encountered by designers in real life. Real world design problems are often presented as a new product proposal or a design change request arising from market needs, product ageing or a host of other factors [17]. These problems contain detailed specifications for functionalities and performance requirements of the product, with quantitative data being associated with most specifications. Deadlines and cost targets also form a part of such design problems, making it a comprehensive document [18]. A real life design problem contains all relevant information related to customer needs, functional requirements, spatial requirements, interface requirements and other details which enable the designer to create a realizable product [19]. Real problems are generally

large and may require long duration for developing a final solution. On the contrary, design problems developed for user studies in creativity (or conceptual problems) are generally designed to be appropriate for subjects of the experiment, are abstract in nature, solution-neutral and within the domain of knowledge of the researchers [11][20]. Conceptual problems are designed to encourage concept generation and creative thinking. Using simplified problems allows the researchers to study cognitive thought process and the impact of ‘treatment’ or design method being tested in a controlled space. This is also essential due to the limited experimentation time available for user studies which are generally conducted in academic settings with student participants [21]. It also reduces the need for the participants to possess multifaceted domain knowledge and expertise while solution formation. The focus of conceptual problems is on creative ideas and concepts with less impetus on inclusion of engineering considerations like design for manufacturability, design for assembly and service. Contrary to real life design problems, design problems used for user studies may or may not be associated with a real life customer or market need. For instance, Hernandez and Shah [22] used the ‘*Ping-pong ball transporter*’ problem in their experiment which does not have reference to an immediate customer who needs it.

“Design a device to ‘transport’ a ping-pong ball the farthest distance powered only by a standard issue compression spring. The device is to be constructed with a limited set of given materials (e.g. balsa wood, wire and Styrofoam).”

Table 1-1 shows examples of two conceptual design problems which have been used in the past. Problem 1 requires the design of solar heating and cooking device while problem 2 seeks the design of a new table for offices. In both examples, generic requirements from the product are specified without delving into details about form or design specifications. This is done to facilitate idea generation amongst participants. Both problems require basic understanding of engineering principles which was expected to be possessed by participants in the respective experiments (undergraduate mechanical engineering students in both case).

Table 1-1: Examples of design problem used in experimental studies in design creativity

Design Problem 1
<i>Develop products that utilize sunlight for heating and cooking food. The products should be portable and made of inexpensive materials. It should be able to be used by individual families, and should be practical for adults to set up in a sunny spot. Note: Specific materials for a targeted temperature can be postponed to a later stage [23].</i>
Design Problem 2
<i>It is asked to design a new table for offices that allows alternate sitting and stand up work. There are a lot of people who must work on sitting position the full day. The possibility to alternate positions during working time could drive to an improvement in health and productivity. The current tables that allow combining positions in work have limited surface, not enough for design, architecture and engineering needs [24].</i>

1.2. Research motivation

Researchers in creativity in engineering design have made significant contributions in furthering the understanding of human creativity and cognition. The engineering design research community is engaged in exploring diverse areas of human creativity. However, in the absence of a common research method, differences in research approach have also

surfaced. The design problem used in creativity experiments is one such source of difference between researchers. This difference can be attributed to the experiment's requirement based on which researchers design their problems. For instance, for studying the effectiveness of TRIZ intervention, Hernandez and coauthors [25] used a '*LED traffic light design*' as their design problem since domain familiarity and technical complexity were important requirements for the participant sample being used. In another creativity study conducted by Hernandez and Shah [22] to understand the key components of ideation method using provocative stimuli, a '*ping pong ball transportation device*' was used as a design problem since this was found to be suitable.

Design problems have been shown to influence the outcome of creativity tasks [15][26]. Their formulation and selection is therefore an important part of experiment design. Although researchers spend considerable time and effort in problem formulation, a research gap exists as far as design problem sharing between different experiments and their reuse in experiments is concerned. Since using the same problem is not always possible due to experiment's requirements, an alternative approach could be to use problems which are similar to each with respect to their contextual meaning, structure, effort needed for solving and motivation imparted to participants. Researchers have used different ways for comparing problems such as problem structure, requirements contained, word counts and results from past experiments [15,27,28]. It is felt that comparing problems based on their contextual meaning and structural elements can further help researchers in problem comparison and formulation. In the long run, these methods can also be used as a means for establishing and comparing benchmarked design problems for

use in experiments in engineering design creativity. Also, similar problems can be used for testing the robustness of design methods and tools being developed. Addressing this research opportunity can therefore enable researchers to compare and use conceptual design problems from the existing pool of problems which could ultimately lead to higher problem reuse in experiments.

1.3. Research objectives

The overall objective of this research is to address the need for using similar design problems in creativity research. The focus of this research is to study design problems and identify the need and ways to reduce variability in design problem usage across experiments.

The main objectives of this research are to:

- Understand and identify the pattern of design problem usage to show how design problems have been used in creativity studies by researchers.
- Enable comparison between conceptual design problems based on problem statement.
- Analyze the influence of design problems on the effectiveness of intervention or method tested in user studies in design creativity using meta-regression. Specifically, the effect of design problems on effectiveness of examples used during experiments is studied.

Figure 1-1 shows the overall objective and supporting objectives for this research.

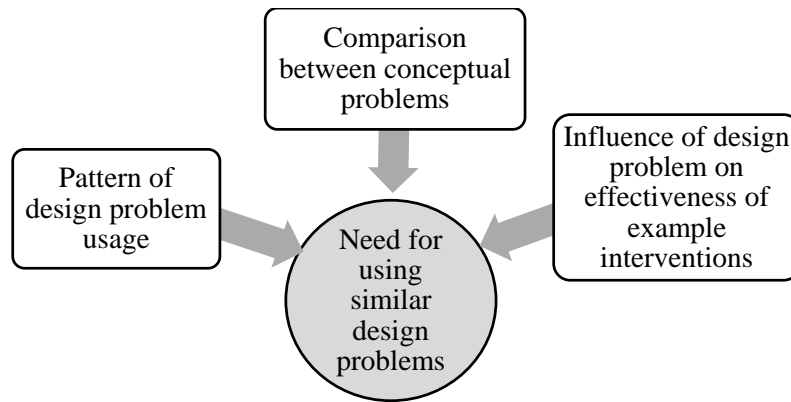


Figure 1-1: Overall objective and supporting objectives for research

1.4. Key research tasks

The approach chosen for addressing the research objectives is discussed in Chapter three. The following research tasks are undertaken for specifically answering these research objectives:

- a. Generation of graph to identify how design problems have been used in the past 15 years by researchers. (Chapter 4)
- b. Postulation and evaluation of methods to compare conceptual design problems. (Chapter 5)
- c. Meta-regression analysis to understand the impact of design problem on effectiveness of intervention (presenting examples during ideation) in user studies in engineering design creativity. (Chapter 6)

1.5. Overview of thesis

Table 1-2 provides an overview of the research objectives and tasks accomplished in this research.

Table 1-2: Overview of objectives and research tasks accomplished in thesis

Overall research objective	To address the need for using similar conceptual design problems for experiments in engineering design creativity.
Sub-objectives	<ul style="list-style-type: none"> • To understand and identify the pattern of design problem usage. • To enable comparison between conceptual design problems based on problem statement. • To analyze the influence of design problems on effectiveness of example intervention in user studies in design creativity.
Research question 1	How can the pattern of design problem usage in creativity research be identified?
Background	Different experiments/user studies use different design problems based on their requirements resulting in low sharing and reuse of design problems.
Research task 1	Graph based analysis of design problem usage
Research question 2	How can conceptual design problems be compared using their problem statements?
Research question 3	How can natural language representation be used to compare conceptual problems?
Background	Comparing conceptual design problems is essential since using the same design problem may not be always possible. Similar problems can also be used for validation and verification of design methods.
Research task 2	Comparison based on structural elements Similarity assessment based on Latent Semantic Analysis
Research question 4	Does the choice of design problem influence the effectiveness of example interventions?
Background	Design problems are important covariates in user studies in design creativity and have been shown to influence the results of creativity experiments.
Research task 3	Meta-regression analysis of user studies in design creativity.

CHAPTER TWO

LITERATURE REVIEW

2.1. Characterizing design problems: a historic overview

There have been several researchers across different domains who have pursued the idea of defining a structure or framework for representing design problems in their respective avenues. Eastman [29] first identified two primary features that distinguished ill-defined problems in architecture from well-defined ones: lack of formal representation language and well-defined specifications for final goal. Eastman used an example of space planning problem used in architectural design to analyze the design process used for solution generation and concluded that search and specification processes together can completely depict a large number of ill-defined problems.

Rittel [30] classified design problems as wicked problems, due to their ill-formulation, conflicting customer requirements and speculative problem solving approach. Rittel used problems in general planning to elucidate the nature of design problems, which he referred to as ‘wicked problems’. According to Rittel, wicked problems are associated with the following characteristics:

- There is no definitive formulation of a wicked problem, which means that an exhaustive formulation containing all information required for understanding and solving the problem can never be built into the problem statement.
- There are no stopping rules for wicked problems, since the problem solver cannot ascertain when the job is complete.

- There are no true or false solutions, but only good or bad solutions for wicked problems. Solutions which are formulated for such ill-structured or wicked problems can only be classified as either good or bad, since the correct solution state is unknown. Solution aptness is subjective and hence open to contention.
- There are no immediate or ultimate test for solutions of wicked problems and the repercussions from solution implementation cannot be determined until all consequences resulting from solution implementation have been studied.
- There is no opportunity to learn by trial – and – error and every implemented solution has irreversible consequences.
- There are no criteria to prove that all solutions to a problem have been identified and considered. Such problems (including design problems) have open ended solutions and do not contain an exhaustive list of solutions.
- All wicked problems are unique, in a sense that there might be additional distinguishing properties between two problems which override the similarities between the two problems.
- There is always a cause associated with a wicked problem. Problems originate when there is a discrepancy between the state of affairs as it is and the ideal state of affairs as it should be.
- Explanation for the existent discrepancy for a wicked problem determines the nature of problem's resolution. For instance, several alternative reasons may exist for the need for a new hair dryer, all of which may be right. Each of these alternative explanations

might be correct, as there are no rules or procedures to determine the correct explanation.

- The problem solver has no right to be wrong when solving a wicked problem, since the aim is not to discover the truth but to solve existing problems of community. A ‘wrong’ solution can have serious consequences.

Simon [8] emphasized on lack of well-defined solution and transformation states as a characteristic of design problems which he also reckoned as ill-structured problems. Simon highlighted the following characteristics of well-defined problems:

- Presence of a description of solution state or a test to determine whether the solution state has been reached or not.
- Presence of a set of terms describing and characterizing the initial state, goal state and intermediate states.
- Presence of a set of operations which define how to transform from one state to another.

Simon then used the problem of ‘designing a house’ which he classified as ill-structured since it lacked the characteristics of a well-structured problem. Simon also elucidated how designers constantly modify and re-structure the problem statement as solutions are derived, leading to co-evolving problem definition.

Most of these characteristics for a wicked problem can be seen in conceptual design problems, such as:

- Conceptual problems do not always contain all required information needed for solving them. Problems are open ended and do not contain specific information related to form,

size, features and interface. This is purposely done by researchers to keep the focus of participants on ideation.

- Like wicked problems, conceptual design problems can have open ended solutions. For example, a problem statement '*design a device which can compost vegetables*' [31] is open ended and does not insinuate towards a particular design solution. A problem statement '*design a hair dryer*' hints towards a known solution but still is open ended as far as the shape, form and interface of the solutions generated is concerned.
- The solution state for design problems used in experimental studies may not be exhaustive. Participants may not be able to determine whether they have identified all solutions to the given problem or not. Wicked problems have similar characteristics.
- The steps needed to transform from problem to solution is generally unknown for participants in a user study which uses a conceptual problem. This is why solutions generated vary from person to person since there are no definitive rules for transformation like wicked problems.

Table 2-1 shows a comparison between wicked problems and conceptual problems based on presence or absence of characteristics identified by Rittel and Simon.

Table 2-1: Comparison between wicked problems and conceptual problems

Characteristic	Wicked problem	Conceptual problem
No definitive formulation	✓	✓
No stopping rules	✓	✓
No true or false solutions	✓	✓
No immediate tests for solutions	✓	✓
No trial and error opportunity	✓	✓
Unknown solution space	✓	✓

Characteristic	Wicked problem	Conceptual problem
All problems are unique	✓	✗
Cause for problem's origin is known	✓	✗
Solver can't generate a wrong solution	✓	✗
Steps to transform to solution unknown	✓	✓

Dixon and coauthors [32] proposed a taxonomical characterization of design problems. Mechanical design problems are classified into five major categories based on the nature of initial and desirable final knowledge states. The basic problem types include:

- Functional problems
- Phenomenological problems
- Embodiment problems
- Attribute problems
- Parametric problems

Dixon and coauthors also sub classified the initial state of knowledge of artifact type by specifying a physical type, an assessment type and coupling. They also defined the size of a problem statement as the number of design variables needed to describe the component which is to be designed. According to Dixon and co-authors, coupling is a measure of interdependence among the performance parameters and the design variables which specify the design itself. Designs are completely uncoupled when each performance parameter are determined by a single design variable alone. The proposed taxonomy by Dixon and coauthors is suited to existent products and problems which have a detailed description associated with them. For a problem to be included in one of the five categories,

a description of initial and final states of knowledge associated with the problem is needed. For instance, initial state of knowledge may be in the form of perceived needs, function, physical phenomenon, embodiment, artifact type or artifact instance. Thus, the research opportunity emerging from Dixon's work is:

Can the size of design problem be evaluated based on problem description?

Ullman [33] extended the taxonomy by including the design environment and the design process in the taxonomy. Ullman characterized design problem based on design's initial and final states, its refinement level and representation language. The modified refinement levels for design include perceived needs, design specifications, functions, physical phenomenon, artifact type and artifact instance. According to Ullman, another aspect of initial and final design states can be associated with the representation language used for problem definition. However, he admits that in order to fully define the design problem, satisfaction criteria and knowledge of attainment of final solution state are needed. Further, Ullman also illustrated the application of his taxonomy for different design stages. In particular, he concurred to the fact that conceptual design could be described by design problems alone without any reference to environment or design process. The initial and final states of a design problem were defined by two characteristics: refinement level and representation. The taxonomy provided by Ullman enables one to add structure to problem statement and provides the research opportunity of using information contained in problem representation as distinguishing factor between two problems. Figure

2-1 shows an overview of Ullman's taxonomy as applied for conceptual design problem classification and the research opportunities arising from his proposed methodology.

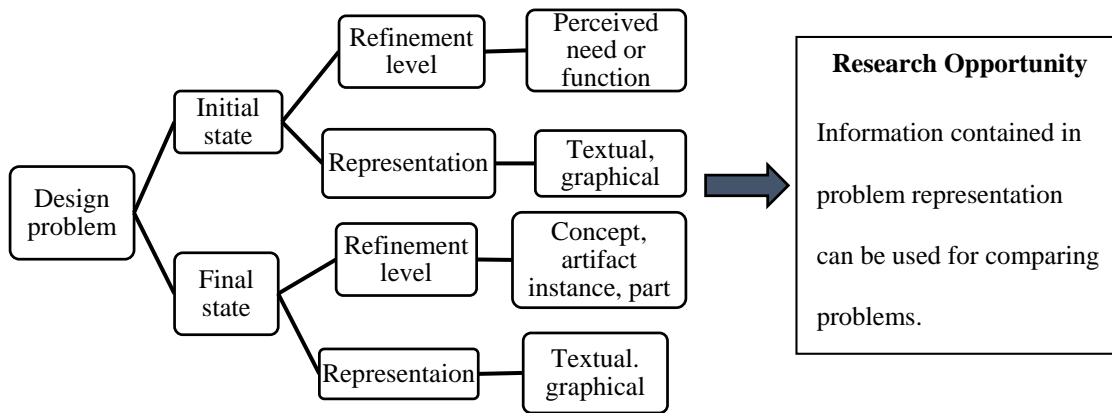


Figure 2-1: Ullman's taxonomy for design problems and research opportunity arising from it [33].

Goel [34,35] proposed twelve common features of design task environments, and suggested that some of these salient characteristics can constitute good examples of design activity. According to them, design tasks comprise of three elements: i) a goal, ii) a problem and iii) other external factors. They identified 12 discernible features of design task environments:

- There is lack of information in all three components of a design problem.
- Two types of constraints are generally present in design task environments, namely nomological and social. While social constraints are negotiable, nomological constraints comprise of natural laws which are non – negotiable.
- Design problems are generally complex and require long time scales for arriving at a solution.
- Design problems are comprised of multiple parts, decomposition of which is determined by the practice and experience of the designer.

- The components of a design problem are not logically interconnected. However, contingent connections and interconnections between the component does exist.
- Inputs to design problems consist of information about end user, the goals and the functions that need to be performed to meet the goals. Outputs from a design problem consist of an artifact specification.
- A feedback loop from the real world during problem solving does not exist. The feedback from end users comes only after the product has been designed and constructed.
- A difference exists between the specification of the artifact and product delivery.
- Product specification always precedes product delivery.

Goel and coauthors later justified the presence of these characteristics in design problem spaces by illustrating their presence in two well – structured problems: crypt – arithmetic and Moore – Anderson logic task. Design problem spaces were generalized as co – evolving spaces, where continuous restructuring of design problem with solution evolution occurs. This work by Goel and coauthors describes the general characteristics of design problems and presents the following research opportunity:

How to quantitatively compare design problem representations based on certain characteristics or features contained in them?

Frost [36] suggested an eight factor non-orthogonal taxonomy for categorization of engineering design problems in order to allow systematic correlations to be made between

different problem classes and appropriate methods for solving them. The eight factors according to Frost's taxonomy include:

- Type of entity being designed: This factor consists of a list of discipline areas which can help in categorization of problems. The type of entity considered could be very general by discipline only (viz. aeronautical, biomedical and others), more specific relating to devices within mechanical discipline (viz. compressors, conveyors, cranes and others) and very specific, relating to components within a device (viz. balance shaft, bearings, bolts and others).
- Degree of innovation involved: According to Frost, the degree of innovation required for solving a problem statement can be gauged from the stereotype of product which is needed. A scale relating to this factor can be regarded as an innovative index can be used for 'quantifying' the innovation level required for solving a problem. For instance, if the problems require designing a revolutionary stereotype with no pre-existing solutions, a high degree of innovation is needed for solving such problems.
- Extent to which designed entity can be decomposed into sub – systems: The number of sub – divisions which a product can be conceptually decomposed into determines the ease with which problem can be solved. Products which can be decomposed into more sub – systems enable more people to work effectively and simultaneously.
- Availability of adaptable solution concepts: This factor is associated with the availability of potentially analogous concepts which can be adapted for use in subject situation. The availability of adaptable solutions makes conceptual design task easier.

This factor can be quantified by the fraction of sub – systems for which analogous solutions can be found.

- Complexity of designed entity: Complexity of the designed entity can be adjudged by the number of sub – systems which the designed entity will contain and the complexity of configuration related to each. Thus, a complete large ship is a complex entity while a small hose fitting can be entitled as a simple entity. The entities are labelled on a scale of 1 – 10 from simple to complex.
- Degree of interaction within solution: This is determined by the extent of connection between the features and parameter values. According to Frost, it is difficult to ascertain an objective measure of this parameter.
- Looseness or tightness of constraint that the design must satisfy: This factor is a measure of how exacting the constraints and requirements are in the design problem. An objective measure of this factor is difficult. However, a problem can be completely constrained or unconstrained.
- Number of artifacts to be built: Since product cost is an indispensable part of design process, the number of parts which need to be manufactured has an impact on decisions made regarding whether a new part should be designed or an existing, off – the – shelf part should be used. The anticipated production volume may be determined from the problem statement based on the type of product being designed. For example, the production volume for a problem seeking design of a power house would be low as compared to a problem which requests for the design of a household refrigerator, the production volume for which can be gauged to be high.

Frost's taxonomy is apt for problems which have existing products and forms associated with them. Assessment of factors like number of artifacts to be built and complexity of designed entity would be challenging in conceptual problems, since such problems are more abstract in nature and are ill defined. However, this work by Frost also provides the opportunity of using some of the elements from his taxonomy for drawing similarity comparison between conceptual design problems.

Durand and colleagues [15] hypothesized nine primary characteristics that make up a design problem based on literature review and problems used by them in the past. According to Durand and coauthors, size for a problem can be estimated based on the number of functional requirements contained in the problem. Likewise, coupling between requirements and constraints can be used to gauge connectedness. However, some of their characteristics including participant's familiarity with design problem and solutions correspond to inherent traits of participants, rather than design problem itself. Other characteristics like effort required in solving and degree of fixation induced are not entirely properties of design problem, but also depends on the nature of participants, their experience and skill sets. Durand and colleagues did not illustrate the presence or absence of these characteristics in the problems chosen by them for their experiment and thus present the following research opportunity:

How can design problems be compared to each other based on the characteristics contained in them?

Table 2-2 shows the characteristics identified by different authors over the years.

Table 2-2: List of characteristics identified by different researchers

Characteristic	Author					
	Rittel	Simon	Dixon	Goel	Frost	Durand
Definitive formulation						
Stopping rules						
True or false solutions						
Immediate or ultimate test of solution						
Trial and error opportunity						
Exhaustive solution space identification						
Problem uniqueness						
Problem cause						
Defined initial state						
Defined goal state						
Defined intermediate state						
Transformation operations						
Functional requirements						
Phenomenological problems						
Embodiment problems						
Attribute problems						
Parametric problems						
Presence of constraints						
Problem complexity						
Interconnected parts						
End user information						
Type of entity						
Degree of innovation						
Decomposability						
Analogous solutions						
Complexity of product						
Number of artefacts						
Coupling between requirements						
Participant's familiarity						
Solvability						

2.2. Methods for design problem evaluation

Summers and Shah [37] evaluated the size, complexity and solvability of design problems and design processes based on six vocabulary elements of a design problem. These six vocabulary elements include design goals, independent design variables, dependent design variables, measures of goodness, design relationships and design constants. Complexity as size can be estimated by the size of language and count of language instances. By counting the number of design variables, functional requirements, non – functional requirements or subassemblies, the size of problem model can be gauged. In order to evaluate design problem complexity as coupling, connection between variables at multiple levels can be used. Problems modeled in graph based representations can be used for measuring coupling, based on the decomposability of the graph. A graph of design problem which is easily separable into distinct sub graphs is not highly connected, hence less coupled. For measuring design problem complexity as solvability, Summers and Shah suggest to measure the degree of freedom that is permitted in a constrained problem. They concluded that complexity as solvability increases with reduction in constraints imposed on design variables. This work is seminal since it enables calculation of problem solvability and coupling which is important while choosing a design problem for an experiment. Researchers generally choose problems which are within the domain of knowledge of the participants and as such, evaluation of problem solvability is essential in selecting an appropriate problem. The methods for problem evaluation discussed by Summers and Shah are more appropriate for parametric and embodiment design problems since such problems contain well defined design variables and design relations needed for complexity

assessment. Therefore, the following research opportunities are realized from this work by Summers and Shah:

Can elements in conceptual problem statements such as functional and non-functional requirement be used for assessing problem size?

Figure 2-2 shows the summary of complexity assessment model proposed by Summers and Shah and the research opportunities arising from it.

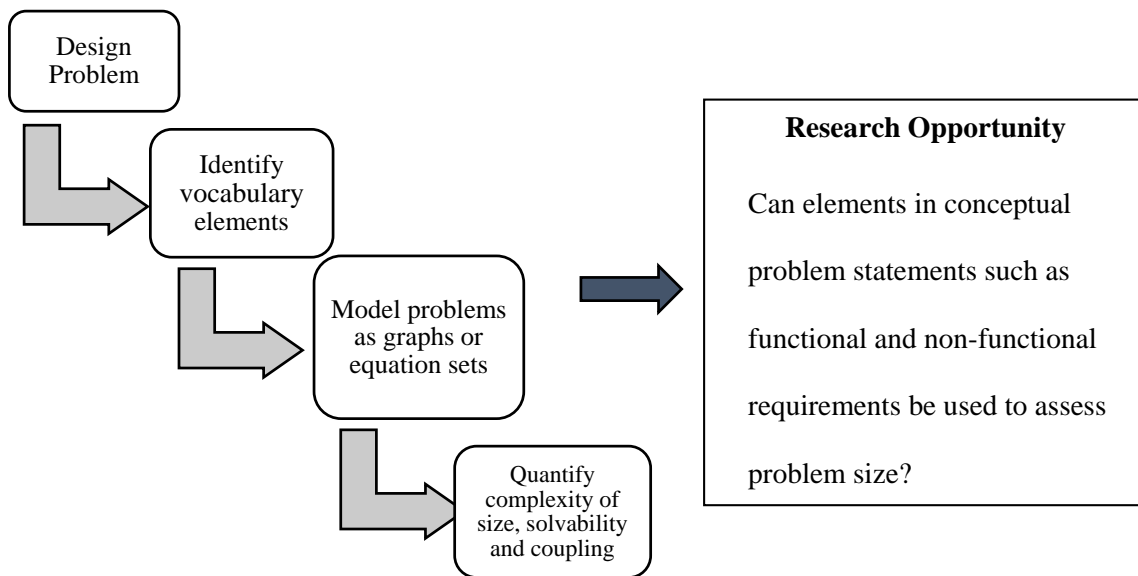


Figure 2-2: Summary of Summers and Shah's model and research opportunity arising from it

Thoe and Summers [38] used a graph based approach to assess the complexity of exam problems. Variables or unknowns in a given problem could be linked to equations or known values used for solving the problem. As a result, a bipartite graph is generated with the variables being connected to the equations in which they are used or derived from. This graph is then analyzed using a variety of graph metric (viz. size, interconnection, centrality and decomposition). The values of metrics obtained from network analysis along with the

assessed effort or point value are then used as inputs and targets for training the neural nets respectively. Thus, the score prediction can be obtained for a given exam problem once the neural network has been adequately trained. The problems used for testing this method were from three mainstream courses in mechanical engineering (viz. heat transfer, mechanisms and mechanics of materials) where correct solution states exist for such problems. On the contrary, design problems (especially the ones encountered in conceptual studies) are open ended with no known solution. Also, there are no defined transformation states associated with design process which elucidate the path that one needs to traverse to arrive at a solution. In such a case, the procedure for connecting design variables given in a problem statement to ‘equations’ used during solution process becomes difficult, if not impossible. The research opportunity arising from this work by Thoe and Summers is whether problem solvability for conceptual problems be assessed by relating certain variables in problem statement with a chosen solution state. Figure 2-3 is a summary of model proposed by Thoe and Summers for evaluation of exam problems and the research opportunity arising from their model.

Durand and colleagues [15] hypothesized nine primary characteristics that make up a design problem. They also proposed a method to compare design problem similarity based on the measure of creativity metrics obtained from experiments using design problems. It was concluded that design problems which lead to comparable creativity metric scores are similar to each other. They compared two problems across a set of four metrics and found each problem generates a different ‘fingerprint’ on each metric. The two problems used by them for this evaluation resulted in similar scores for quantity of ideas

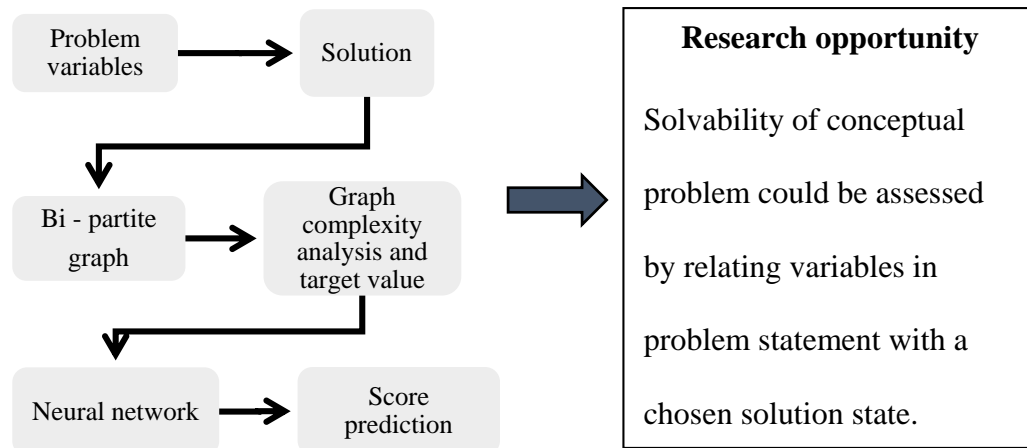


Figure 2-3: Summary of Thoe's model for exam problem evaluation and research opportunities arising from it.

generated but different quality and variety scores. However, this procedure makes it difficult to ascertain which problems are similar since they produce varying results across the four metric. This work by Durand and others also provides a research opportunity as to how design problem can be compared to each other based on their statements by eliminating the need for using the problems in user studies before comparison is made possible. Figure 2-4 shows a summary of model proposed by Durand and coauthors and the research opportunity arising from it.

Can problem similarity be assessed based on its representation alone without using the results of experimental study?

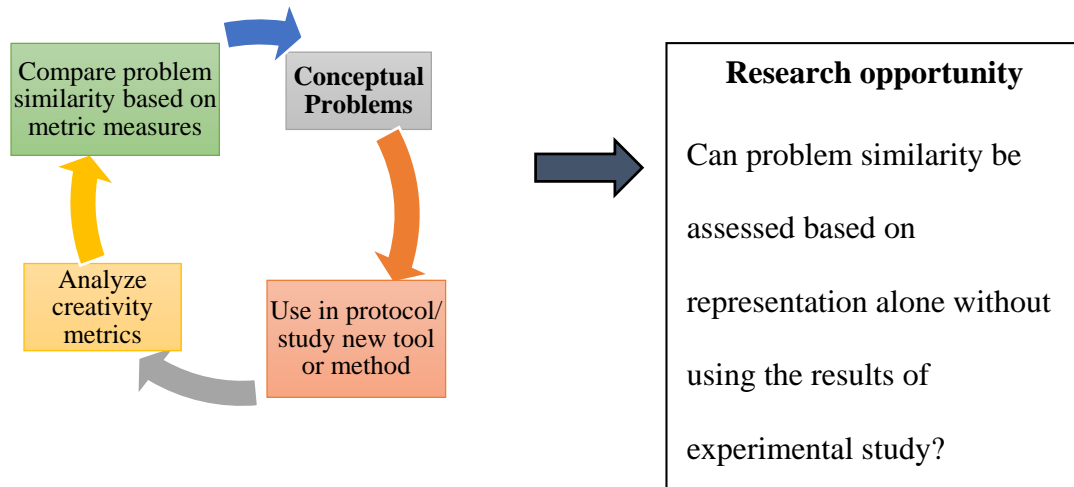


Figure 2-4: Summary of Durand's model and research opportunities arising from it.

Table 2-3 summarizes the contribution from various authors over the years towards characterization of design problems and the potential research opportunities present in their works.

Table 2-3: Summary of literature review on design problems and open research opportunities.

Author/Year	Contribution	Research opportunity
Eastman, 1969	<ul style="list-style-type: none"> Identified two features which distinguish ill –defined problems from well – defined ones. Problem solution and specification processes can depict ill-defined problems. 	Some characteristics of ill – defined problems may be used for characterizing conceptual problems.
Rittel, 1973	<ul style="list-style-type: none"> Classified planning problems as wicked problems. Identified 10 characteristics of wicked problems. 	Some characteristics of wicked problems may be associated with conceptual design problems.
Simon, 1977	<ul style="list-style-type: none"> Identified the characteristics of well – defined problems. Emphasized the idea that design problems co-evolve during design process 	Conceptual problems can be differentiated from well – defined problems based on characteristics.

Author/Year	Contribution	Research opportunity
Dixon, 1988	<ul style="list-style-type: none"> Proposed a taxonomy for design problems based on initial and final states of knowledge 	Size of design problem be evaluated based on the number of design variables.
Ullman, 1992	<ul style="list-style-type: none"> Proposed taxonomy for problem characterization based on initial and final states of design, its refinement level and representation language. 	Information contained in problem representation can be used for comparing design problems.
Goel, 1992	<ul style="list-style-type: none"> Suggested twelve features of design task environments Emphasized on co-evolution of design problems during design process 	Is it possible to quantitatively compare design problems?
Frost, 1994	<ul style="list-style-type: none"> Proposed an eight factor taxonomy for design problem 	Some of Frost's factors can be used for characterizing conceptual design problems
Summers and Shah, 2010	<ul style="list-style-type: none"> Provided a way to determine complexity of design problem in terms of size, coupling and solvability 	Can language representation be used to compare problems for similarity?
Thoe and Summers, 2013	<ul style="list-style-type: none"> Proposed a graph based approach for complexity measurement of exam problems 	Can conceptual problems be evaluated based on elements present in problem statement when solution state is unknown?
Durand and Linsey, 2015	<ul style="list-style-type: none"> Hypothesized nine characteristics of design problems. Elucidated the impact of problem type on results of creativity studies. Use creativity scores to compare design problems for similarity 	How can problem similarity be assessed based on representation alone without using results from experimental study?

2.3. Differences between user studies in design creativity

Study of creativity in engineering design can be challenging since a creative ‘event’ may or may not occur during the design process [11]. This makes it difficult to study the phenomenon in its natural setting (in a design office of an organization, for instance). Design researchers often rely on user studies with human participants to understand various aspects of design creativity [16,22,23,39–52]. User studies help in reducing the associated complexity of a design process by manipulating one or more independent factors, thereby making it possible to measure the required dependent variable and controlling other variables [53]. Frey [54] advocated an analogous comparison between clinical trials and tests for design methods. According to Frey, design methods can be developed and validated using five approaches which include:

- Controlled field evaluation of design methods
- Case studies of industrial practice
- Experiments with human subjects in laboratory setting
- Detailed simulation of design methods
- Theoretical understanding of cognitive sciences and organizational behavior

Experimental studies provide a controlled environment to analyze and comprehend different facets of design creativity. Eight elements make up these experiments and are shown in Figure 2-5.

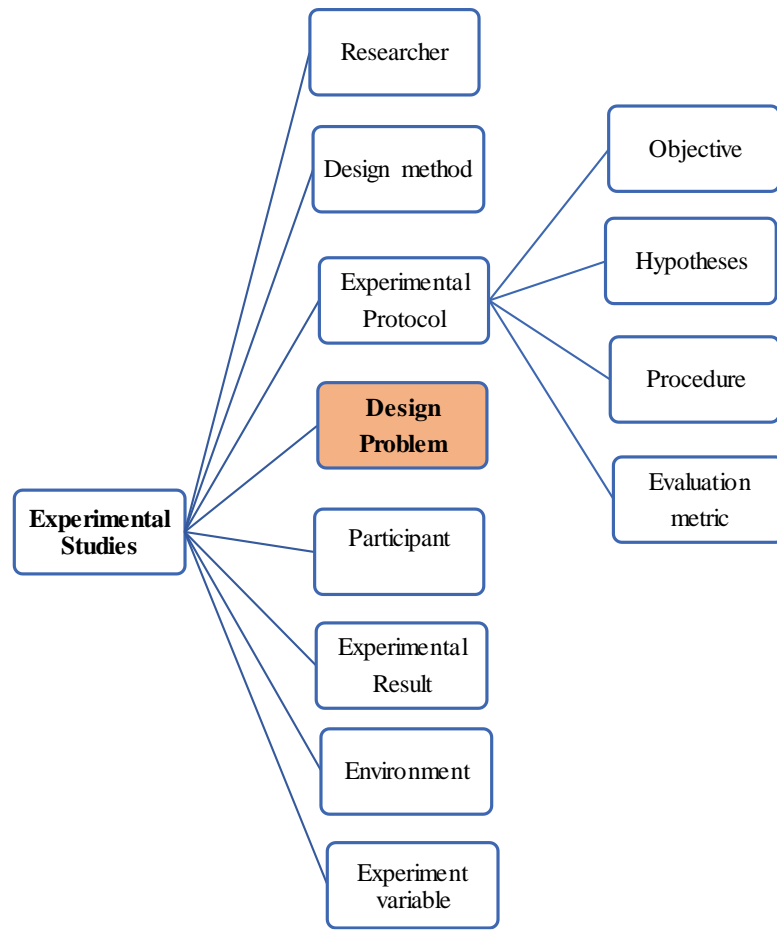


Figure 2-5: Elements of an experiment for creativity studies in engineering design

- **Researcher:** Someone who conducts an organized and systematic investigation into an area of interest. In conceptual design research, people with academic affiliations are generally involved in systematic studies on design creativity. Major tasks of a design researcher include hypothesis formulation, selection of participants, study design, choice of design method or tool to be tested, problem selection, implementation of study protocol and result analysis.

- Design methods: Sequence of activities to be followed in order to improve particular stages of design process [55]. For instance, brainstorming is a method used during conceptual design stage which has been shown to improve solution creativity [56,57].
- Participants: The subjects of the user study who take part in the experiment. The influence of design method used in experiment is evaluated based on the results generated by participants in the process. Often, students in academic institutions are chosen as subjects since they are easy to access and inexpensive [58,59].
- Experiment protocol: A predefined, written list of step to be followed for implementing the experiment. Every experiment has a set of rules and a procedure designed in order to render it useful for the study. Protocols are the guidelines and procedures to be followed for conducting an experiment.
- Design problem: Problem statements provided to the designer containing the requirements, needs, functions, or objectives which the design needs to satisfy [1]. These design problems are sometimes called design briefs or tasks, and can vary widely in content and form. Problems used in creativity studies are generally ‘small’ problems aimed at instigating idea generation [11].
- Experiment results: Measurable outcomes from a study resulting from the operations performed in the experiment. Conclusions about the cause and effect relationship are drawn based on results obtained. User studies in design creativity usually generate result in one of the following forms: i) sketches of ideation process, ii) written notes, iii) audio or video recordings, iv) participant survey or v) reflection notes [60].

Experiment results, along with data collected from literature search should ideally contain all information required to answer the research questions and hypotheses.

- Environment: Surrounding in which the experiment is executed. This may be a classroom or a shop floor where the study is being conducted.
- Experiment variable: Any factor, trait or conditions that exists in an experiment. They may be of three types: independent, dependent and controlled. Researchers manipulate independent variables to observe the effect on dependent variables while trying to keep the controlled variables as constant.

Researchers determine the elements to be used based on experimental requirements. Some of these differences arise because of constraints associated with user studies in creativity. For instance, since most studies in design creativity are conducted in an academic setting, it is common for researchers to use students as test subjects. Researchers choose participants based on their availability and accessibility, hence the choice of participants is likely to be different between studies. Likewise, different design methods and experimental protocols are used since the objectives of these studies are different. In order to highlight the variation in choice of elements, an analysis of user studies published in the past is presented in following section.

Analysis of published user studies

For illustrating between-study differences in engineering design creativity, user studies published between 2000 and 2014 are collected and analyzed to identify different design problems, participants, metrics for assessment, method and experiment variables that have been used in them.

Collecting studies

The following steps are used for study collection:

1. Online search engines Web of Science, Google Scholar and EBSCOhost are searched using the following keywords: '*creativity + engineering design*', '*ideation + engineering design*' and '*experiment + creativity + engineering design*'. Using these keywords in succession brings down the original number of results from 644,000 to 11,400.
2. Filtering is done based on the year of publication. A range between 2000 and 2014 is chosen based on a prior search result since the year range between 1960 and 1999 produced results which were less focused on engineering design or user studies. This filter brings down the results to 8080 studies.
3. Further filtering is done based on the journals in which the studies are published. The search results are limited to Design society, Design studies, ASME journals and other journals relevant to engineering design. The total number of studies remaining after this step are 392.
4. The next level of filtering is done by analyzing the abstract of 392 studies resulting from step 3. The abstracts are reviewed for presence of an experiment in the paper. This further reduced the total number of studies from 392 to 129.
5. The next level of filtering is done by analyzing the 129 resulting studies in step 4. Here, studies which did not report or use a design problem are removed. The total number of studies remaining after this step is 93.

6. The next level of filtering chosen is whether the study contains a user study or not. Case studies and design projects are thus eliminated from the list. This brings down the list from 93 to 41.
7. The last filter level is to check for studies which have published duplicate experiments. This helps to bring down the number of studies from 41 to 34. Thus in total, 34 user studies in the area of engineering design creativity are collected through this online database search and filtering process. Table 2-4 shows a portion of filtering process which was used for collecting studies. Table 2-5 shows the list of studies and the different elements contained in them.

Table 2-4: Example of filtering process chosen for study collection

Reference no of article	Published between 2000-2015	Related to engineering design	Contains a design problem	User study	New experiment
[61]	✓	✓	✓	✗	✓
[44]	✓	✓	✓	✓	✓
[62]	✓	✓	✗	✗	✗
[24]	✓	✓	✓	✓	✓
[63]	✓	✓	✓	✗	✓
[45]	✓	✓	✓	✓	✓
[64]	✓	✓	✓	✓	✓
[65]	✓	✓	✗	✗	✗
[66]	✓	✓	✗	✗	✗

Table 2-5: List of studies and elements of experiments used in them

Reference no.	Author/s	Participant	Problem name	Result	Metric used
[44]	S. Kim, Y. Kim	Freshman engineering students	Subway improvement	Sketches	CPSS, IPA
[24]	Sonseca, Mulet, Chakrabarti	16 PhD candidates+2 Industrial Designers	1) New table, 2) Tubular map case 3) System for gathering wire, 4) Table for offices	Sketches	CPSS
[45]	S. Kim, H. Kim, Jin	Engineering undergraduate	Wearable binocular	Sketches	Gough Creativity Index
[64]	Rogers, Salustri	Graduate students, Industry professionals	Bi/tri cycle	Sketches, Report, identification matrix, system diagrams,	Novelty, usefulness, cohesion
[67]	Naim, Lewis S. Schmidt, Viswanathan, Linsey, McAdams, Campbell, Poppa, Robert	Undergraduate students in final year	1) Water lifting device 2) Peanut shelling machine	Sketches, Written scripts	Number of unique ideas
[25]	Hernandez, C. Schmidt, Okudan	Engineering students	Traffic light using LED	Sketches	Quantity, Novelty, Variety
[68]	Toh, Miller	Engineering students	Milk frothing device	Sketches, Interview questionnaire	Quantity, Quality, Novelty, Variety
[69]	Linsey, Markman, Wood	Senior undergraduate students	1) Peanut shelling machine 2) Commercial product	Morph Matrices, survey results, sketches	Number of analogies

Reference no.	Author/s	Participant	Problem name	Result	Metric used
[23]	Daly, Christian	Engineering students	Solar device	Sketches, Answers to questions	Diversity, Number of concepts
[50]	Glier, S. Schmidt, Linsey, Mcadams	1) Second semester students 2) Senior level undergrads	Peanut shelling machine	Sketches, Written notes	Quantity, Quality
[42]	Linsey, Clauss, Kurtoglu, Murphy, Wood, Markman	Students from senior mechanical design course	Peanut Shelling machine	Sketches	Quantity, Quality, Novelty, Variety
[70]	Acuna, Sosa	Industrial Design students	Counter top stand	Sketches, Models	Originality, Functionality
[22]	Hernandez, Shah, Smith	Mechanical Engineering undergraduate students	1) Ping pong ball transporter 2) Tool for alien species	Sketches	Quantity, Quality, Novelty, Variety
[71]	C. Schmidt, Hernandez, Kremer, Linsey	Undergraduate Engineering students	Biomass cooking device	Sketches, recorded data on smart pen	Novelty, Quantity, Variety
[72]	Weaver, Wang, Kuhr, Crawford	Undergraduate students	UAV	Sketches, internet search list, documents	Number of concepts, diversity, novelty
[73]	Doboli, Umbarkar	Senior undergraduate students	Innovative remote controller	Sketches, comment, writings	Quality, Novelty, variety
[74]	Kudrowitz, Te, Wallace	Industrial designer, Engineering student, middle school students	1. Doodle toaster 2. Coffee maker 3. Horizontal toaster 4. Crumb tray toaster	Sketches	Online survey
[75]	Lopez, Linsey, Smith	Undergraduate mechanical engineering student	Peanut Shelling machine	Sketches, survey answers	Total number of ideas

Reference no.	Author/s	Participant	Problem name	Result	Metric used
[76]	Okudan, Hernandez, Jablokow, C. Schmidt, Lin	Engineering students	Traffic light using LED	Sketches, Comments	Novelty, Variety
[77]	Jin, Benami	Mechanical engineering students	Oars	Design sketches, video records	Number of design concepts
[78]	Cheong, Chiu, Shu	Fourth-year engineering students	Waste sorting device	Causal relation template	Novelty, usefulness, cohesiveness
[79]	Kurtoglu, Campbell, Linsey	Graduate students	1) Bottle capping device 2) Soda maker	Sketches, Design notes, Survey	Quantity, Quality, Novelty, Variety
[80]	Tsenn, Atilola, McAdams, Linsey	Senior Mechanical Engineering undergraduates	Peanut Shelling machine	Sketches, Notes and comments	Quantity, Quality, Novelty, Variety
[81]	Toh, Miller, Okudan	Engineering design students.	Electric toothbrush	Sketches, notes	Quantity, quality, novelty, variety
[16]	Johnson, Genco, Paul Seepersad, Otto	Senior mechanical engineering students	1) Next Generation alarm clock 2) Litter collection device	Sketches, notes	Originality,
[82]	Shorachi, Goncalves	Industrial design students	Alternative alarm clock	Sketches, notes	Fluency, Rarity, Originality, usefulness, Feasibility, Creativity
[83]	Lujun	Undergraduate juniors in Mechanical Engineering	Petroleum pumping unit	Sketches	Quantity, Quality
[84]	Cardoso, Goncalves, Badke-Schaub	Industrial design students	Human transportation system	Sketches	Fluency, Flexibility

Reference no.	Author/s	Participant	Problem name	Result	Metric used
[85]	Chan, Fu, Schunn, Cagan, Wood, Kotovsky	Mechanical Engineering undergraduates	Device to collect energy from human motion	Sketches	Quantity, Novelty, Quality
[86]	Cardoso, Goncalves, Badke-Schaub	Bachelors and masters student from industrial design course	Human transportation system	Sketches	Fluency, Flexibility, Originality
[87]	Chan, Fu, Schunn, Cagan, Wood, Kotovsky	Graduate and undergraduate students	Device to collect energy from human motion	Sketches	Quantity, Quality, Novelty
[88]	Wilson, Rosen	Undergraduate mechanical engineering students	Leg immobilization device	Sketches	Novelty, Variety
[89]	Vishwanathan, Linsey	Senior undergraduate students	Peanut Shelling machine	Sketches	Quantity
[90]	McKoy, Hernandez, Summers, Shah	Undergraduate engineering students	1) Bicycle lock 2) Mechanism to grab books	Sketches	Quality, Novelty

Analysis of the 34 user studies collected shows that each study has at least one characteristic element which is different from the other studies. Table 2-5 also shows the different design problems that have been used by researchers in experiments. In 34 user studies, a total of 37 different design problems have been used. The list of problems and their statements is can be found in Table A-1. Thus, problems used by researchers in experiments constitute one source of difference between studies. Since design problems are amongst the few variable which can be controlled by the researchers, an opportunity exists for the design research community to enable researchers to locate and reuse design

problems to help reduce one source of variation between studies. However, reusing existing problems may not be viable in every situation. Other alternative ways to achieve this objective may be by:

- Using benchmarked design problems which have been used and tested in experiments and proven to be useful in studies to act as standard problems.
- Using problems which are similar to ones used in other studies where experiment conditions and requirements are similar.

Establishing a repository of benchmarked design problems for experiments is an opportunity which the design community should address. This would essentially require a large collaborative effort from the community since these problems need to be tested and accepted after validation in experiments. Alternatively, similar problems can be used in situations where an existing problem cannot be reused. Such an approach would be extremely beneficial for researchers to help them find and design problems suitable for their study. For example, a researcher who knows the requirements needed in his/her design problem can compare the requirements to an existing problem which has been used earlier in an experiment. This would ultimately help in problem selection process and reduce the chances of the problem not being appropriate for the task at hand.

2.4. Research questions

Based on literature reviewed and analysis of collection of user studies, the following research questions have been identified:

User studies published in the area of design creativity differ from each other on a number of fronts such as metrics for evaluating results, participant type and use of design problems to name a few. For this research the pattern of design problem usage by researchers is used to identify the need for using same or similar problems in creativity research. Analysis of design problem usage pattern can also help in identifying design problems which have been frequently used by researchers which could be used as benchmark problems if needed. This germinates the first research question (RQ):

RQ1: *How can the pattern of design problem usage in creativity research be identified?*

An alternative to reusing the same design problem in experiments can be to use problems which are similar to each other in some respect. Researchers have used justified conceptual problems as being similar considering different aspects such as number of requirements, word count, expected characteristics in solution, problem domain, interconnectedness amongst functional units and results from previous studies [15][28]. It is felt that a systematic method for comparing two conceptual design problems can be useful for researchers trying to compare problems before use in experiments. Based on prior work in problem evaluation and the need for a systematic method for comparing problems based on their statements, the following research questions are formulated:

RQ2: *How can conceptual design problems be compared using their problem statements?*

RQ3: *How can natural language representation be used to compare conceptual problems?*

Table 2-5 illustrated the fact that design problems have been a source of difference between user studies in engineering design creativity. While this may be due to difference in experiment's requirements, analysis of the impact of this difference on conclusions drawn from creativity experiments can help bring into perspective the importance of using existing or similar problems for experiments in engineering design creativity. The following research question is formulated to study whether design problems can affect the conclusions that are drawn from experiments in creativity:

RQ4: *Does the choice of design problem have any influence on effectiveness of interventions used in experimental studies in design creativity?*

CHAPTER THREE

RESEARCH APPROACH

The following research tasks are used to answer the questions pertinent to this research:

- a. Graph based analysis of design problem usage in creativity research.
- b. Comparison between design problems based on structural elements of a problem.
- c. Comparison between design problems based on Latent Semantic Analysis of problem statements.
- d. Meta-regression analysis with design problem size as a moderator variable.

Table 3-1 summarizes the various research tasks to be accomplished and corresponding research questions answered by each task.

Table 3-1: Overview of research tasks and corresponding research questions answered

Research task	Research question answered			
	RQ1	RQ2	RQ3	RQ4
Graph based analysis of design problem usage (Chapter 4)	✓			
Comparison based on structural elements (Chapter 5)		✓	✓	
Comparison based on LSA (Chapter 5)		✓	✓	
Meta-regression analysis of user studies (Chapter 6)				✓

3.1. Graph based analysis of design problem usage

For illustrating the pattern of design problem usage by researchers in the area of design creativity (**RQ1**), a graph based representation is used. The collection of 34 user studies shown in Table 2-5 is used for generating a graph depicting how different design problems have been used by researchers. For this, the list of authors from all 34 studies is extracted and the corresponding design problem used by them is identified. The list of authors is further refined by separating student authors from faculty authors on these 34 user studies. Thereafter, an incidence matrix is generated between faculty authors and design problems used by them. A bipartite graph is then generated to show how different researchers have used the 37 design problems extracted from these studies. Thereafter, a projection of this bipartite graph is used to generate a one mode graph where design problems represent the nodes of this graph, while the common researcher between two design problems act as links/edges between two nodes. A second graph showing how design problems have been used in the 34 articles published is also used for addressing this research question. Here, a bipartite representation of connection between the 34 papers and design problems used by them is generated.

3.2. Comparison of design problems based on structural elements

Based on design problem characteristics identified by different authors in the past and assessment of collection of design problem statements obtained from the 34 user studies, five structural elements of a design problem (viz. goal, functional requirement, non – functional requirement, end user information and reference to an existing product) are

used to compare conceptual design problems with each other [91]. These elements can be identified in problem statements, which can then be compared to each other based on either the number of elements present or the presence or absence of these elements. For example, the following problem statement contains two goals (*design an end connector and a socket*), two functional requirements (*attachment and removal*), two non – functional requirements (*quickly and safely*) and a reference to an end user (*ground crew of commercial airliners*).

‘Large commercial aircraft are refuelled between flights from mobile tankers containing aviation kerosene. The ends of the hoses from tankers are fitted with special connectors that are attached by ground crew. A hose end connector and socket are to be designed to allow the attachment and removal of hoses to be carried out quickly and safely [92]’.

Using this approach, the size of a problem statement can also be evaluated by counting the number of elements (goals, functional requirements and non – functional requirements) present in the problem statement. Thus, this research task would help answer two research questions about comparison of design problems based on natural language representation of problem statements (**RQ2 and RQ3**).

3.3. Comparison of design problems based on Latent Semantic Analysis

As a second approach to design problem comparison, linguistic analysis of problem statements using Latent Semantic Analysis (LSA) is used. LSA uses statistical computations to a large corpus of text in order to extract and represent the contextual usage

meaning of words, sentences and documents [93]. LSA enables comparison between phrases of words based on their contextual meaning obtained from text corpus. The underlying assumption for this method is that words and phrases that have similar meanings will be used in similar pieces of texts. Similar words or phrases are represented by values close to one, while values close to zero or less indicate dissimilarity. The problem statements for design problems extracted from the collection of user studies used for answering RQ1 (Section 3.1) are compared against each other for their semantic similarity using LSA. Thus, LSA provides a second method for problem similarity assessment based on its natural language representation and helps answer research questions RQ2 and RQ3.

3.4. Meta-regression analysis with problem size as a potential moderator

Meta-analysis, also known as analysis of analyses, is a statistical technique commonly used in medical science and psychology to compare and assimilate results from different studies [94–97]. It helps to combine and compare the observed effects from treatments or interventions used in different studies to draw a summarized picture of the state of a research field. User studies in design creativity can be seen analogous to medical trials, where a treatment in the form of a design method or tool is used in an experimental setting and its efficacy is evaluated based on the observed effects on human participants. Meta-analysis can help summarize the effect of different creativity enhancing ‘interventions’ or methods which have been tested in the past, thereby enabling an indirect comparison between different methods tested. This task is a precursor to meta-regression, which would be used for answering the following research questions: Does the choice of design problem influence effectiveness of methods being tested in user studies in design

creativity (**RQ4**). Specifically, user studies using examples as an intervention will be used for understanding the influence of design problems on the effectiveness of example interventions. A regression model with design problem size as predictor of effect size obtained from studies using examples as interventions will be used to analyze if a problem size is related to effectiveness of example interventions.

CHAPTER FOUR

DESIGN PROBLEM USAGE GRAPHS

Objective: To understand and identify the pattern of design problem usage.

Research task: Generation of graph diagrams showing design problem usage.

A graph consists of two finite sets, a set of points called vertices and a set of connecting lines called edges such that each edge connects two vertices called the endpoints of the edge [98]. Graphs provide a useful way of representing how things are connected to each other (or how things are disconnected). Table 2-5 shows how different studies have used different design problems which also provides an opportunity to study the pattern of design problem usage. This would help identify problems which have been reused and shared between researchers and between different studies. In the long run, such problems could be tested and used as standard problems. For this, a bipartite graph between different researchers and the design problems used by them is used.

4.1. Graph based representation of design problem usage

In order to generate the bipartite graph between researchers and problems used by them, 37 design problems used in the 34 studies shown in Table 2-5 is used. Each author's name on the study published is extracted and associated with the design problem that was used by them. Since authors on papers may be one time authors who may not be active in design research community, additional pruning is used to filter out student authors from faculty. The final graph representation contains a relationship between different faculties

and the design problems used by them. Table 4-1 shows a portion of coding process used to separate the students and faculties for the network. A complete list of authors extracted from the 34 paper, the design problems used in each study and researcher-problem relationships can be found in Table B-1. All students in the list of authors were eliminated from the graph since most of them are expected to be one time authors who may not be active in design research again. In the remaining list of authors, each faculty was associated with the design problem used when conducting an experiment.

Table 4-1: Sample coding procedure for researcher - design problem graph

Reference no.	Author/s	Problem name
[44]	S. Kim (Faculty) Y. Kim (Faculty)	Subway improvement
[24]	Sonseca (Faculty), Mulet (Faculty), Chakrabarti (Faculty)	1) New table, 2) Tubular map case 3) System for gathering wire, 4) Table for offices
[45]	S. Kim (Faculty) H. Kim (Student) Jin (Faculty)	Wearable binocular
[64]	Rogers (Student) Salustri (Faculty)	Bi/tri cycle
[67]	Naim (Student) Lewis (Faculty) S. Schmidt (Student) Viswanathan (Faculty) Linsey (Faculty) McAdams (Faculty) Campbell (Faculty) Poppa (Faculty) Robert (Faculty)	Water lifting device Peanut shelling machine
[25]	Hernandez (Faculty) C. Schmidt (Faculty) Okudan (Faculty)	Traffic light using LED

4.2. Bipartite graph between different researchers and design problems

With the dataset containing relationship between different researchers and design problems used by them, a bipartite graph representation of the pattern of design problem usage is obtained. This graph is shown in Figure 4-1. For brevity, the vertices of this graph have been numbered sequentially starting from 1. The list of vertices and the corresponding researchers and design problems used in graph is shown in Table B-2. An edge between a researcher and a design problem exists if the researcher has used the design problem in the list of 34 studies shown in Table 2-5. Weights are assigned to edges based on the number of times a problem has been used by an author. Thus, if author ‘A1’ has used problem ‘DP1’ three times, the corresponding weight assigned for the edge between the two nodes is three. This helps in highlighting the propensity of different researchers for using design problems chosen by them in the past.

Observations from the bipartite graph

The following observations can be made through visual examination of the bipartite graph:

- Researchers have used different design problems for experiments. Researchers possibly design and use problems based on their experience and experiment’s requirements. However, every time a new problem is used for an experiment, it becomes a responsibility of the researcher to determine its appropriateness for the study. If a new problem has not been tested earlier, establishing the fact that design problem chosen will not convolute the results obtained from experiments is difficult.

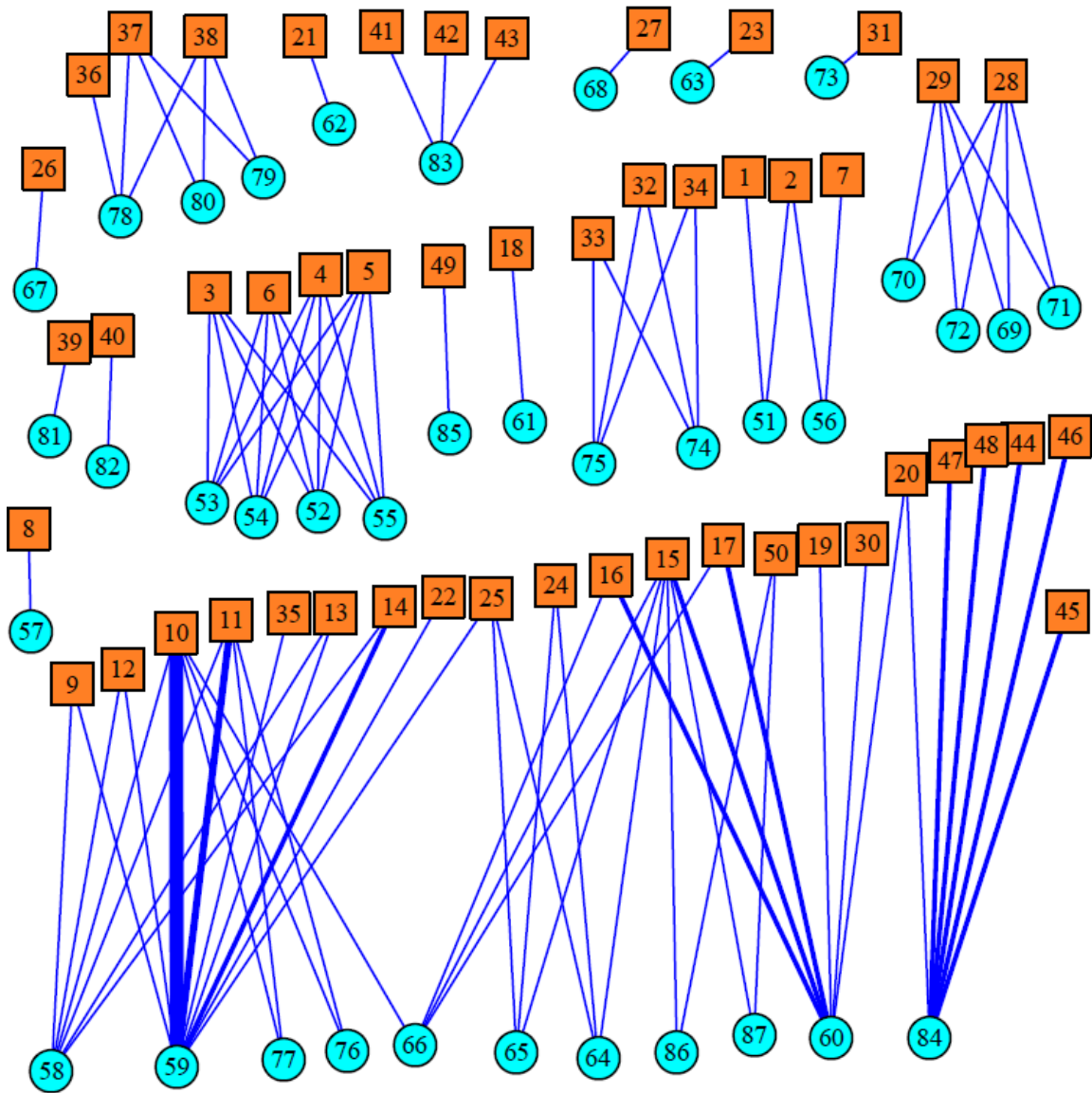


Figure 4-1: Bipartite graph showing design problems used by different researchers. Square vertices represent different researchers while circles represent design problems used by them.

- Some design problems have been used multiple times by researchers in their experiments. For instance, researcher Julie Linsey (node 10) has used the ‘Peanut shelling machine’ problem seven times in different experiments. This design problem

also had the highest node degree value of nine as compared to other design problems as shown in Figure 4-2. This means that a total of nine researchers have shared this design problem in experiments. However, analyzing the organizational affiliations of these nine researchers indicates that they share or have shared the same affiliation with researcher Julie Linsey at some point of time. This indicates that problem sharing and reuse is practiced by researchers possibly within their academic affiliations or research groups. On the other hand, design problem 10 ‘Traffic light using LED’ (vertex 60) has been used by six researchers with different academic affiliations in two different experiment giving some indication of problem sharing between researchers with different affiliations as well.

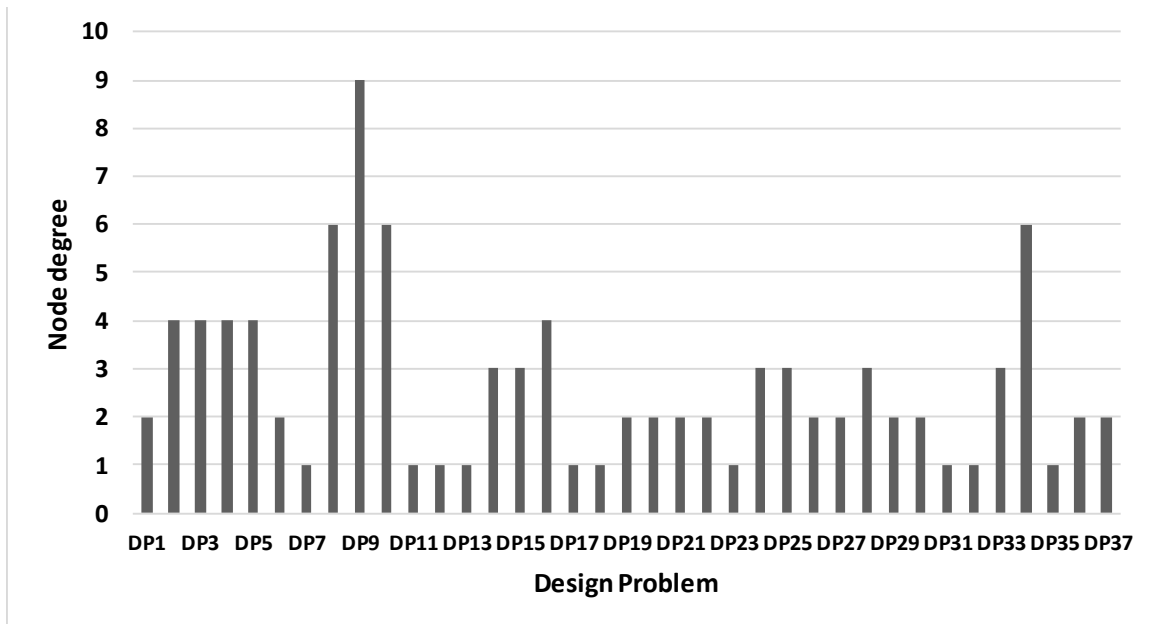


Figure 4-2: Node degree for different design problems.

4.3. One mode representation of design problem graph connected by researchers

A one mode graph projection for the bipartite representation of connection between design problems and researchers can be used to obtain a graph of design problems which are connected to each other by the researchers using them. The projected graph is shown in Figure 4-3. The graph shows that there is a group of problems which have been shared by researchers in their experiments. Such problems have a higher chance of being reused in experiments as compared to the other ‘isolated’ problems since more researchers have used this group of problem in their experiments.

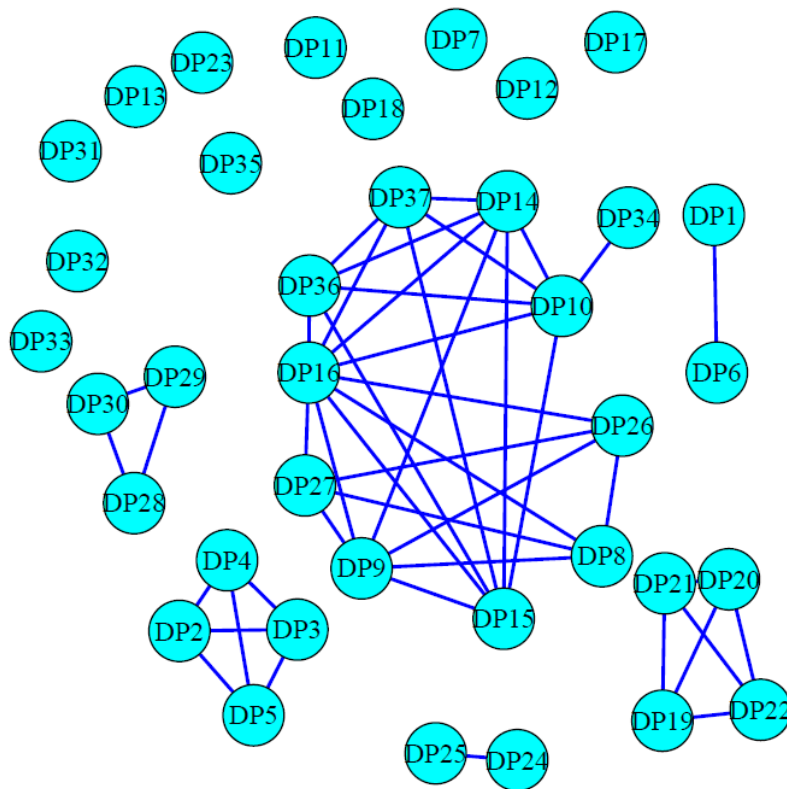


Figure 4-3: One - mode graph projection showing connection between different design problems.

Analysis of one – mode design problem graph

Three network metrics have been used to analyze the one mode design problem graph [99]. These are:

- *Node degree*: The number of edges connected to a vertex. The degree of a node signifies the importance of that node in the given graph. For an undirected graph with n nodes, the degree of node i can be expressed in terms of adjacency matrix A_{ij} as:

$$k_i = \sum_{j=1}^n A_{ij} \quad (1)$$

Figure 4-4 shows a bar graph of node degree evaluated for all nodes in the one mode design problem graph. Node degree distribution shows that sixteen nodes (design problems) have a node degree of either zero or one. This means that approximately 43% of design problems in the graph have either one or no connection with other problems. Based on the sample studies used for this analysis, this indicates that such problems have probably not been used by other researchers in their experiments.

- *Communities*: Communities, which are also referred to as clusters or modules are group of vertices which have very high edge density with vertices within the group, but low densities between these groups [100]. They are location of fault lines, about which graphs tend to separate. Community detection in a graph can be accomplished by several algorithms [99]. One such algorithm is based on betweenness centrality which is defined as the number of shortest paths which pass through a vertex in a network. Analysis of design network using this algorithm identified 17 different communities in the network, as shown in Figure 4-5 (including the isolated vertices, which are

considered as separate groups)¹. This high number indicates that design problems are reused and connected within certain groups of researchers.

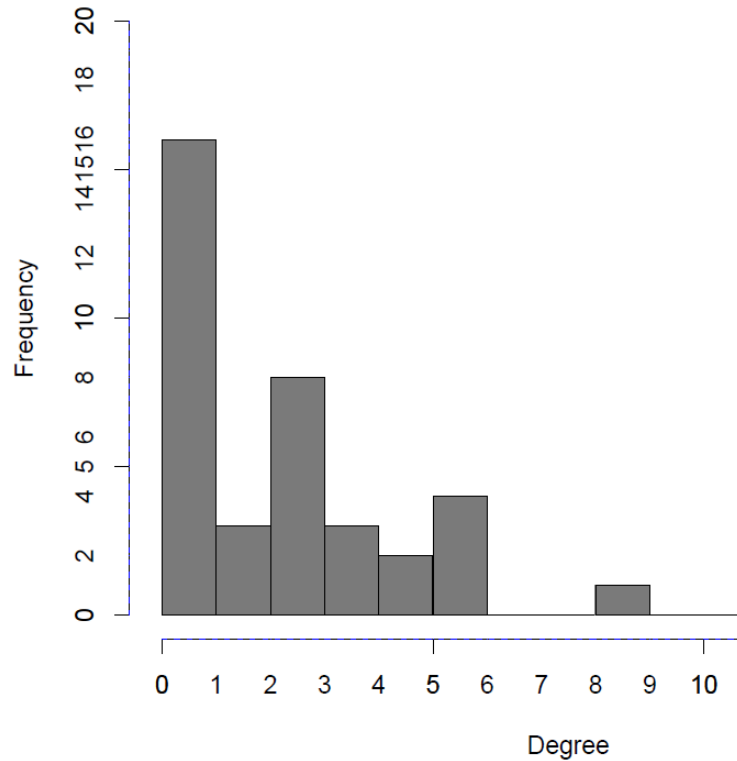


Figure 4-4: Bar graph of node degrees for all vertices of one-mode design problem network

- *Assortative mixing*: Assortative mixing refers to the likelihood of association between similar vertices in a network diagram. For instance, a graph which shows signs of high degree nodes tending to show affinity towards other high degree nodes and vice versa is said to show assortative mixing by degree [99]. The level of assortative mixing in a network is evaluated using the assortative coefficient defined by

¹ The algorithm was implemented using the available package 'igraph' in R.

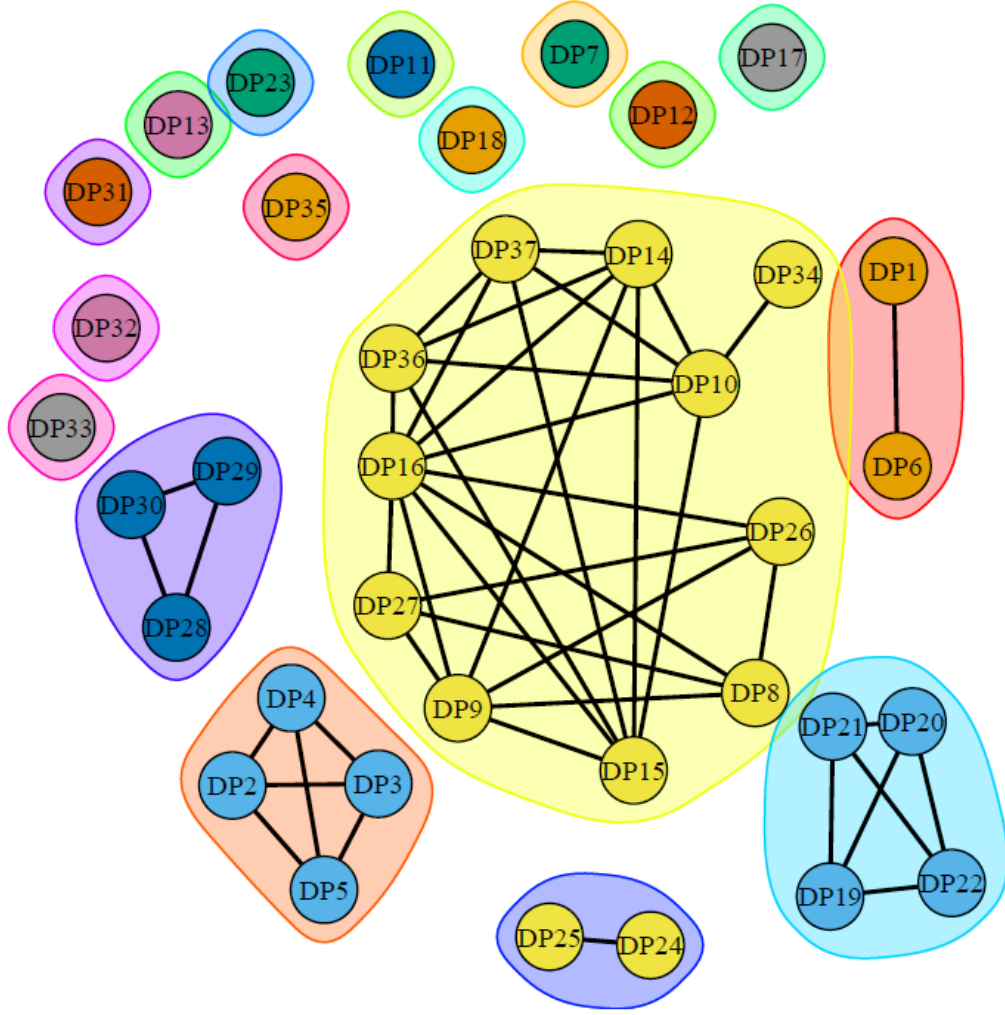


Figure 4-5: Communities identified in design problem network using edge betweenness algorithm

$$r_{degree} = \frac{\sum_{ij} (A_{ij} - k_i k_j / 2m) k_i k_j}{\sum_{ij} (k_i \delta_{ij} - k_i k_j / 2m) k_i k_j} \quad (2)$$

where A_{ij} is the adjacency matrix, k_i and k_j represent the node degree of vertices i and j , m is the total number of edges present in network and δ_{ij} is the Kronecker delta.

The value of assortative coefficient for the design problem network was 0.70, which indicates a high affinity between design problems with similar degree scores. A high value of assortative mixing means that problems which are shared more often clump around each other and hence have a higher probability of connecting together. The reason for this can be attributed to the presence of separate communities or sub-groups in the network, where problem sharing is practiced within some communities.

4.4. Bipartite graph between design problems and papers using them

The bipartite graph between 34 studies shown in Table 2-5 and design problems used in them is shown in Figure 4-6. An edge between two sets of vertices exists when the problem has been used in a paper. The graph shows that problem reuse between experiments is not common for the 34 studies used for generating this graph. The frequency of use of design problems in papers published is shown in Figure 4-7. Design problem represented by node 43 (Peanut shelling machine) is one problem which has been used seven times in studies considered for this analysis. Three other problems represented by nodes 44, 67 and 68 have been used in two experiments. Apart from these, other design problems have one experiment associated with them indicating low problem sharing between experiments.

4.5. Discussion

As a part of the process of identifying the level of connection between researchers in design problem usage, a graph based representation is used. The bipartite graph between researchers and design problems used in 34 user studies published between the years 2000

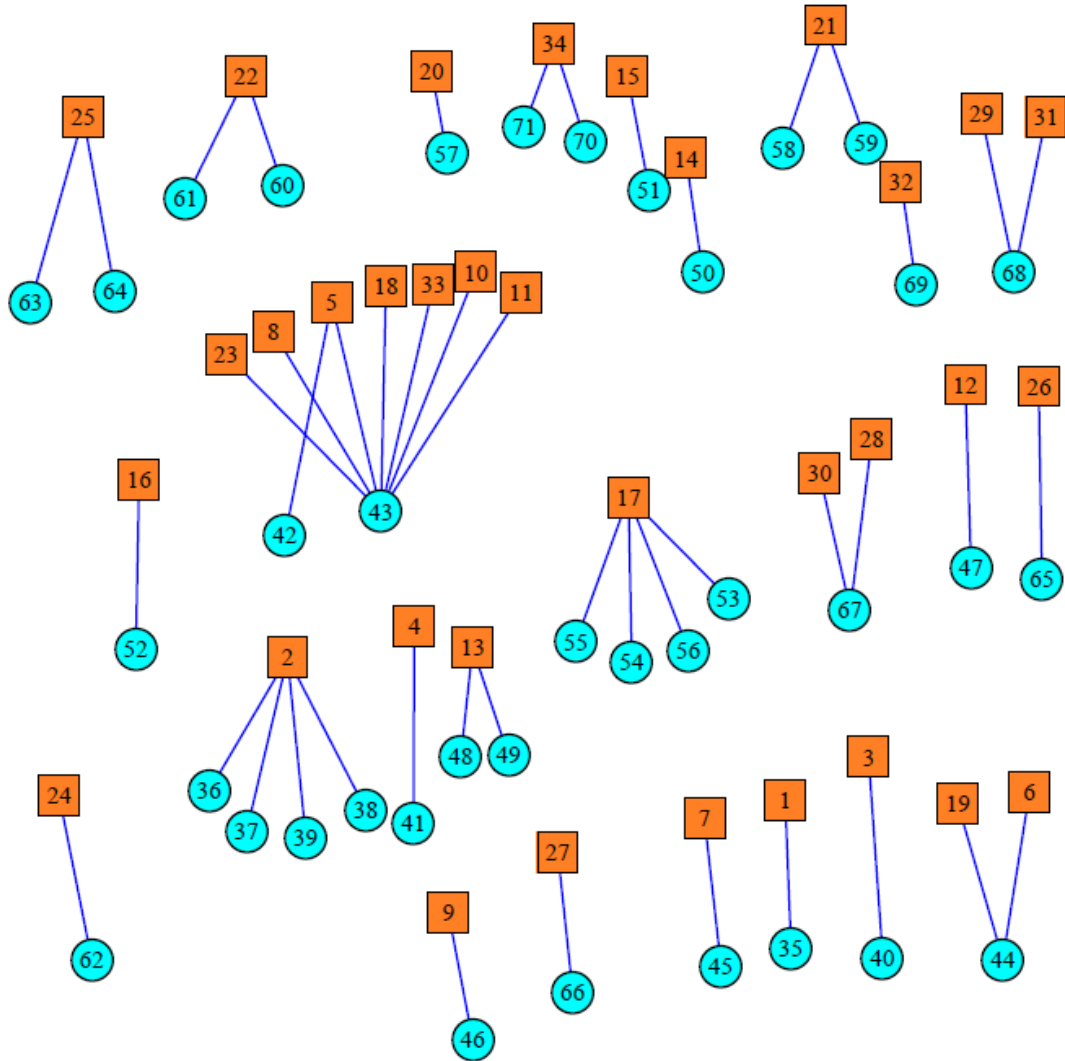


Figure 4-6: Bipartite network showing design problems used in different papers published. Square vertices represent different papers while circles represent design problems used by them.

and 2014 indicates that some problem sharing between researchers occurs within their academic affiliations and research groups. The ‘Peanut shelling machine’ problem has the highest number of researchers associated with it. All nine researchers who have used this design problem belong to the same academic affiliation or have shared the same affiliation at some point of time. However, the one mode representation of design problems in

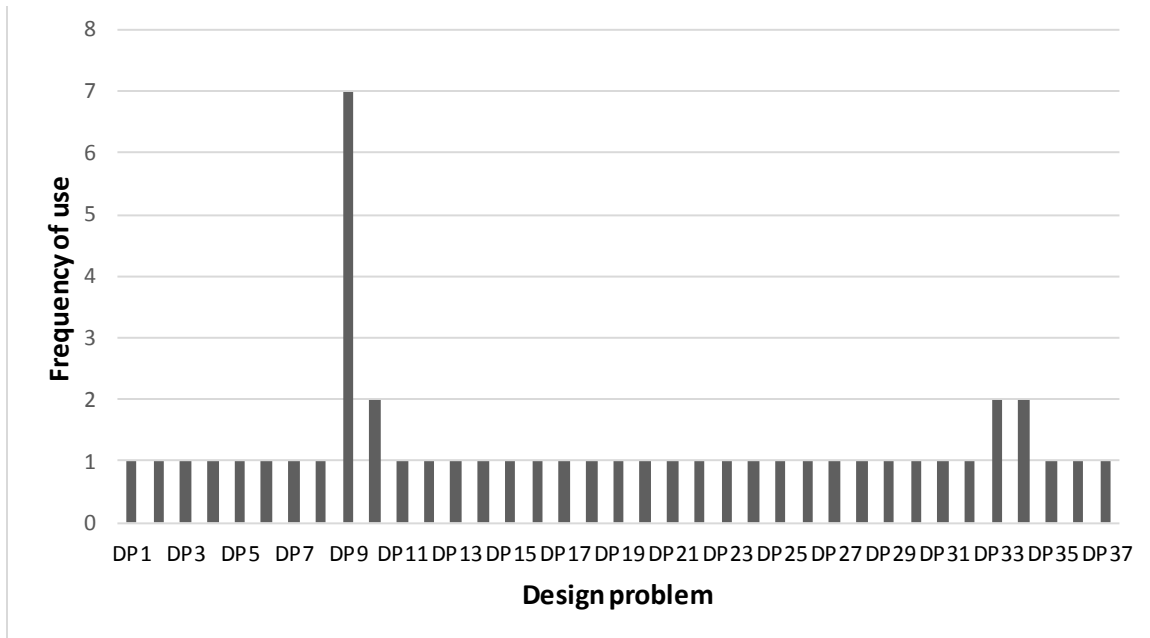


Figure 4-7: Design problems and their frequency of use in user studies published

Figure 4-3 indicates that approximately 43% of the design problems have a degree score of zero or one. This indicates low problem reuse between fifty researchers in 34 studies considered. At an individual level, some problems have been used multiple times by a researcher for different experiments. In Figure 4-1, design problem represented by node 59 (Peanut shelling machine) has been used by researcher represented by node 10 (Julie Linsey) multiple times indicating an inclination towards reusing the same problem in experiments at an individual level. Similarly, design problem represented by node 10 (Traffic light) has also been used twice by three researchers indicating some degree of individual preference for reusing an existing problem. The one mode representation for design problems obtained by projecting the bipartite graph between researchers and problems used indicates the presence of 17 communities in the graph. It is possibly a result

of small research groups within the academic or professional work environment of the researcher. A high value assortative connection (0.70) between design problems is indicative of the fact that problems used within a community or research groups have higher probability of being reused by members of that community, while other research groups continue using problems prevalent in their respective groups. Another bipartite graph between papers and design problems used in them (Figure 4-6) also indicates limited problem sharing between studies. 33 out of 37 total design problems considered for analysis have been used once in an experiment and not reused again.

4.6. Comments and recommendations

Based on the graph based analysis of design problem usage, the following observations can be made:

- Sharing of design problems between researchers for experiments in user studies is probably not a widespread practice. Thus, an opportunity exists for design community to reduce the number of variables between different studies by sharing same design problems across studies.

Recommendations: Establishment of a design problem repository containing ‘standard’ design problems suitable for user studies. Problem standardization should be accomplished through consensus between researchers after experimental testing. Guidelines should also be established to enable researchers to choose a suitable design problem for their study based on experimental requirements.

- An opportunity exists for the research community to adopt a collaborative research approach to enhance element sharing between different studies and experiments.

Recommendations: A common design research methodology is needed which can act as a guideline for researchers. The methodology should include various research aspects including problem selection, participant selection, method of assignment to condition and choice of metric. The methodology should also include the idea of selecting problems which are similar when the use of an existing design problem is not feasible.

CHAPTER FIVE

ASSESSMENT OF DESIGN PROBLEM SIMILARITY

Objective: To enable similarity comparison of conceptual design problems based on problem statement.

Research task: Methods to determine problem similarity based on their statement.

Durand and coauthors [15] highlighted the interaction between design methods and design problems, and observed that the type of design problem used affects the results across a set of creativity metrics. For this, they used two different design problems and tested them on two groups whose creativity scores for metrics quantity, quality, novelty and variety were compared. The results indicated that while the two problems produce same quantity and variety of ideas, the quality and novelty of ideas produced were different. Thus, evaluating the similarity between two problems based on results obtained from experiments may be difficult. This provides the opportunity of comparing design problems based on their representation which will be useful for a researcher who is trying to select a problem for his/her study. Researchers have used similarity of problem domain, number of requirements, concepts expected in the solutions for two problems, domain similarity of two problems, the size of design problem in terms of number of functional units and structural similarity of problem statement [15][28] as a means to justify how two problems are similar. Yet, it is felt that a systematic approach to compare conceptual design problems would be helpful for researchers while selecting problems appropriate for their

study or when the experiment's requirements demand selection of two similar problems. In order to facilitate this, two approaches are proposed and discussed in subsequent sections:

- Approach 1: Estimating size of a design problem by identifying structural elements of design problem.
- Approach 2: Semantic analysis of problems based on their representation.

5.1. Approach 1: Estimating size of a design problem by identifying structural elements of design problem

Based on the characteristics of a design problem identified by researchers shown in Table 2-2 and analysis of 37 design problem statements collected from 34 user studies (Table 2-5), a list of five characteristics are identified as structural elements of conceptual design problems [91]. These five elements have been shown in Figure 5-1. These elements are: i) goal of the problem, ii) functional requirements in problem statement, iii) non – functional requirement in problem statement, iv) information about end user of product and v) reference to an existing product. These elements help estimate the size of design problem.

Summers and Shah [37] defined problem size as a sum of different variables which are present in specific problem representation. For conceptual problems, however, these variables are not always stated in the problem statement. An alternative way for assessing problem size of conceptual problems is to use the count of number of goals, functional

requirements and non-functional requirements or the presence or absence of information about end user and reference to an existing product.

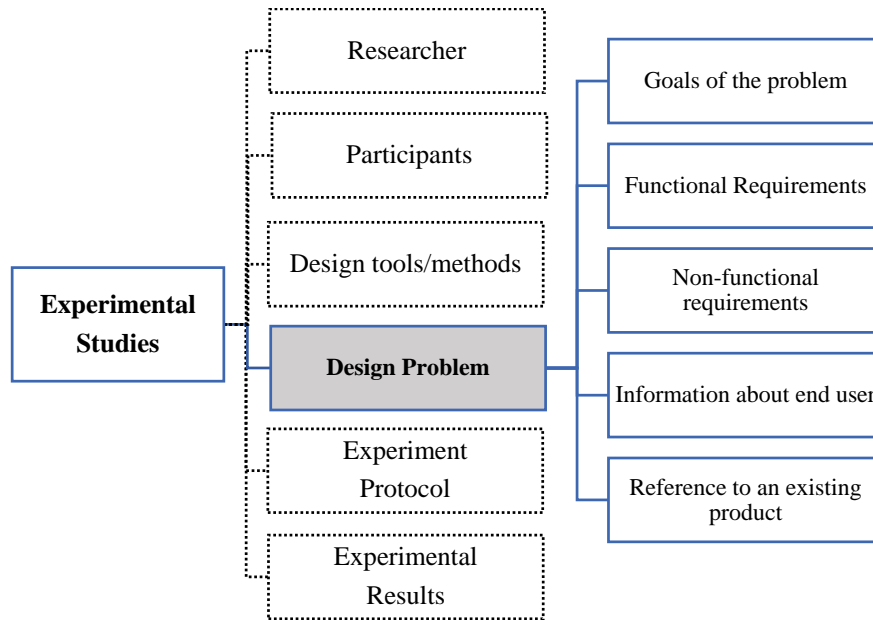


Figure 5-1: Structural elements of a design problem

For identifying these elements, a set of rules or definitions is required, to serve as a protocol for identifying the five elements in a conceptual design problem. The definitions of these structural elements of a design problem are explained in the subsequent sections.

- **Goals of the problem:** The final objective of every design task is determined by the problem goal. Ideally, every design task should be associated with at least one goal. Goals help to establish the final outcome expected from the solution space explored. The problem statement may require the reader to design more than one artifact, in case of which two goals are associated with a problem statement.

- **Functional requirements (FR):** Functional requirements are the primary functions of the design, or the tasks that the product needs to perform without giving any reference about how the task should be accomplished [101–103]. They define behaviors and actions that the designed artifact needs to support [104]. Functional requirements are generally associated with action verbs linked to an object (a noun generally) [4]. Nouns derived from verbs like a washing machine, reading device etc., can also be used to represent functional requirements of a product in a problem statement. Sometimes, additional functions may be added to an existing product, in case of which these additional function becomes a part of the product's functional requirement. For example, an additional function of washing dishes may be appended to a washing machine, thereby increasing the functional requirement. Since problem representations sometimes contain implied requirements, in order to reduce subjectivity in identification, selection of functional requirements in a problem statement should only be based on explicitly stated requirements. To identify FRs, one should look for:
 - Action verbs like move, work etc. associated with objects (objects include nouns on which the action verbs act like sprinkle water, dry hair).
 - Primary functions of the design (eg. move objects, lift, transport objects).
 - These could also be nouns derived from verbs (eg. washing machine, toaster)

Additionally, if there are two objects associated with one primary function, it should be counted as two separate FRs (eg. move object X & object Y is counted as two FRs).
- **Non-functional requirements (NFR):** These are 'non-functional' requirements, which do not determine the primary functions of the product, but cast a bound on the overall

shape, size, cost, operation and selection of the design [104–109]. They specify external constraints that the product must meet and place restrictions on how the user requirements are to be met [110]. Typical NFRs include requirements that describe the non-behavioral (non-action, non-function) aspects of the product such as performance targets, usability, reliability, durability and other physical specifications [111]. Identification of non – functional requirements should also be based on explicitly stated texts in the problem statement, to reduce chances of subjective interpretation when implied requirements are stated.

- Information about end user: A new product need or requirement arises only when someone is in need for it. In other words, every new product being designed or developed is meant for a customer or user, who is in need for it [112]. If the end user of the product is specified in the problem statement, it enables the problem solver to identify the final customer, and develop solutions keeping in mind their likes and dislikes. It also helps the problem solver in understanding additional unspecified attributes that the product must possess to satisfy end user.
- Reference to an existing product: Whether the problem statement contains any reference to an existing product or not is another characteristic which can be used for comparing two design problems. For instance, if a design problem requires ‘design of a hair dryer’ some degree of constraint is imbedded in the problem solver’s thought due to the fact that he/she might have seen or used such a device and may not be able to generate a novel design for it. This characteristic can at times be a perplexing entity since realizing the existence of the product stated in problem statement depends on

personal attributes of the problem solver like demography, culture and educational standing.

5.1.1. Examples showing identification of elements in design problems

Two examples are presented to illustrate the identification of design problem elements.

Example 1: The first example problem requires the problem solver to design a new hair dryer [92], shown in Figure 5-2 with annotations for elements identified in the problem.

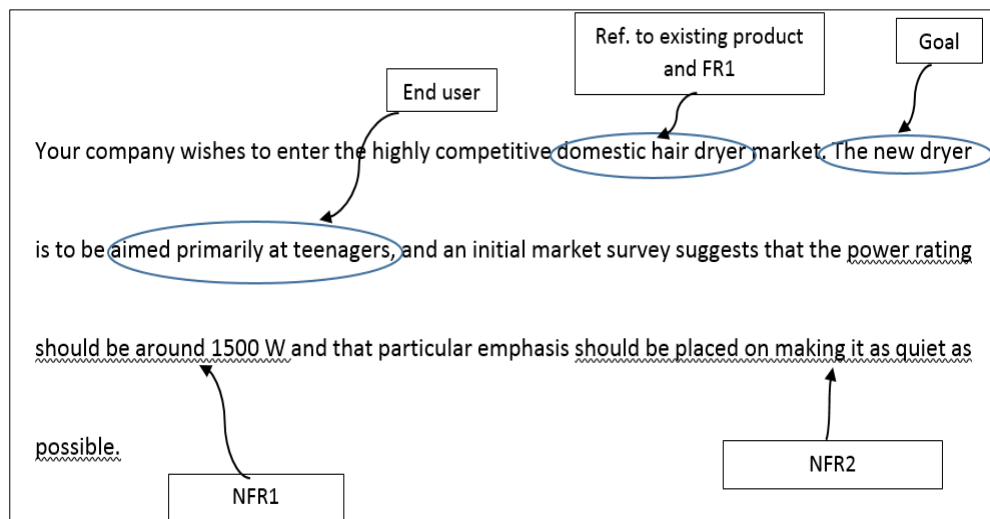


Figure 5-2: Example 1 showing characteristic identification in a design problem

Example 2: The second example problem requires the problem solver to design products that use sunlight for heating and cooking food [92], shown in Figure 5-3 with annotations for elements identified in the problem.

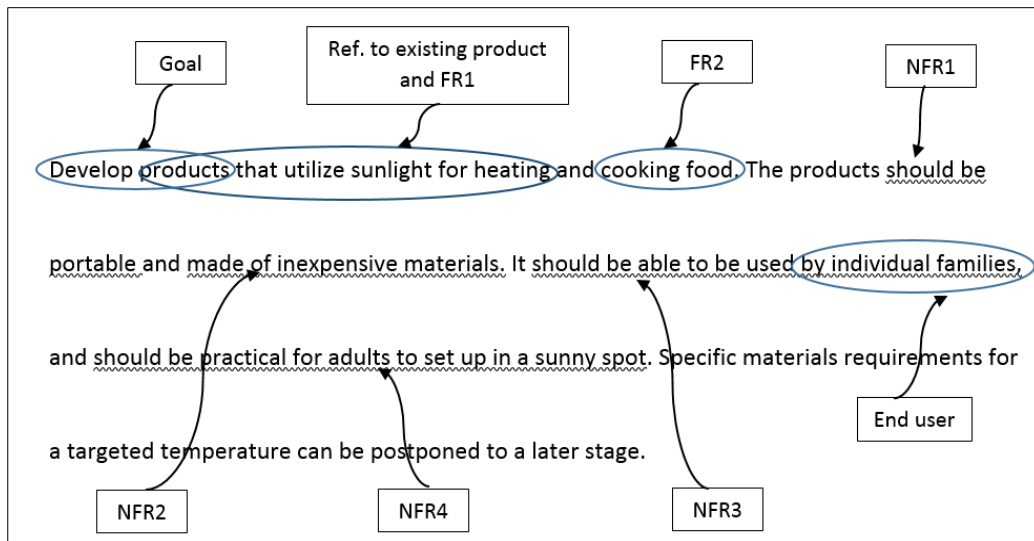


Figure 5-3: Example 2 showing characteristic identification in a design problem

5.1.2. Protocol for identification of elements in design problems

Comparing two design problems is possible after evaluating their size based on the five structural elements. The number of each element or its presence or absence provides an indication of how big the problem is. However, a limitation of this method is that it cannot capture the knowledge and solvability of the problem. For instance, a problem seeking ‘design of a pen’ may contain the same number of elements as another problem seeking ‘design of a vacuum cleaner’, yet both problems are different with respect to the knowledge content and effort needed in solving them. Still, this approach can help researchers to obtain a preliminary measure of problem similarity, after which other methods to measure problem ‘complexity’ can be used for detailed similarity assessment [37]. The proposed approach for similarity comparison is aimed at ascertaining some level of deliberation during problem selection by researchers for user studies in design research. For illustrating the proposed approach, two design problem examples shown in Figure 5-2

and Figure 5-3 are used again. The three step procedure for design problem similarity assessment using this approach is as follows:

Step 1: The problem statements are provided to the researcher/evaluator who carries out the process of element identification.

Steps 2: For each characteristic identified in the problem statement, the appropriate score is entered into a table (see Table 5-1).

- A count of the *number of goals, number of functional requirements and number of non-functional requirements* in the problem statement are recorded. For example, the ‘hair dryer’ problem in Figure 5-2 contains 1 goal, 1 functional requirement and 2 non – functional requirement.
- If there is information about the end user a value of 1 is assigned, if there is no information about the end user a value of 0 is assigned for *End user*. The ‘hair dryer’ problem used in example 1 (Figure 5-2) has information about the end user and thus a score of 1 is assigned under the ‘end user’ column.
- If the problem statement references an existing product or a modification or improvement to an existing product, a score of 1 is assigned else a 0 is assigned. Again, in the ‘hair dryer’ example, an improvement to an existing product is sought, which means that the problem contains a reference to a product which already exists. Thus, a score of 1 is assigned to the ‘reference to existing product’ column.

Step 3: The two design problems are compared based on the scores assigned for each characteristic in the table. Table 5-1 shows an example of such a comparison table obtained

for ‘hair dryer’ (Example 1: Figure 5-2) and ‘solar device’ problems (Example 2: Figure 5-3).

Table 5-1: Example of scoring system and problem comparison based on approach 1

Design Problem Element	Design Problem	
	Example 1	Example 2
Number of goals	1	1
Number of Functional requirements	1	2
Number of non-functional requirements	2	4
End user information (Yes=1/No=0)	1	1
Reference to an existing product (Yes=1/No=0)	1	1

Comparison of the two design problems based on the scores obtained in Table 5-1 reveals that example 2 (solar device) has more ‘information’ contained in the form of functional and non – functional requirement. Thus, the two design problems can be considered as ‘dissimilar’ as far as the information contained in the form of functional and non – functional requirements are concerned.

5.1.3. Inter-rater reliability test for approach one

The definitions for the structural elements used for comparing problems may sometimes be interpreted differently by different people. For instance, the way a problem is stated can have an impact on whether a requirement is considered as a functional or a non – functional requirement [105]. People may have individual perspectives about what constitutes functional and non – functional requirements, which may lead to differences in the result of problem comparison approach discussed above. Therefore, in order to verify the agreement between people with respect to the procedure for comparing problems, an

inter-rater reliability study is conducted with four independent evaluators (three graduate students with understanding of design methodology and related concepts, one associate professor teaching design methodology). The procedure used for this study is outlined below:

- a. All Evaluators are trained about the context of the experiment and the nature of design problems under consideration.
- b. Evaluators are then introduced to the five elements which typically constitute a design problem, accompanied by definition of each characteristic presented in the form of a tabulated rubric (see Table C-1).
- c. A completed example of element identification procedure, such as the one used in Figure 5-2 is shown, with emphasis on how the prescribed definitions are used in identifying the elements.
- d. The Evaluators are then given four design problems to assess and evaluate the scores for each element. They are also asked to annotate their choices, to verify and eliminate any ‘outlier’ choices made.

The four design problems used and the annotated results for all four Evaluators are shown in Figure C-1 to Figure C-16 in Appendix C. To assess the consistency and agreement between the four Evaluators, an inter-evaluator reliability test is conducted for each characteristic separately. A high reliability for each characteristic would result when the evaluators agree to the definitions and identify the same number of elements in the problem statement. Agreement for each characteristic is evaluated based on the scores or counts assigned by evaluators for the four design problems. Thus, for assessing whether

the four evaluators identify the same number of functional requirements in a problem, the counts assigned to the characteristic ‘functional requirement’ is used.

Since this study resulted in scores from multiple Evaluators on different data scales, inter-evaluator reliability was assessed using Krippendorff’s alpha [113–115]. Krippendorff’s alpha is widely used in content analysis where trained readers are used to categorize textual data. In its general form, Krippendorff’s alpha is represented as [113]

$$\alpha = 1 - \frac{D_o}{D_e} \quad (3)$$

where D_o is the observed disagreement among values assigned to units of analysis and D_e is the disagreement one would expect when the coding of units is attributable to chance rather than to the properties of these units. In case of perfect agreement, $D_o = 0$ and $\alpha = 1$. When observers agree by chance, $D_o = D_e$ in which case, $\alpha = 0$. The choice of satisfactory α value depends on the importance of conclusions that would be drawn based on it.

The alpha value for each characteristic was evaluated separately, based on the ‘scores’ assigned by Evaluators. Observed scores for each element/characteristic is summarized in Table C-2 to Table C-6 in Appendix C. Krippendorff’s alpha was evaluated for each element and has been summarized in Table 5-2.

Table 5-2: Summary of alpha values for five characteristics identified by four evaluators

Problem Characteristic	Krippendorff’s Alpha
Number of goals	1.0
Functional requirement	1.0
Non-Functional requirement	0.505
End user information	0.184
Reference to existing product	0.598

A high Krippendorff's alpha value of 1.0 for characteristic 'Number of goals' indicates that Evaluators concur to the definition of a goal, and were easily able to identify it from the problem statement. Analysis of the annotated problems also shows that each Evaluator was indeed able to identify the same phrases used for describing the problem goal. It should also be noted here that all problems used for this study had an explicitly stated objective, which probably made identification easier.

An alpha value of 1.0 for characteristic 'Number of functional requirements' insinuates the fact that Evaluators were able to comprehend the definition and use the guidelines to identify this characteristic in the problems. By examining their annotated work, it was evident that the Evaluators were able to relate functional requirements to action verbs, as well as to nouns derived from action verbs.

A moderate value of 0.505 for the characteristic 'Number of non-functional requirement (NFR)' indicates that there may be some ambiguity between Evaluators in identifying this element. Assessment of the annotated results show that Evaluators are able to identify the same set of NFRs in the design problem in most cases. Otherwise, the choices made are consistent and correlated well amongst all four evaluators. The low agreement may also have resulted due to:

- a. Lack of Evaluator training: A rigorous training session was not used prior to this study and the effectiveness of the training protocol was not proven through pre – post assessment of Evaluators. The differences between Evaluators could have resulted due to the subjective views that people have about NFRs.

- b. Robustness of protocol: The protocol used for element identification may not have been robust enough to enable evaluators to identify elements consistently. Further validation of the protocol used might be helpful in improving reliability for this element.

A lower reliability coefficient of 0.184 for the characteristic 'End user information' indicates discrepancy amongst Evaluators as far as identification of this characteristic was concerned. Since this was a categorical choice, a low value clearly indicates that some Evaluators failed to identify references being made to the end customer or users. This could be improved by increasing the robustness of the protocol used for this approach. A low value for this coefficient is also due to the nature of measurement scale. Since this characteristic is evaluated using a binary (yes/no) scale, the chances of expected disagreement by chance are less (D_e in equation (3)) since only two outcomes are possible. This in turn reduces the value the reliability estimate α .

Finally, a moderate alpha value of 0.598 indicates good cohesion amongst Evaluators as far as the 'Reference to existing product' characteristic is concerned. This was also a categorical choice, where Evaluators had to assess whether a reference exists in the problem statement or not.

5.1.4. Discussion for approach 1: similarity assessment by identification of structural elements in design problems.

The first approach to compare problems is based on identifying the five structural elements in a design problem. The method requires a count or presence/absence of elements in the problem statement based on which problem size can be estimated. This count of elements can then be used for comparing two design problems with each other to

determine the difference in number of elements contained. This approach allows comparison based on problem model. A protocol is developed for the comparison process between problems using this method. In order to verify this method, an inter-rater reliability study is conducted with four human Evaluators who were assigned the task of identifying the five structural elements from a set of four design problems. Evaluation of reliability coefficient for each characteristic indicated a strong correlation between the characteristics: functional requirements and goals of the problem. A 'moderate' correlation was observed for characteristics: non – functional requirement and reference to existing product. In particular, Evaluators differed to a certain extent in identifying the non – functional requirements. This could be associated with the inherent subjectivity with which people view non – functional requirement. A low reliability was observed for the characteristic 'end user information' which was attributed to the implicit statements used for describing the end user in some of the design problems used in this study. The results of the protocol study give preliminary indications that people agree relatively well as far as identification of these elements are concerned.

This method based on identifying structural elements can serve as a starting guide for researchers, who want to identify and compare the size of design problems they want to use in their experiments. For instance, if a researcher wants to use a problem which should be similar to an existing problem, he/she can compare the size of information of the two problems based on the number of elements contained or the presence or absence of elements. This would benefit the research community since a systematic process for justification for problem selection is an opportunity that exists. The limitations of this

method include the need for a robust protocol for identifying the five elements with minimal difference between people. Another limitation of this approach was that the method did not capture the knowledge content of problems, and was based only on the number of structural elements identified. The complexity of the problem in terms of how easy or difficult the solution process is not accounted for by this method. In other words, a problem which seeks the design of a pen may have the same number of elements as another problem seeking the design of a hovercraft. An inter-rater reliability test conducted to verify the protocol insinuates that people are able to identify four characteristics with reasonable agreement between them. However, the protocol for this approach needs to be improved further to enable people to identify all five elements with higher reliability.

5.2. Approach 2: Similarity assessment through semantic analysis of design problems

The second approach for comparing problems is based on linguistic analysis of problem representation. Semantic analysis is the process of relating syntax structures from the level of phrases, clauses, sentences and paragraphs to the level of writing as a whole. Semantic analysis requires syntactic parsing of some kind. Semantic similarity refers to the likeness of meaning or semantic content of a set of terms, phrases, sentences or documents. A wide variety of similarity metrics are used to estimate the strength of semantic relationship between different language instances [116]. The aim of these measures is to assess the similarity of these language instances based on their meaning (semantic) rather than their syntactic similarity. For instance, the words tea and coffee are semantically similar while toffee and coffee are not [116]. Semantically similar words or phrases are used in the same way and context [117].

Semantic similarity can be evaluated at different levels. Measures used to evaluate sentence level similarity are of particular importance here since one of the objectives involve similarity assessment of design problems based on their representation. The second approach proposed to meet this objective is to assess the semantic similarity of different problem statements collected. The idea behind this approach is to assess whether design problem representations can be compared to each other based on their semantic similarity. For example, a problem statement '*design a pen*' is expected to be semantically similar to the problem '*design a pencil*' but dissimilar to the problem '*design a space shuttle*'. Thus, it seems interesting to assess if comparison between problem statements can be made based on their semantic content and whether or not the different semantic measures are capable of differentiating between problem statements.

5.2.1. Some measures for sentence level semantic similarity assessment

A pair of sentences is considered to be semantically similar if they have similar meaning or are used in the same context [118]. A few important techniques used to evaluate sentence similarity include:

- a. Word overlap measure: The intuition behind this approach is that if two sentences are semantically similar if they have more words in common. For a query sentence A and a reference sentence B, the similarity value is represented as

$$S(A, B) = \frac{n(A \cap B)}{n(A)} \quad (4)$$

This method relies on counting the number of terms which are common between two sentences to calculate similarity score. However, it does not differentiate between the

‘grammar’ of words used. Also, synonyms are considered as different words during similarity evaluation.

- b. TF – IDF method: An abbreviation for *Term Frequency – Inverse Document Frequency*, is used to reflect how important a word is to a document in a collection or corpus [119]. It is a product of term frequency and inverse document frequency. Term frequency is defined as the number of times a term occurs in sentence being evaluated. The inverse document frequency is a measure of how common or rare the term is across all documents. It is logarithmically scaled inverse ratio of total number of documents and number of documents containing the word [116,118]. In its general form, TF – IDF score is given by

$$(tf, idf) = f_{t,d} \cdot \log\left(1 + \frac{N}{n_t}\right) \quad (5)$$

where $f_{t,d}$ represents the term frequency, N is the total number of documents and n_t is the inverse document frequency.

- c. Jaccard similarity coefficient: This coefficient measures the similarity between two finite sample sets and is defined as ratio of size of intersection divided by the size of the union of the two sets. Mathematically, it can be represented as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (6)$$

This coefficient is similar to the word overlap method, the only difference being that the denominator contains a union of both sets. However, like the word overlap method, its estimate is based on the number of common words between the two sentences and

cannot differentiate between the ‘*grammar*’ of words used. Two sentences which contain a high number of common prepositions are likely to be regarded as similar by this approach.

- d. Latent Semantic Analysis: Latent Semantic Analysis (LSA) relies on applying statistical computations to a large corpus of text in order to extract and represent the contextual usage meaning of words, sentences and documents [93]. LSA enables representation and comparison of meaning of words and passages from analysis of text alone. The underlying assumption for this method is that words and phrases that have similar meanings will be used in similar pieces of texts. Similar words or phrases are represented by values close to one, while values close to zero or less indicate dissimilarity. LSA applies singular value decomposition (SVD) which is a form of factor analysis for analyzing texts. For example, comparing two sentences ‘*Design a pen*’ and ‘*Design a pencil*’ using LSA yields a similarity score of 0.81 indicating high semantic similarity between the two statements. When ‘*Design a pen*’ is compared with ‘*Design a car*’, similarity score obtained is 0.20 indicating low semantic similarity between the two statements.

Amongst the several alternatives available for semantic similarity analysis, Latent Semantic Analysis is a suitable choice for the task at hand for two reasons:

- a. LSA compares statements based on their contextual meaning derived from a large corpus of text. This provides a more robust measure of similarity as compared to other methods, which rely on the collection of problem statements used as input. Even TF –

IDF does not take into account words which are semantically similar and treats them as different words thereby reducing similarity scores.

- b. LSA has been shown to be more efficient than other methods especially when long statements and large set of texts are compared to each other [117,120].

5.2.2. The LSA approach to problem similarity evaluation

For evaluating design problem similarity using LSA, all 50 problem statements shown in Table A-1 are compared against each other. In order to evaluate LSA similarity scores, the online LSA tool² was used [121]. This online interface contains precomputed semantic spaces and tools to manipulate those spaces. One such option is 'Matrix comparison', which enables the similarity comparison of all design problems with each other, and generates a square matrix with similarity scores. All design problem statements are fed as input text in the LSA toolbox for matrix comparison. The choice of corpus is selected as 'General reading up to 1st year college' with 300 factors. This corpus has the closest relevance for the area under study from the available options. The number of factors is the number of dimensions of the reduced sparse matrix that should be kept for analysis. LSA scores have been shown to vary with number of factors chosen and often, better similarity estimates are obtained when high order factors are used [122]. This however, depends on the size of text to be analyzed and computational power. Also, the kind of comparison between different design problems is chosen as a 'paragraph-paragraph' type comparison, which applies a weighing function to enhance retrieval results [123]. LSA

² Available at <http://lsa.colorado.edu/>

generates scores on the scale of -1 to 1, with -1 being completely dissimilar instances while 1 representing complete similarity (or equivalence). The similarity score matrix for design problem statements obtained from LSA is shown in Table C-7.

5.2.3. Discussion for LSA results (approach 2)

A snippet of the LSA results has been included in Table 5-3.

Table 5-3: A snippet of similarity scores. Top row and first column represent examples of problem statements used for analysis

Problem statement to Problem statement comparison	Design an urban bi or tri cycle for use by white collar workers	Design a concrete mixer which can operate using a bicycle pedal mechanism.	Design a device which can compost waste vegetables.	Design a reading device for old people which can read the newspaper for them.	Design a water lifting device.
Propose alternative solution to coal pile problem at thermal plant, since the plant may not have enough land nearby to store the coal on ground.	0.02	0.01	0.06	0.01	0.03
Design of a next generation alarm clock which ensures easy operations like change of time and alarm stop unlike conventional clocks.	0.09	0.26	0.15	0.13	0.12
Design a litter collection device for volunteers.	0.23	0.48	0.6	0.25	0.39
Redesign an electric toothbrush for increased portability.	0.01	0	-0.01	-0.01	0.03
Design alternative means to manually propel boats which are easy to maneuver, don't rock the boat or splash water.	0.09	0.15	0.15	0.06	0.52

Observation of similarity scores reveals that semantic similarity assessment can identify problem statements which resemble each other contextually to a certain extent. For instance, the similarity score for the problem statements 'Design a device which can compost waste vegetables' and 'Design a litter collection device for volunteers' was found

to be 0.6. Using subjective interpretation, it can be argued that since both problems are related to waste or trash in some respect, their contextual meanings in the text corpus used for analysis could have indicated a similarity.

Similarly, a moderately high similarity score is observed for problems ‘Design a water lifting device’ and ‘Design alternative means to manually propel boats’ which is possibly because both problems, in context are somewhat related to water. On the contrary, the semantic similarity between the problem texts ‘Redesign an electric toothbrush for increased portability’ and ‘Design a water lifting device’ is as low as 0.03, which indicates a very low contextual similarity between the design problems. A negative similarity score of -0.01 for problems ‘Redesign an electric toothbrush for increased portability’ and ‘Design a device which can compost vegetables’ indicates that these problems are contextually dissimilar.

5.2.4. Graphical representation of semantically similar problems

LSA provides a matrix of similarity scores between different design problem statements which were compared against each other. This matrix can therefore be used for generating a network graph where semantically similar design problems are connected to each other. In other words, problem 1 and problem 2 contain an edge between them if they are semantically similar based on LSA results. Likewise, problem 1 and problem 3 are disconnected if they are not similar semantically. However, a cut off score to determine contextual similarity of problem statements is not established yet and needs verification by applying the LSA process to a bigger set of problems. Here, network of design problems based on LSA similarity results is generated using the following procedure:

1. An adjacency matrix is generated using the LSA similarity matrix shown in Table C-7 of Appendix C by replacing all the negative LSA scores by zero which signifies that these sets of problems (same as the ones used in chapter four earlier) are disconnected.
2. Using this matrix, a network is generated where design problems are connected to each other if the adjacency matrix contained a value other than zero. For example, design problem 1 and 2 have an edge in between since their LSA similarity score is 0.14 while problem 2 and 27 are disconnected since their similarity score is 0. Network density is evaluated for the network thus generated.
3. Step 2 is repeated but this time, the adjacency matrix is generated by replacing all cells with LSA score less than 0.01 by zero to indicate disconnectedness. This results in a new network and network density is calculated again. This process is repeated till the network becomes completely disconnected and all design problems appear as separate nodes. This happens when the maximum LSA score (0.63) is replaced by zero for form an adjacency matrix.
4. Densities obtained for network in this series of steps are plotted against the LSA cut off score which was chosen for that network. For instance, a network density of 0.94 was obtained when problems with LSA score of less than zero were chosen as disconnected, 0.91 when problems with LSA score of less than 0.02 were chosen as disconnected and so on. The resulting plot is shown in Figure 5-4 which helps identify the LSA score after which the network density levels out.

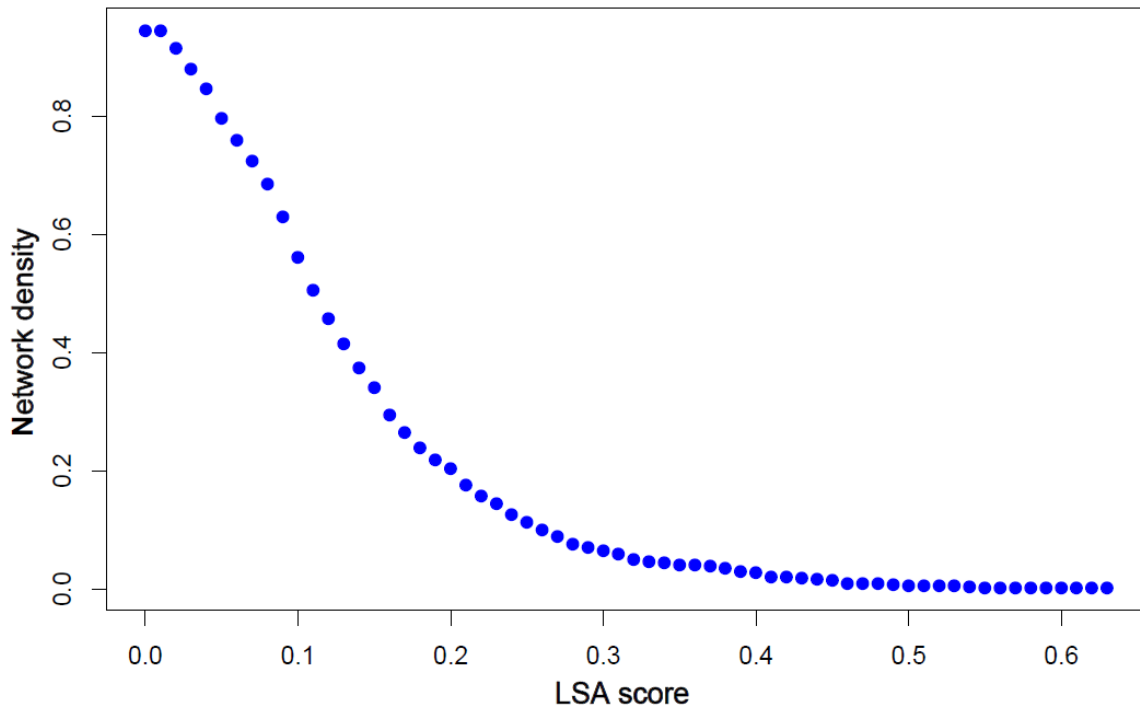


Figure 5-4: Variation of network density with choice of similarity cut off score

Network density attains a constant value after LSA score of 0.35. This plot gives preliminary information about LSA cutoff scores which can be used for gauging problem similarity based on this approach. Figure 5-5 to Figure 5-8 show design problems which are connected to each other when their LSA similarity scores are more than 0.35, 0.40, 0.45 and 0.50 respectively. These networks help represent the connection between various problems based on their LSA scores.

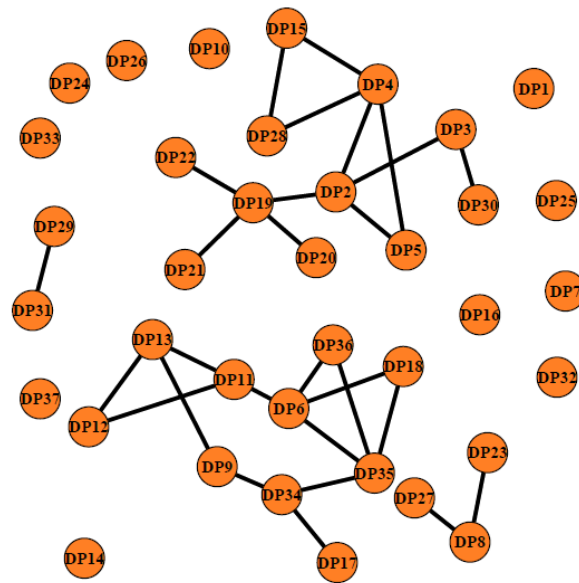


Figure 5-5: Design problems connected at cut off score of 0.35

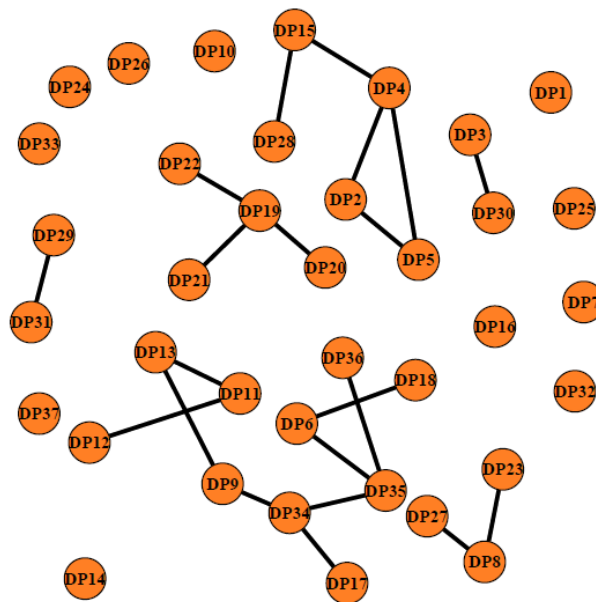


Figure 5-6: Design problems connected at cut off score of 0.40

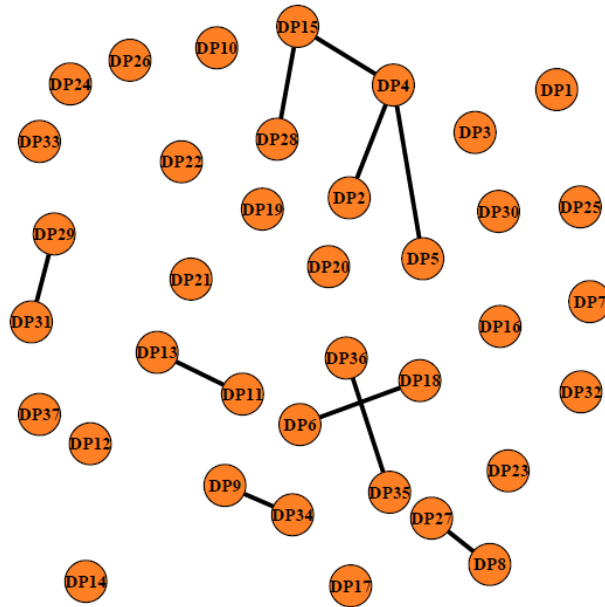


Figure 5-7: Design problems connected at cut off score of 0.45

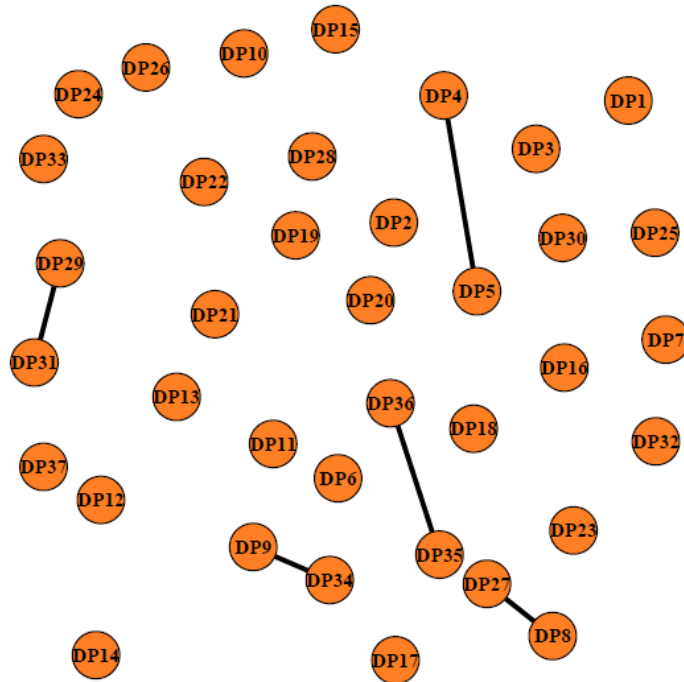


Figure 5-8: Design problems connected at cut off score of 0.50

5.2.5. Limitations of approach 2 (LSA)

LSA provides a convenient way of comparing the semantic similarity of design problems. However, in its present form, it does not use word order to derive syntactic relations or logic [93]. Particularly, in this case when design problem representations are analyzed, it was found that this method is unable to identify the ‘nuances’ and solvability of problem statements. LSA scores for the texts 'design a car to accelerate from 0 to 60 in 60 seconds' and 'design a car to accelerate from 0 to 60 in one minute' was found to be 0.86, which is high but still not equal to one. This however, is a contradiction with how an engineer perceives these two statements. For a human designer, both these problem statements are equivalent and same. Also, the results from LSA depend on the text corpus used for estimating the cosine products for similarities, which means that better and more relevant scores can be obtained from text corpuses which are more exhaustive and relevant to the subject under study. Here, general body reading up to first year of college is chosen as the text corpus because it is the most relevant option available. A corpus which is pertinent to the task at hand should include different articles and literature published in the field of design, so that better contextual similarity can be obtained. Also the similarity scores from LSA does not enable establishment of an objective cut – off score below which problems can be considered as dissimilar. Human interpretation of the results obtained from LSA is still required. Lastly, LSA comparison does not address the solvability of design problem or their requirements. The effort needed for a designer to solve a problem cannot be ascertained based on the similarity scores obtained from LSA.

5.3. Possible application of the two similarity assessment approaches in creativity experiments

The two approaches discussed for evaluating similarity of conceptual design problems can be used together to help researchers justify similarity (or dissimilarity) between two design problems used for a study. The application of these two methods is illustrated through two examples below.

Durand and coauthors [15] studied the impact of design problems on the four creativity metrics: quantity, quality, novelty and variety through two experiments using four design problems. The first experiment used ‘alarm clock’ and ‘corn shucking device’ as comparator problems (Problem 1 and Problem 2 in Figure 5-9) where the two groups were asked to use inspiration from nature during idea generation. Examination of results obtained from the experiment indicated no significant difference between the quantity and variety of ideas generated by the two groups for the two design problems assigned to them. However, the group assigned to the ‘Alarm clock’ problem generated higher quality ideas as compared to the group which worked on ‘Corn shucking device’ problem. On the contrary, the ‘Corn shucking device’ group generated solutions with higher mean novelty score as compared to group which worked on ‘Alarm clock’ problem. Table 5-4 shows the comparison between the two problems using structural elements and LSA. Although both problems contain the same number of functional requirements, the ‘*alarm clock*’ problem contains more number of non-functional requirements as compared to the ‘*corn shucking device*’ problem. This might have resulted in the ‘*alarm clock*’ problem generating higher quality solutions as compared to the ‘*corn shucking device*’ problem [109]. Additionally,

Problem 1

Alarm clocks are essential for college students, however often times they will wake up a roommate and those around them as well. Design an alarm clock for individual use that will not disturb others. The clock should be portable for use in a variety of situations such as on the bus, in the library, or in a classroom.

Customer Needs: Must wake up individual with no disturbance to others.

Must be portable and lightweight.

Electrical outlets are not available as a constant power source.

Low cost

Problem 2

Corn is currently the most widely grown crop in the Americas with the United States producing 40% of the world's harvest. However, only the loose corn kernels are used when bought canned or frozen in grocery stores. An ear of corn has a protective outer covering of leaves, known as the husk, and strands of corn silk threads run between the husk and the kernels. The removal of husk and silk to clean the corn is known as shucking corn. Design a device that quickly and cheaply shucks corn for mass production.

Customer Needs: Must remove husk and silk from corn cob with minimal damage to kernels.

A large quantity of corn must be shucked quickly.

Low cost

Figure 5-9: Alarm clock and corn shucking device problems used by Durand and coauthors

the ‘alarm clock’ problem contains a reference to an existing product (an alarm clock) which is more likely to be familiar with participants than the ‘corn shucking device’. This may have resulted in the difference in novelty scores between the two problems. However, at this point, this idea cannot be conclusively stated and needs further investigation. The semantic similarity score for these two problems from LSA was 0.30 indicating that the problems have low contextual similarity. Thus, the two problems may be regarded as different problems both structurally and semantically.

Table 5-4: Structural and LSA similarity for alarm clock and corn shucking device

Structural similarity			LSA similarity
	Alarm clock	Corn shucking	0.30
Number of goals	1	1	
Number of Functional requirements	1	1	
Number of non-functional requirements	4	3	
End user information (Yes=1/No=0)	1	0	
Reference to an existing product (Yes=1/No=0)	1	0	

In the second experiment, peanut shelling machine and blind cup device (Problem 3 and Problem 4 in Figure 5-10) were used as comparators to assess their impact on creativity scores. In the second experiment, participants were free to use any design method for generating solutions. The problem statements are shown in Figure 5-10. In the second experiment, the two problems differed from each other since they generated different quantity, quality and variety scores.

Problem 3

In places like Haiti and certain West African countries, peanuts are a significant crop. Most peanut farmers shell their peanuts by hand, an inefficient and labor-intensive process. The goal of this project is to design and build a low-cost, easy to manufacture peanut shelling machine that will increase the productivity of the African peanut farmers. The target throughput is approximately 50 kg (110 lbs) per hour.

Customer Needs: Must remove the shell with minimal damage to the peanuts.

Electrical outlets are not available as a power source.

A large quantity of peanuts must be quickly shelled.

Low cost.

Easy to manufacture

Problem 4

Design a volume-measuring apparatus for use while cooking by a person who is blind. It needs to be easy to operate and able to be used for both powders and liquids without splattering during operation. The apparatus needs to measure graduated quantities from 1/4 to 2 cups.

Customer Needs: Prevent waste of food products.

Easy to clean.

Low cost.

Figure 5-10: Peanut shelling device and blind cup problems used by Durand and coauthors

The two problems differ from each other in terms of the number of functional and non-functional requirements contained in them. The ‘peanut shelling machine’ problem contains one functional requirement (‘to shell peanut’) whereas the ‘*blind cup device*’ problem contains two functional requirements (‘to measure graduated quantities’ and ‘to be useful for powders and liquids with no splattering’). The ‘*blind cup device*’ problem generated higher quality solutions as compared to the ‘*peanut shelling machine*’ problem in spite of having less number of non-functional requirements. The semantic similarity score for these two problems obtained from LSA is 0.40 indicating some contextual

similarity between the problems. This may have been because both problems have a reference to food or an item which is related to a food in some respect (peanut). However, when both structural and semantic similarity are used in conjugation, the two design problems may not be considered as similar to each other.

Table 5-5: Structural and LSA similarity for peanut shelling machine and blind cup device

Structural similarity			LSA similarity
	Peanut shelling machine	Blind cup device	0.40
Number of goals	1	1	
Number of Functional requirements	1	2	
Number of non-functional requirements	5	3	
End user information (Yes=1/No=0)	1	1	
Reference to an existing product (Yes=1/No=0)	0	1	

In a recent work on functional modeling, Patel and coauthors [28] used two similar design problems to assess the impact of using partial function structures in arriving at a solution for assigned design problems. For this, Patel and coauthors used two different but similar design problems: the automatic cloth ironing device and automatic recycling machine (Problem 1 and Problem 2 in Figure 5-11). The number of requirements in the two problems, the type of solutions expected, number of words in both problems and sentence formation were used as justifications for the two problems to be similar. The results from the experiment also indicated that the two problems are similar since the mean number of functions added for each problem was not statistically different.

Problem 1

Design an automatic clothes-ironing machine for use in hotels. The purpose of the device is to press wrinkled clothes as obtained from clothes dryers and fold them suitably for the garment type. You are free to choose the degree of automation. At this stage of the project, there is no restriction on the types and quantity of resources consumed or emitted. However, an estimated 5 minutes per garment is desirable.

Problem 2

Design an automatic recycling machine for household use. The device should sort plastic bottles, glass containers, aluminum cans, and tin cans. The sorted materials should be compressed and stored in separate containers. The amount of resources consumed by the device and the amount of space occupied are not limited. However, an estimated 15 seconds of recycling time per item is desirable.

Figure 5-11: Automatic clothes ironing device and recycling device used by Patel and coauthors

Table 5-6 shows the comparison between the two problems based on structural elements identified and the semantic similarity between two problem statements. Both problems contain almost similar number of functional and non-functional requirements. Both problems contain information about who the end user is going to be. Semantically, the two problems indicate a high similarity of 0.52 probably because both problems require the design of an automatic device and contain time limit as a desirable operating characteristic. Thus, by using the two approaches together, similarity of the two problem statements used can be further justified.

Table 5-6: Structural and LSA similarity for automatic ironing device and recycling device

Structural similarity			LSA similarity
	Automatic clothes ironing	Automatic recycling machine	0.52
Number of goals	1	1	
Number of Functional requirements	2	3	
Number of non-functional requirements	1	1	
End user information (Yes=1/No=0)	1	1	
Reference to an existing product (Yes=1/No=0)	0	0	

5.4. Comments and recommendations

- The two approaches for similarity assessment of conceptual problems help in comparison between problems based on the problem statement. The effectiveness of both these methods can be enhanced by simultaneously using both methods to assess problem similarity.

Recommendation: LSA and element identification can be used in conjugation to compare both semantic and size similarity between design problems. This can be achieved by first evaluating the semantic similarity between design problems using LSA and then comparing their similarity based on size. Figure 5-12 shows the connection between the 50 problems used in this study based on their LSA and size similarity. The cutoff score chosen to consider problems to be similar for this graph was 0.35. The color of the node represents the size of problem. Here, the size of a design problem is equal to the number of functional and non-functional requirements contained in the problem statement. This graph shows how problems with different sizes are connected to each other semantically. For instance, DP6 is similar to DP 35 with respect to size and semantic content together. Thus, using both methods in conjugation can be a useful way of selecting or justifying problems which are similar.

- Solvability of problem cannot be addressed by either of the two methods. Problem solvability is partially related to the existence of a known solution for the problem which is difficult to ascertain for conceptual problems due to their abstract nature.

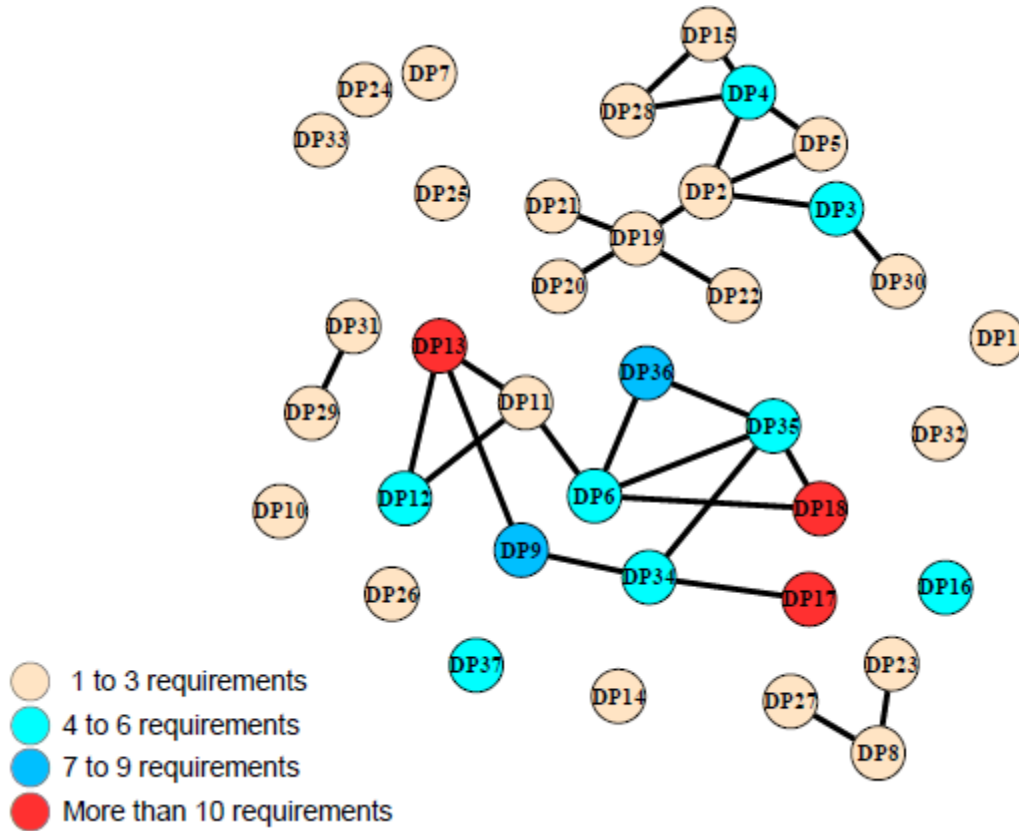


Figure 5-12: Design problems connected by size and LSA scores (>0.35)

Recommendation: Problem solvability is an important measure that should be assessed before selecting a problem for user study. Solvability of problems can be evaluated if a known benchmark solution to the problem by method prescribed by Thoe and Summers [38]. Requirements in problem statement can then be related to the design parameters in solution to determine the coupling between requirements. This can be used to estimate how difficult the problem is. This method will, however require researchers to use design problems which have an existing solution.

- LSA provides an objective evaluation of problem similarity based on problem statement but the results depend on the text corpus used for extracting the contextual

meaning of phrases. A corpus which is relevant to the area of design and creativity may provide better similarity comparison.

- The robust protocol for identifying the five elements in a design problem statement could be useful in improving the inter-rater agreement between people for the first comparison approach based on element identification.

Recommendations: The protocol used for element identification could be improved by repeating the inter-rater agreement study with an improved set of questions for each element. Since a low agreement was observed for element ‘reference to an existing product’, the questions asked for identifying this element in this study should be improved further.

CHAPTER SIX

META-ANALYTIC REVIEW OF CREATIVITY STUDIES

Objective: To understand whether design problems has a relationship with effectiveness of examples used as interventions in creativity experiments in engineering design.

Research task: Meta-regression analysis of user studies using examples as intervention.

6.1. Meta-regression analysis in creativity studies

Meta-analysis and meta-regression have been widely used in clinical trials and social sciences over the years. Studies in design creativity are, in many ways analogous to clinical trials in medical research [54]. While meta-analysis is used for systematically combining results from relevant studies to obtain an overall conclusion with greater statistical power as compared to individual studies, meta-regression is used for identifying potential covariates or moderators in experiments and their impact on effectiveness of method or intervention being studied [124]. For example, a meta-regression analysis conducted by Johnson and others in 1999 helped establish the importance of aspirin dose in risk of stroke [125] which had not been noticed in previous trials.

Use of examples before or during idea generation process has been shown to improve creativity of ideas generated in experiments [88,126–128]. However, other studies have shown that providing examples constraints the solution space explored to example-related domains, thereby reducing the creativity of solutions generated [85,129,130]. Thus,

it is difficult to ascertain whether presenting examples before or during idea generation tasks is helpful or not. One important difference between all such studies using examples as intervention is the design problem used. Since design problems have been shown to influence the results of creativity experiments [15] [26], understanding whether the design problem used in these studies had any influence on effectiveness of example intervention is important in trying to address the question of whether examples are useful or not. In order to address this, meta-regression can be used to ascertain if design problems moderated the effectiveness of interventions used in these studies (**RQ4**).

6.2. Basics of meta-analysis

Meta-analysis is a systematic method for statistically integrating the results of a set of related studies on a given topic [131]. It provides a standardized approach for examining existing literature on a particular topic to determine whether a common conclusion can be reached regarding the observed effects of a treatment from different studies [132]. Meta-analysis can be used for summarizing, integrating and interpreting selected sets of scholarly works [133]. Inferences drawn from most experimental studies are restricted by the small sample size used. By combining several studies together, meta-analysis aims to levy the resulting large sample size to draw better generalized conclusions than the individual studies comprising it. In short, meta-analysis aims at combining the results of different treatments or interventions published in separate studies to obtain an integrated result. To achieve this, studies fulfilling certain criteria are obtained and a treatment effect size is calculated for each study. The overall effect size for the collection of studies can then be obtained by combining the individual effect size estimates from different studies [134].

Meta-analysis applies only to empirical research studies which produce quantitative findings and report descriptive or inferential statistics to summarize the results [133]. It cannot be used for summarizing theoretical works which are qualitative in nature.

Treatment effect size, commonly known as effect size in social sciences, is a name used for family of indices that measure the magnitude of treatment effect or intervention. It carries information about either the direction or magnitude of quantitative research finding, or both. Different metrics are used for estimating the effect size of treatment (viz. odds ratio, risk ratios, mean difference, mean gain, proportion difference and others), depending on the type of measure reported and nature of relationship between reported variables [133]. The choice of effect size metric is based on three considerations:

- Their ability to measure (approximately) the same thing from different studies.
- The ability to compute them from reported information in studies.
- Its meaningfulness for the area under study.

For instance, if the summary data reported in primary studies included for meta – analysis are based on means and standard deviations in two groups, mean difference (standardized or unstandardized) or response ratio can be used for estimating effect size.

In addition to effect size, the precision of the effect size estimate is also needed. This is because effect size values resulting from studies using larger samples are more precise than those resulting from studies using smaller samples. That is, sampling error is smaller for effect sizes estimated from larger samples than from smaller samples [133]. Thus, every effect size value is not equivalent in terms of the amount of information it carries. This ‘difference in precision’ of effect sizes prohibits obtaining a summarized

result by calculating the arithmetic mean of these effect sizes. To resolve this, each study in a meta-analysis is assigned a weight based on its precision and a weighted summation is used for calculating the overall effect size estimate. These weights are obtained by calculating the standard error of the effect size and taking its inverse since higher standard error means lower precision [133].

6.3. Models for effect size estimation

Two statistical models have been developed for drawing inferences about effect size from a collection of data namely fixed effect and random effect model. The fixed effect model is used for making inferences about effect size parameters in the studies under observation while the random effect model is aimed at making inferences about distribution of effect size parameters in a population of studies from a random sample of studies [135]. Fixed effect models are appropriate for making conditional inferences which apply to the collection of studies included but not to studies to be conducted in the future or the ones not included in analysis. Random effect models are used for making unconditional inferences which embodies generalized conclusions beyond the studies included in analysis. The statistical procedure associated with these two models are different and the choice of which model to use depends on the existent differences between studies under observation. For instance, if the studies included for meta-analysis contain random variations in settings, participant characteristics or nature of treatment methods employed, a random model should be assumed since the variation in effect size distribution can be attributed to ‘special’ differences between these studies, apart from the natural sampling variations [135].

6.3.1. Fixed effect model

The simplest fixed effect model is based on the assumption that the variation in effect size estimation arises due to ‘chance’ variations associated with subject – level sampling error present in the included studies [133]. The average effect size is estimated by combining the effect size estimates across all studies in the sample. True effects size in a study is the effect size of underlying population and is the size which would result when an infinitely large sample size is used. Observed effect size is the effect size which is actually observed in studies [97]. If θ_i and T_i denotes the true effect size and observed effect size in the i^{th} study respectively, then the true effect size estimate can be modeled as

$$T_i = \theta_i + \varepsilon_i \quad (7)$$

where $\varepsilon_i \sim N(0, \sigma_i^2)$ represents the sampling error and σ_i^2 is the variance associated with the i^{th} study included in analysis. The true effect size parameter θ_i is determined by a mean effect size β_0 , whose weighted estimate is given as

$$\hat{\beta}_0 = \frac{\sum_{i=1}^k w_i T_i}{\sum_{i=1}^k w_i} \quad (8)$$

where $w_i = 1/\sigma_i$ and k is the number of studies included in the analysis. The sampling variance for $\hat{\beta}_0$ is the reciprocal of sum of weights for the k studies involved and given by

$$\sigma_*^2 = \left(\sum_{i=1}^k w_i \right)^{-1} \quad (9)$$

The confidence interval for β_0 using this model can be obtained as

$$\hat{\beta}_0 - t_{\alpha/2} \sigma_* \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2} \sigma_* \quad (10)$$

which can be used to determine whether the observed effect size is significant or not.

6.3.2. Random effect model

For analyzing effect sizes from different studies which are systematically different from each other, the assumption of a common true effect size for all studies might be impractical. The random effect model assumes that each observed effect size differs from the true effect size of the population by a subject level sampling error (within-study variation due to difference in participant characteristics) and another source of variability resulting from random variations between studies. Here, the variance associated with each effect size has two components: one associated with natural sampling error (same as the fixed effect model) and a second component associated with random variance between studies. In other words, this model assumes that there is a greater residual variation in effect sizes than would be present due to natural sampling variations alone. The observed effect size for each study can be represented as

$$T_i = \beta_0 + \varepsilon_i + \eta_i \quad (11)$$

where $\eta_i \sim N(0, \tau^2)$, τ^2 represents the degree to which true treatment effects vary between studies as well as the extent to which individual studies give biased assessment of treatment effects. Other terms have same meaning as discussed in equation(7). Here, the true effect size, θ_i is not fixed and the variance of T_i incorporates the variance of θ_i as well. The sampling variance of T_i , assuming that the sampling error ε_i and random error η_i are independent can be represented as

$$\sigma_{total}^2 = \sigma_i^2 + \hat{\tau}^2 \quad (12)$$

where $\hat{\tau}^2$ is an estimate of the between – study variance component. By replacing σ_i by σ_{total} in equation(8), the weights associated with individual studies can be evaluated and the estimate of true effect size can be assessed using equations (8) to (10) thereafter.

The difficulty associated with a random effect model is obtaining a good estimate of the random variation, $\hat{\tau}^2$. Two different methodologies are used for this, one based on method of moments and the other based on maximum likelihood which uses an iterative scheme [136]. The methods based on moments is adequate for most purposes and is easier to implement but produce results which are less accurate as compared to iterative methods. Iterative methods can be used to get estimates of random variation without making assumptions about the distribution of random effects [137][138].

6.4. Heterogeneity in effect size

One of the assumptions of random effect model is that the true effect size is not ‘fixed’ or constant between studies, but varies from one study to another. When the dispersion of effect sizes around their mean is no greater than expected from within – study sampling errors, the distribution of effect sizes is said to be homogeneous [133]. In such a case, the observed effect sizes vary within some range of the common effect size due to within – study errors. A distribution is said to be heterogeneous when the variability of effect sizes is larger than what would be expected due to sampling errors alone [139]. Heterogeneity refers to the excess variation in observed effect sizes over that expected from the imprecision of results within studies [137]. Heterogeneity generally arises because

studies differ in their design and conduct, apart from differences in participants, interventions, conduct and outcome measures [138]. This diversity is known as methodological or clinical heterogeneity. Statistical heterogeneity exists when true effect is different between studies and can be detected if the between – study variation in effect size is greater than what would be expected by chance.

A statistical test which rejects the null hypothesis of homogeneous effect sizes is based on the Q statistic, which is distributed as a chi – square with $k-1$ degree of freedom where k is the number of effect sizes and is given as

$$Q = \sum_{i=1}^k w_i (ES_i - \overline{ES})^2 \quad (13)$$

where w_i is the weight associated with each study, ES_i is the individual effect size for each study and \overline{ES} is the weighted mean effect size for k effect sizes [133]. If Q exceeds the critical value for $k - 1$ degrees of freedom, null hypothesis of homogeneity of effect size is rejected. A statistically significant Q therefore indicates the presence of heterogeneity. It is also accepted now that test for heterogeneity have low power and hence a non – significant test for heterogeneity should not be taken as implying homogeneous effect size distribution [140,141].

Understanding the cause of heterogeneity increases the scientific value and relevance of results from meta-analysis. However, there are always competing explanations for heterogeneity which makes it difficult to identify the exact cause. Meta-regression is one approach which can be used to identify relationships between effect size and one or more characteristics of the study involved. Regression analyses are commonly

used for identifying relationship between independent and dependent variables in studies. Meta-regression is based on similar principles, except that the dependent variables is the effect size observed for different studies and the covariates (or predictors) are variables which differ between studies [142].

6.5. The meta-regression approach

Meta-regression is often used for identifying one or more ‘moderators’ or covariates which may influence the heterogeneity of effect sizes. Moderators are qualitative or quantitative variables that affect the direction and/or strength of the relation between an independent variable and a dependent variable [143]. In meta-regression, moderators are defined as covariates which are measurable characteristics in studies which may or may not explain the heterogeneity observed in effect sizes. For instance, in clinical trials, the dosage of drug administered may yield larger observed effects of treatment and may be a moderator which influences the effect size of different studies. A meta-regression model can be expressed as

$$\delta_{iR} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n x_{in} + \eta_i + \varepsilon_i \quad (14)$$

where δ_{iR} is the predicted effect size, x_{ij} is the j^{th} moderator variable for the i^{th} study, β 's are regression coefficients and $\eta_i \sim N(0, \tau^2)$, where τ^2 denotes the amount of residual heterogeneity which cannot be accounted for by the moderators in the regression model and $\varepsilon_i \sim N(0, \sigma_i^2)$ represents the within – study error [144]. In order to predict the regression coefficients, a modified weighted least square method is used by most statistical packages (SPSS, R, STATA and others), where each effect size is weighed by the inverse of its

variance. A significant regression coefficient for any moderator variable indicates that the variable has a contribution in the observed heterogeneity in effect sizes observed.

Two indices are used for determining the overall fit of the weighted meta-regression model: Q_M which denotes the heterogeneity due to regression model and Q_E which measures the variability unaccounted for by the regression model. These metrics represent the sum of squares for the regression model and residual sum of squares respectively [133]. Both Q_M and Q_E are distributed as chi – square. Q_M has n degree of freedom where n is the number of predictor variables in regression model. A significant value for Q_M indicates that at least one of the predictor variables in the regression model has a coefficient significantly different from zero. Q_E has $k - n - 1$ degrees of freedom, where k is the number of effect sizes used in the regression model. A significant Q_E means that variability beyond subject – level sampling still remains even after removing the variability due to moderating variables used in the regression model. In other words, Q_E is an estimate of heterogeneity in effect size which cannot be accounted for by the moderating variables and some other sources of variations exist between studies.

Residual heterogeneity which cannot be explained by covariates in regression model can exist and should be acknowledged in analysis. Hence, the proper approach for determining the weights of studies is to use a random effect model which estimates the weight of each study by calculating the inverse of sum of within – trial and residual between – trial variance [145]. However, estimation of residual between – trial variance is challenging and the estimates are usually imprecise because of the limited number of studies it is based on. Different methods have been developed for estimating the residual

between – study variance like empirical Baye’s estimate [146] or restricted maximum likelihood (REML) estimate [137] [134].

6.6. Steps for meta-regression analysis of user studies using examples as intervention

The following procedure is used for conducting a meta-regression for user studies published where examples have been used as an intervention:

Step 1: Collect published literature in the area of design creativity where examples for problem solution have been studied.

Step 2: Develop a coding procedure to extract relevant information for meta-regression from the collection.

Step 3: Estimate the effect size for each study.

Step 4: Estimate the variance in effect size for each study (or treatment).

Step 5: Generate regression model with design problem size as moderator.

Step 6: Analyze the significance of these regression model to determine whether the covariate has an impact on effect size seen in studies or not. Figure 6-1 shows the steps involved in meta-regression.

6.6.1. Step 1: Collecting studies for meta-regression

The first task required for a meta-regression analysis is to collect which are relevant to the study question. The objective of this study is to assess whether the design problem has any relationship with effectiveness of examples used as interventions. Out of the 34 studies collected in chapter four and five, nine contained examples as interventions. However, in

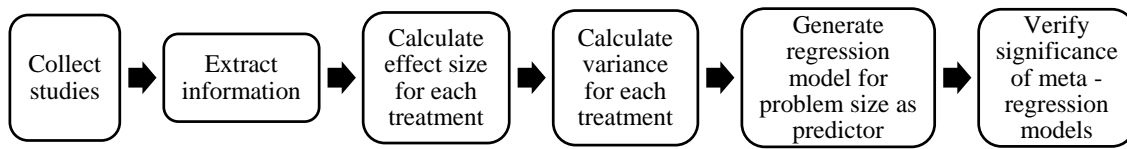


Figure 6-1: Steps for meta-regression analysis

order to get a good sample size for regression analysis, additional search is needed. Electronic databases including Google ScholarTM, Web of ScienceTM, Academic One file, EBSCohost, Journal of Engineering Design, Design studies journal and conference proceedings of ASME – DETC and ICED are used for retrieving additional relevant studies published which use presenting examples to participants as an intervention. The keywords used for search are “design” intersected with either *creativity*, *idea generation*, *concept generation* or *ideation*. The relevant studies should, therefore be related to design creativity which contain a design problem task and use presenting solution examples as an intervention. The relevant references and citations used in the search results are also sought. In total, 189 studies were found which matched the search criteria. The studies are further filtered using the following criteria:

1. The study includes a between – subjects experiment design, with a treatment/intervention group contrasted against a control group which receives no treatment.
2. The study uses a design task for testing the efficacy of an intervention in the form of examples and reports the problem statement used.

3. The study reports quantitative information which enables computation of effect size. Qualitative studies were subsequently eliminated from the collection.
4. The study measures one of the following aspects of creativity: quantity or fluency, quality, variety or novelty of resulting ideas.
5. Studies were limited to English language and focus was on studies published between the years 1995 to 2015.

Criterion 1 is used because it provides a baseline against which the treatment conditions can be compared for their performance. Criteria 2 is used because the objective of this analysis is to study the influence of design problems on effectiveness of examples presented to participants during experiment. For this, it is essential that the problem statement used are reported. Qualitative studies in design creativity are eliminated because methods used in meta-analysis cannot be applied for such studies. A wide variety of metrics have been used for measuring design creativity over the years [147,148]. Hence, in order to reduce this source of variation, only the studies which measure quantity, quality, novelty and variety of design solutions are considered for analysis [149]. Apart from this, studies which used fluency as a measure of creativity are also included since it is also based on similar principles as quantity metric proposed by Shah [149]. Through filtering, the number of studies which meet all the criteria is narrowed down to seventeen. However, not all studies report all four metrics for creativity assessment. Also, some studies use multiple experiments or multiple treatments, which allows for a reasonable number of sample point for regression analysis. In addition, attempts to collect unpublished data is not made due to

difficulties associated with retrieving them. The list of studies used for analysis and metric reported by them is shown in Table 6-1.

Table 6-1: Studies included for meta-regression and creativity metrics reported by them

Reference number	Metric reported			
	Quantity	Quality	Novelty	Variety
[80]	✓	✗	✗	✗
[130]	✓	✗	✓	✗
[82]	✓	✗	✓	✗
[83]	✓	✓	✗	✗
[84]	✓	✗	✓	✗
[150]	✓	✗	✓	✗
[151]	✓	✓	✓	✗
[152]	✓	✗	✓	✗
[153]	✓	✗	✗	✓
[86]	✓	✗	✓	✗
[154]	✓	✗	✗	✗
[155]	✓	✗	✓	✗
[156]	✓	✓	✗	✗
[157]	✓	✓	✓	✓
[158]	✓	✗	✓	✗
[87]	✗	✓	✓	✗
[88]	✗	✗	✓	✓

6.6.2. Step 2: Extracting information

The second step necessary for conducting a meta-regression analysis is extracting relevant information. For this, a coding procedure is developed which helped in gathering the relevant information needed for analysis. From the selected set of studies, the following information are extracted:

- Name of author/authors
- Year of publication
- Title of study

- d. Method of participant allocation to condition
- e. Participant characteristic
- f. Sample size, reported mean and standard deviation for both treatment and control group for the four metrics concerned.
- g. Design problem used and its problem statement.
- h. The result of t – test or F – test reported³.

Four different datasets result due to the fact that different values for means and standard deviations would be reported for each of the metric. Each intervention condition was considered as an independent study with one example and one control condition. For instance, if a study reports the use of two treatments T1 and T2 on treatment groups and compares it to a control group C, it was broken down into two separate studies: T1 vs control group C and T2 vs control group C⁴. Thus, a total of $N=45$ sample points are obtained for metric ‘quantity’, $N=13$ sample points for metric ‘quality’, $N=32$ sample points for metric ‘novelty’ and $N=6$ sample points for metric ‘variety’. Generally, a minimum of ten sample points is required for conducting a meta-regression analysis [159], hence a regression model for effect sizes reported for metric ‘variety’ is not used.

³ This was essential to allow effect size calculation when means and standard deviations were not reported in studies.

⁴ The complete dataset can be viewed at [Extracted dataset for Meta-regression](#)

6.6.3. Step 3: Effect size estimation

For each study which is included for meta-regression, the standardized mean difference (SMD) value is calculated to estimate the effect size for each measure (quantity, quality and novelty). This statistic is chosen because some studies report creativity outcomes on different scales (although they measure the same construct). For SMD, the standard deviation for studies is used for standardizing the mean differences to a common scale and is given as

$$SMD = \frac{\text{Difference in mean outcome between treatment and control group}}{\text{Pooled standard deviation}} = \frac{\bar{X}_{\text{treatment}} - \bar{X}_{\text{control}}}{s_p} \quad (15)$$

The pooled standard deviation, s_p is given as

$$s_p = \sqrt{\frac{(n_{G1} - 1)s_{G1}^2 + (n_{G2} - 1)s_{G2}^2}{(n_{G1} + n_{G2} - 2)}} \quad (16)$$

where n_{G1} and n_{G2} refer to the group sizes, s_{G1} and s_{G2} are the standard deviations for the two groups [133]. However, this effect size produces effect sizes which are biased upwards when small sample sizes are involved particularly when sample size of 20 or less are involved [160]. An unbiased estimate of effect size can be obtained using

$$SMD' = \left[1 - \frac{3}{4N - 9}\right] \times SMD \quad (17)$$

where N is the total sample size ($n_{G1} + n_{G2}$). The standard error for the unbiased SMD is given as

$$SE' = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1}n_{G2}} + \frac{(SMD')^2}{2(n_{G1} + n_{G2})}} \quad (18)$$

By convention, a positive sign is assigned to an effect size where the treatment group does better than the control group in terms of output measured. This sign convention has been kept consistent for all SMD (effect size) calculations in this study. Some studies only report the t-value or F-value resulting from significance test between two groups. The standardized effect size in such cases can be calculated from the algebraically equivalent formulas for SMD given as [133]

$$SMD = t \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad \text{and} \quad (19)$$

$$SMD = \sqrt{\frac{F(n_1 + n_2)}{n_1 n_2}} \quad (20)$$

6.6.4. Step 4: Estimation of variance in effect size

It has been assumed that the variance in effect size estimates among studies is not only due to sampling error but also due to difference in settings of each study (study protocol, choice of design problem, experiment time and others). As discussed earlier, a meta-regression model has two sources of variation (equation(14)): one due to within – study variation and the other resulting from residual heterogeneity which cannot be accounted for by the regression model. Before conducting a regression assessment to test the impact of these covariates on effect size, the component of within – study variance needs to be evaluated. This component can be calculated using equation (18) which gives the standard error (and hence variation) for each treatment and control condition. Thus by

using equations (15) through (20), the effect sizes and corresponding variance for each effect size can be estimated for each of the four elements of creativity metric. Table D-1 to Table D-3 in Appendix D shows the unbiased standard mean differences and unbiased within – study standard errors for all studies for each of the three metrics.

6.7. Meta-regression analysis with problem size as moderator

In order to understand whether the choice of design problem has any correlation with the effectiveness of examples in experiments, a meta-regression model is needed where some quantitative representation of the design problem is used as a predictor variable. Size of design problem is one way using which design problems can be quantified. As discussed in earlier sections, size of a design problem can be defined in different ways [37]. For this research, size of a design problem has been defined as the sum of number of functional (FR) and non – functional requirements (NFR) contained in the problem representation. Thus, problem size can be expressed as

$$\text{Problem size} = \text{No. of FRs} + \text{No. of NFRs} \quad (21)$$

This metric is chosen in accordance with the method discussed in earlier section (Chapter 5) where design problems are compared to each other based on the five structural elements of a design problem. Also, the number of requirements given to participants have been shown to influence creativity of ideas generated [109]. Worinkeng identified that the type of requirement and the number of requirements used in a design problem has different effects on the four creativity metrics [109]. For calculating problem size, the design problem statement for all studies included in the meta-regression analysis are collected as

a part of data abstraction process. The procedure for evaluating problem size is similar to the process of identifying structural elements as discussed in previous section on problem similarity assessment. Here, only the number of functional and non-functional requirements used in the problem statement are evaluated. The list of studies included in analysis, the design problems used by them, the number of functional requirements, number of non-functional requirements and problem size is shown in Table 6-2. A detailed list of problem statements can be found in Table D-4 in Appendix D.

Table 6-2: Design problems used in seventeen studies used for analysis and problem size

Reference number	Problem name	No. of FR	No. of NFR	Problem size
[80]	Peanut shelling machine	2	5	7
[130]	New toy	1	1	2
[130]	Alien creature	1	1	2
[82]	Alarm clock	1	1	2
[83]	Petroleum pumping unit	1	0	1
[84]	Future transportation	1	0	1
[150]	Hen's egg	1	0	1
[151]	Device to pick books	1	3	4
[152]	Device to harness human power	2	2	4
[153]	Automatic watering device	3	0	3
[86]	Future transportation	1	0	1
[154]	Bike rack	1	4	5
[154]	Coffee cup	4	2	6
[155]	Fabric display	3	1	4
[156]	Peanut shelling machine	2	5	7
[157]	Peanut shelling machine	2	5	7
[158]	Alternative clock	2	0	2
[87]	Device to harness human power	2	2	4
[88]	Device to immobilize joints	1	4	5

6.7.1. The meta-regression model

For assessing whether design problems moderate the effect size or not, a linear weighted regression model is used with standardized effect size as the outcome variable and design problem size as predictor variable (covariate). This regression model is represented as

$$\delta_i = \beta_0 + \beta_1 \times size + \varepsilon_i + \eta_i \quad (22)$$

where δ_i is the predicted value of standard mean difference, β_0 and β_1 are the coefficients of the regression equation, $\eta_i \sim N(0, \tau^2)$ represents the residual variance which cannot be explained by the covariate used in regression model and $\varepsilon_i \sim N(0, \sigma_i^2)$ represents the within – study variance which can obtained using the standard error values from Table D-1 to Table D-3 for each of the three metrics. However, to conduct a weighted regression, the weights associated with each effect size needs to be evaluated. The weight for an effect size under random effect model can be estimated by

$$w^* = \frac{1}{\sigma_i^2 + \hat{\tau}^2} \quad (23)$$

where the estimate of $\hat{\tau}$ can be obtained from one of the several estimation methods [137,161]. The method based on restricted maximum likelihood estimates (REML) is a commonly used method for estimating $\hat{\tau}^2$ through an iterative scheme [162]. The statistical

software ‘R’ is used for conducting this meta-regression analysis using REML as the method for estimating residual heterogeneity [163]. Three regression models are built, one

for each of the three creativity metric. Additionally, all three effect size datasets (quantity, quality and novelty) are also assessed for presence of heterogeneity without fitting the moderating variables so that a comparison between original heterogeneity and heterogeneity with regression model can be made. Meta-regression results for each of the three metrics are discussed in following sections.

6.7.2. Quantity

For the $n=45$ effect sizes obtained for the metric ‘quantity’ (Table D-1), the model shows a statistically significant coefficient for the problem size at 95% confidence level ($p<0.01$). This indicates that there is a linear relationship between effect size (SMD) and covariate, problem size. The results of the linear regression model with problem size as moderator is shown in Table 6-3.

Table 6-3: Coefficients of regression model for metric quantity

Variable	Coefficient	SE	95% Confidence Interval	P
DP size	-0.10	0.02	-0.15 to -0.05	<0.01
Constant	0.29	0.10	0.11 to 0.48	<0.01

A significant and negative coefficient for the problem size in the regression model indicates that with increasing problem size, the standard mean difference of the treatments reduces. In other words, larger problems may reduce the effectiveness of examples as instigators for idea generation. This reduction in standard mean difference with problem size is shown in Figure 6-2. With increasing problem size, the quantity of ideas generated by participants in treatment groups who were presented with examples is lower than that of control group participants. This might be a result of the larger size of design problem

that was used in these studies and not due to examples fixating the treatment group participants. Thus, a potential confounding between problem size and effectiveness of example intervention may exist which needs to be verified through experiments before making conclusions about whether examples are useful or not.

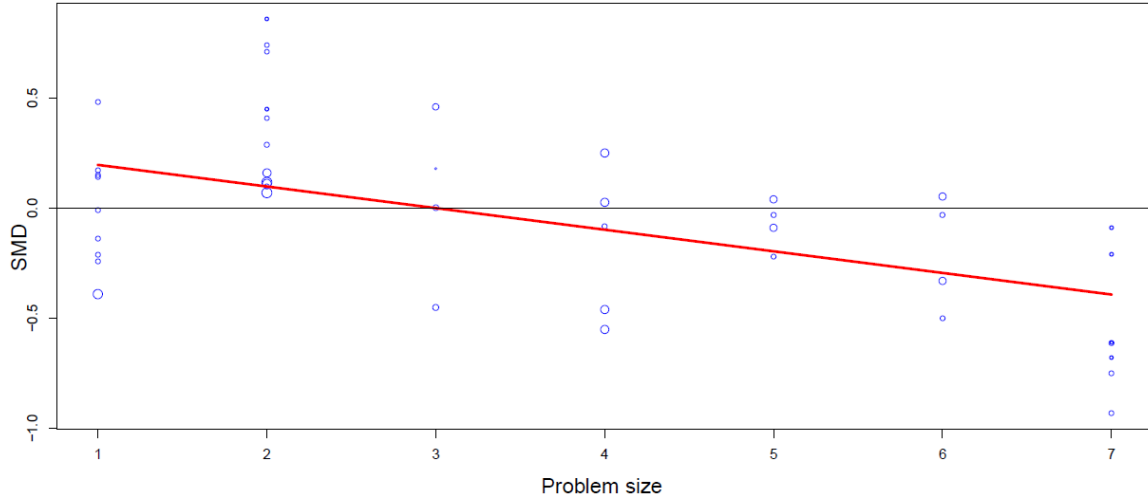


Figure 6-2: Standard mean difference for treatments versus design problem size for metric quantity. Plotted line shows the regression line while the size of each circle represents the weight assigned to effect size during regression analysis.

It is also essential to identify the amount of residual heterogeneity remaining in effect size distribution which could not be explained by the moderator used in the regression analysis. As discussed earlier, the τ^2 in a meta-regression model is an estimate of amount of residual heterogeneity which could not be explained by the moderator variable. Two other metrics which are frequently used to assess residual heterogeneity and heterogeneity accounted for by the regression model are the I^2 and R^2 . I^2 in meta-regression is defined as the ratio of residual variability to total unaccountable variability and is expressed as

$$I^2 = \left(\frac{Q_E - k + 1}{Q_E} \right) \times 100\% \quad (24)$$

where Q_E is the residual heterogeneity and k is the number of studies (effect sizes) involved. For meta-analysis model without moderators, I^2 is defined as the ratio of component of random variation to total variation and expressed as

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \sigma_e^2} \quad (25)$$

where $\hat{\tau}^2$ represents the random variance component. R^2 is defined as ratio of variance explained by the regression model to the total variance in effect size distribution and is given by

$$R^2 = \frac{\sigma_{\text{explained}}^2}{\sigma_{\text{total}}^2} = \frac{\tau_{\text{total}}^2 - \tau_{\text{residual}}^2}{\tau_{\text{total}}^2} \quad (26)$$

where τ_{total}^2 is the total random variance component without moderators. Table 6-4 and Table 6-5 show the estimated heterogeneity and other metrics for two models: meta-analysis without moderators and meta-regression analysis with design problem size as moderator for metric ‘quantity’.

Table 6-4: Summary of heterogeneity metrics for SMD distribution without moderator

τ_{total}	I^2	Q_{total}	p – value
0.21	32.95%	67.55 (dof = 44)	<0.01

Table 6-5: Summary of heterogeneity metrics for SMD distribution with moderator

τ_{residual}	I^2	R^2	Q_{residual}	p – value
0.13	16.05%	61.20%	49.21 (dof = 43)	0.24

A significant value of Q_{total} ($p < 0.01$) for the standard mean difference distribution for metric ‘quantity’ means that significant heterogeneity exists. R^2 value shows that the moderating variable ‘problem size’ could account for approximately 61% of the existing total random variance in the effect size distribution for creativity metric ‘quantity’. The remaining 39% residual variance could possibly be due to other variables that exist in these studies.

Key conclusions:

- With increasing problem size, participants who were presented with examples as interventions generate less number of ideas or solutions as compared to control group participants. Problem size, therefore may be influencing the effect of presenting examples.
- Meta-regression identified problem size as a potential moderator in studies using examples as intervention. This however, needs to be verified through experiments where problem size is varied and example intervention is used simultaneously.

6.7.3. *Quality*

For the $n=13$ effect sizes obtained for the metric ‘quality’ (Table D-2), the model fails to reject the null hypothesis of coefficient for the moderator variable being zero at 95% confidence level ($p=0.23$). This indicates that there is no linear relationship between effect size (SMD) and covariate, problem size for the metric ‘quality’. The results of the linear regression model with problem size as moderator is shown in Table 6-6.

Table 6-6: Coefficients of regression model for metric quality

Variable	Coefficient	SE	95% Confidence Interval	P
DP size	-0.09	0.07	-0.23 to 0.06	0.23
Constant	0.85	0.39	-0.08 to 1.62	0.03

A negative coefficient for design problem size indicates that the standard mean differences reduce with increasing problem size. However, since the coefficient of design problem size is non – significant for metric ‘quality’, enough evidence does not exist to say there is a linear relationship between problem size and effect size (SMD). In other words, whether problem size affected the relationship between presenting examples and quality of idea generated by treatment group participants cannot be established with available evidence. Figure 6-3 shows the regression line for predicted values for SMD with respect to design problem size.

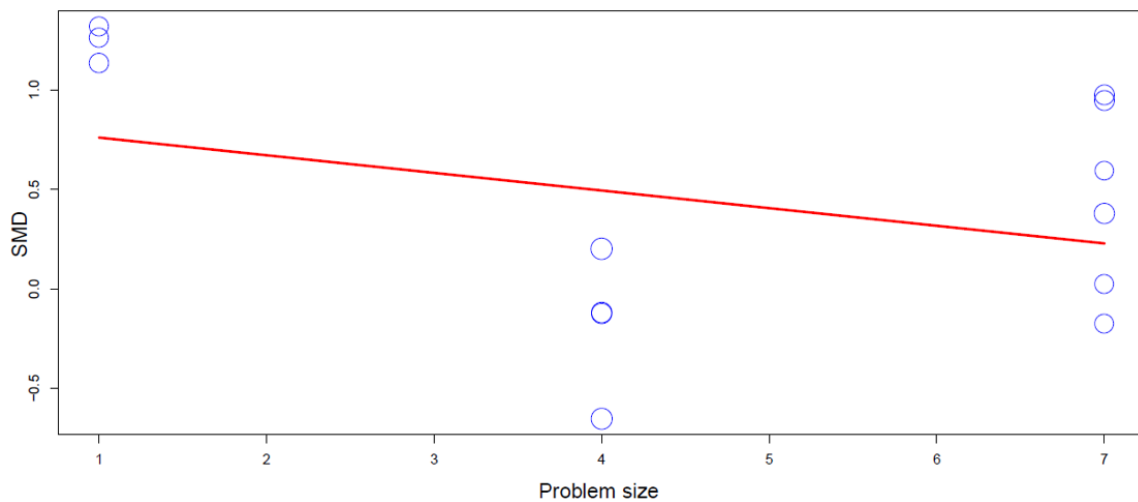


Figure 6-3: Standard mean difference for treatments versus design problem size for metric quality

Table 6-7 and Table 6-8 show the estimated heterogeneity and other metrics for two models: meta-analysis without moderators and meta-regression analysis with design problem size as moderator. A significant value of Q_{total} ($p < 0.01$) for the standard mean difference distribution for metric ‘quality’ means that significant heterogeneity exists. R^2 value shows that the moderating variable ‘problem size’ could account for approximately 3% of the existing total random variance in the effect size distribution for creativity metric ‘quality’. Also, a significant $Q_{residual}$ ($p < 0.01$) means that the existing heterogeneity in SMD distribution could not be accounted for by the moderator ‘problem size’ and that there are other sources of variations between studies which need to be explore further.

Table 6-7: Summary of heterogeneity metrics for SMD distribution without moderator

τ_{total}	I^2	Q_{total}	$p - value$
0.55	72.96%	44.72 (dof = 12)	<0.01

Table 6-8: Summary of heterogeneity metrics for SMD distribution with moderator

$\tau_{residual}$	I^2	R^2	$Q_{residual}$	$p - value$
0.54	72.49%	3.05%	41.05 (dof = 11)	<0.01

Key conclusion:

- Significant evidence does not exist to suggest that problem size affects the effectiveness of examples as interventions.
- There may be other variables which might be able to explain the residual heterogeneity in effect size distribution.

6.7.4. Novelty

For the $n=32$ effect sizes obtained for the metric ‘novelty’(Table D-3), the model fails to reject the null hypothesis of coefficient for the moderator variable being zero at 95% confidence level ($p=0.44$). This indicates that there is no linear relationship between effect size (SMD) and covariate, problem size for the metric ‘novelty’. The results of the linear regression model with problem size as moderator is shown in Table 6-9. Figure 6-4 shows the regression line for predicted values for SMD with respect to design problem size.

Table 6-9: Coefficients of regression model for metric novelty

Variable	Coefficient	SE	95% Confidence Interval	P
DP size	0.05	0.03	-0.08 to 0.17	0.44
Constant	-0.18	0.12	-0.61 to 0.26	0.43

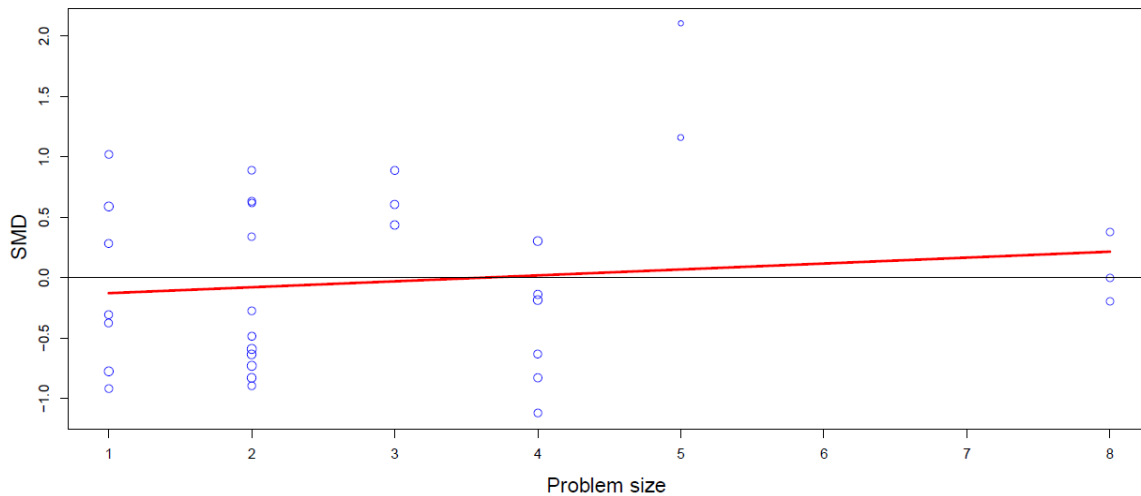


Figure 6-4: Standard mean difference for treatments versus design problem size for metric novelty

Table 6-10 and Table 6-11 show the estimated heterogeneity and other metrics for two models: meta-analysis without moderators and meta-regression analysis with design

problem size as moderator. A significant value of Q_{total} ($p < 0.01$) for the standard mean difference distribution for metric ‘novelty’ means that significant heterogeneity exists and the assumption of a random effect model is justified. R^2 value shows that the moderating variable ‘problem size’ could not account for any portion of the random variance in the effect size distribution for creativity metric ‘novelty’. Also, a significant $Q_{residual}$ ($p < 0.01$) means that all heterogeneity in SMD distribution could not be accounted for by the moderator ‘problem size’ and that there are other sources of variations between studies which need to be explored further.

Table 6-10: Summary of heterogeneity metrics for SMD distribution without moderator

τ_{total}	I^2	Q_{total}	$p - value$
0.62	81.25%	147.20 (dof = 33)	<0.01

Table 6-11: Summary of heterogeneity metrics for SMD distribution with moderator

$\tau_{residual}$	I^2	R^2	$Q_{residual}$	$p - value$
0.62	81.25%	0%	147.20 (dof = 40)	<0.01

Key conclusion:

Significant evidence does not exist to suggest a relationship between problem size and effectiveness of examples as intervention for novelty of ideas generated.

6.7.5. Verifying assumptions in regression models

Like other regression methods, meta-regression models also need to be verified for homoscedasticity and residual normality assumptions. Homoscedasticity assumptions can be verified by plotting standardized residual values against predicted values and have been

shown in Appendix D (Figure D-1 to Figure D-3). Visual assessment of these graphs do not show any discernible pattern in the distribution of residual values with predicted values of effect size for any of the four creativity metrics and the residuals seem to be more or less randomly distributed.

Normality of residuals for design problem regression model have been verified using normal q-q plot for all four creativity metrics. The results are shown in Figure D-4 to Figure D-6 in Appendix D. Most points in the q-q plots are scattered around a straight line and lie within the ideal confidence intervals (indicated by dotted lines in figures) which indicate that the normality assumptions for residuals of this regression model are not violated for quantity, quality and novelty metrics. However, some deviation of quartiles from the straight line can be seen for metric ‘variety’. Still, none of the plotted points lie beyond the confidence interval lines so as to severely affect the normality assumptions for this regression model.

6.7.6. Discussion of results

Meta-regression analysis with design problem size as moderator indicates that problem size has different implications on the effectiveness of examples on the three creativity metrics quantity, quality and novelty. The following conclusions can be made from this analysis:

- Larger problem size may reverse the effectiveness of examples on ‘quantity’ aspect from positive to negative. Using larger problems may cause fixation to be induced in participants in treatment group due to which they start generating less number of ideas.

- Enough evidence did not exist to show a relationship between problem size and effectiveness of examples for ‘quality’ and ‘novelty’ aspect. Problem size, therefore may not be an influencing factor in determining whether participants exposed to examples generate better ‘quality’ or ‘novel’ solutions or not.
- The regression results for effect size distribution for ‘quality’ and ‘novelty’ showed significant residual variance which could not be accounted for by problem size. This indicates that there are other variables which could be affecting the heterogeneity in effect size distribution which need to be investigated further.

6.8. Influence of requirement type on example interventions

The preceding section analyzed the effect of problem size of effectiveness of examples as interventions in creativity studies. However, the two components of problem size, functional and non-functional requirements have been shown to have different influences on results of creativity studies [26]. Thus, it is felt that analyzing the impact that these two requirement types have on effectiveness of examples would be useful in determining whether both components of design problem affects the results of example interventions or not. In order to analyze this, meta-regression models between effect sizes observed for the three creativity metrics and number of functional (FR) and non-functional requirements (NFR) present in the design problem used is studied separately. Thus, two regression models are used for each creativity metric:

$$\delta_i = \beta_0 + \beta_1 \times FR + \varepsilon_i + \eta_i \quad (27)$$

$$\delta_i = \beta_0 + \beta_1 \times NFR + \varepsilon_i + \eta_i \quad (28)$$

where symbols have the same meaning as discussed in earlier sections.

6.8.1. Quantity

The results for the two regression models with number of functional and non-functional requirements as moderators is shown in Figure 6-5 and Figure 6-6 respectively. Both functional and non-functional requirements have a significant regression coefficient (shown in Table 6-12) indicating the presence of linear relationship between effect size and the two variables. Number of non-functional requirements, however can account for 49% of the existing random variation in effect size distribution whereas number of functional requirement can account for 24%. This indicates that the difference in number of non-functional requirements in different problems used in these studies has a higher impact of heterogeneity present in effect size distribution as compared to functional requirements.

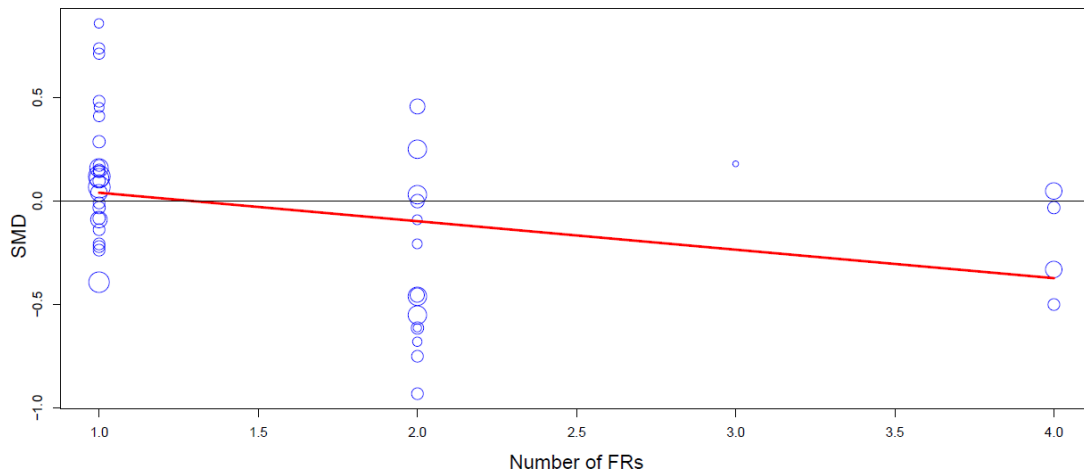


Figure 6-5: Standard mean difference for treatments versus number of FRs for metric quantity

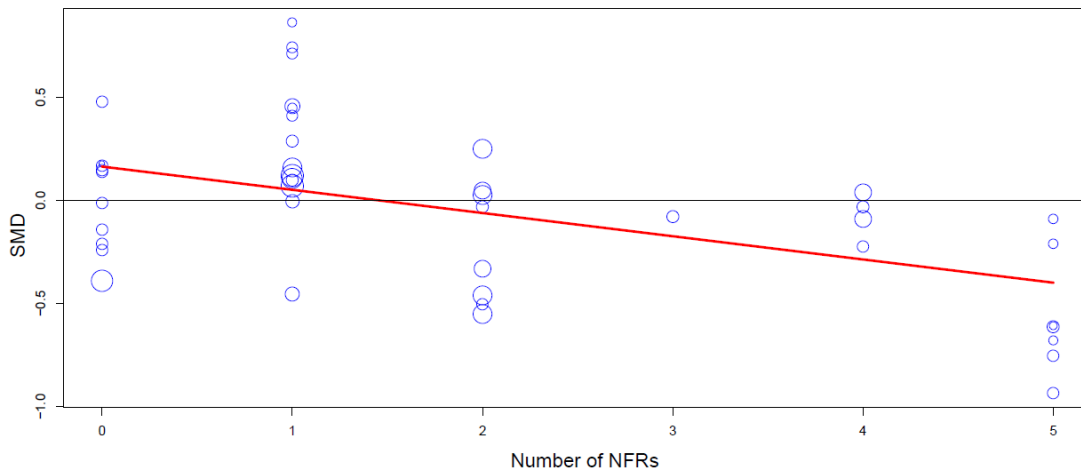


Figure 6-6: Standard mean difference for treatments versus number of NFRs for metric quantity

Table 6-12: Summary of results for regression for metric quantity

	FR	NFR
Regression coefficient	-0.14*	-0.11*
R²	24%	49%
Residual τ	0.18	0.15

*p<0.05

Key conclusion:

Both functional and non-functional requirements may influence effectiveness of example interventions. However, this claim needs to be verified through experiments.

6.8.2. Quality

The results for the two regression models with number of functional and non-functional requirements as moderators is shown in Figure 6-7 and Figure 6-8 respectively. Both functional and non-functional requirements have a non-significant regression

coefficient (shown in Table 6-13) indicating that a linear relationship between these two variables and effect size may not exist. Also, the amount of heterogeneity which could be explained by the two models was 1% and 2% respectively for number of functional and non-functional requirements respectively, which indicates that there are other variables which might be the reason for existing heterogeneity in effect size distribution in data for metric quality.

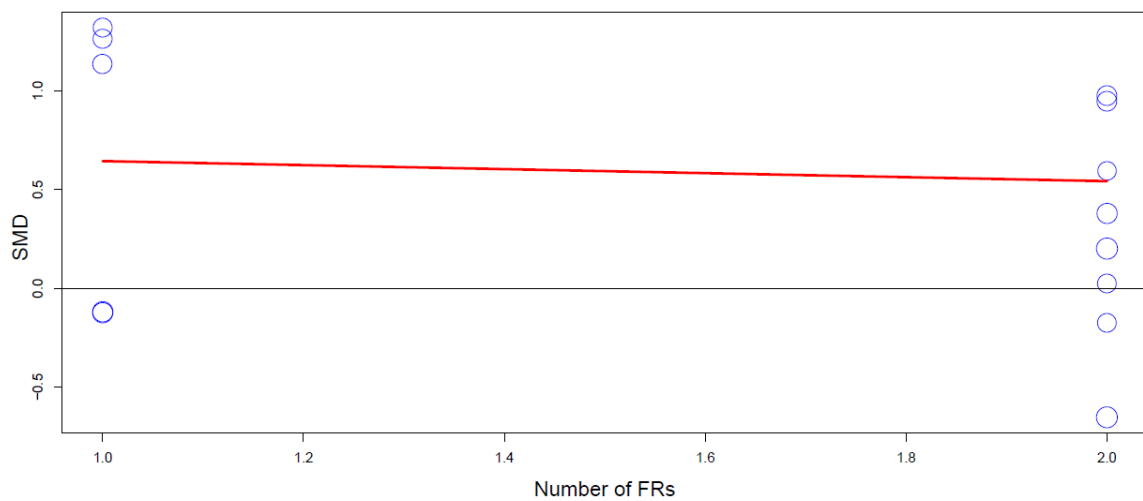


Figure 6-7: Standard mean difference for treatments versus number of FRs for metric quality

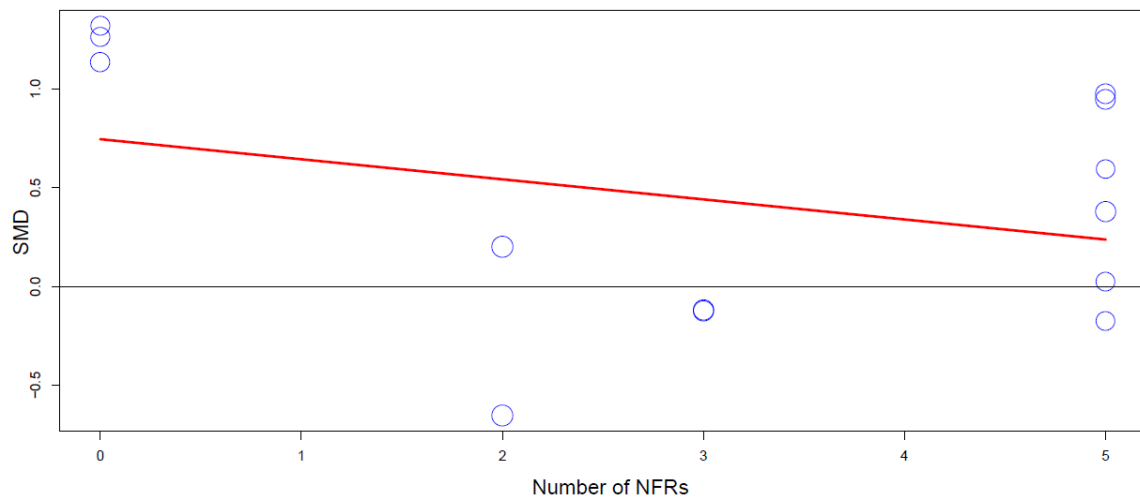


Figure 6-8: Standard mean difference for treatments versus number of NFRs for metric quality

Table 6-13: Summary of results for regression for metric quality

	FR	NFR
Regression coefficient	-0.04	-0.10
R²	1%	2%
Residual τ	0.55	0.54

Key conclusion:

Significant evidence does not exist to suggest a relationship between number of functional and non-functional requirements and effectiveness of examples as interventions.

6.8.3. Novelty

The results for the two regression models with number of functional and non-functional requirements as moderators is shown in Figure 6-9 and Figure 6-10 respectively. Both functional and non-functional requirements have a non-significant regression coefficient (shown in Table 6-14) indicating that there is not enough evidence to suggest a relationship between number of functional and non-functional requirements to effectiveness of example interventions. Also, the amount of heterogeneity which could be explained by the two models was 10% and less than 1% respectively for number of functional and non-functional requirements respectively, which indicates that there are other variables which might be the reason for existing heterogeneity in effect size distribution in data for metric novelty.

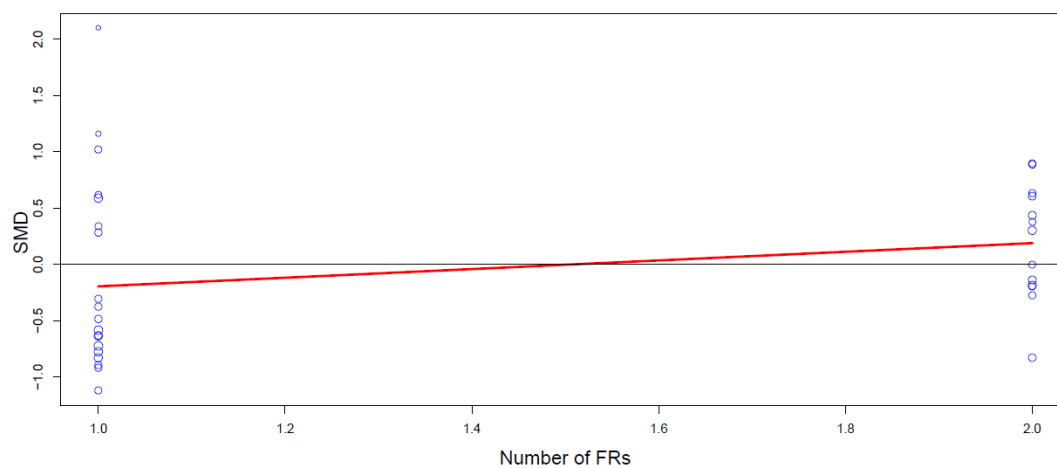


Figure 6-9: Standard mean difference for treatments versus number of FRs for metric novelty

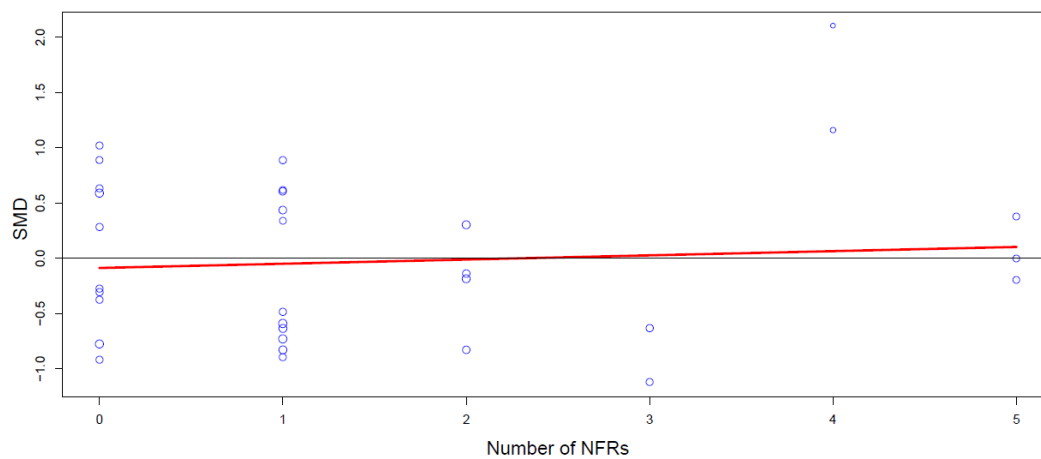


Figure 6-10: Standard mean difference for treatments versus number of NFRs for metric novelty

Table 6-14: Summary of regression results for metric novelty

	FR	NFR
Regression coefficient	0.38	0.04
R²	10%	<1%
Residual τ	0.58	0.63

Key conclusion:

Significant evidence does not exist to suggest a relationship between number of functional and non-functional requirements and effectiveness of examples as interventions for novelty of ideas generated by treatment groups.

6.8.4. Discussion of results

- Meta-regression models for both number of functional and non-functional requirements indicate that both variables might have an influence on the effectiveness of examples as interventions. The regression model seems to indicate a linear relationship with effect size for quantity of ideas generated in experiments using examples as intervention. Since the coefficient for regression model for both variables was negative, increasing the number of functional or non-functional requirement may possibly reverse the effect of examples from positive to negative (inspiration to fixation).
- Meta-regression models for the two variables for metrics ‘quality’ and ‘novelty’ did not provide enough evidence to suggest a relationship between effect size and the two predictor variables.

6.9. Comments and recommendations

- There are other sources of difference between studies which may be important covariates including participant characteristics and instructions provided during ideation task.

Recommendation: Meta-regression can be used to analyze the impact of some of these characteristics on results of interventions used in creativity. For instance, in order to investigate whether the type of participants used in experiments has an influence on the effectiveness of examples, a sub-group analysis with different participant types as predictor variable can be conducted to verify if this variable is a moderator or not.

- Studies published in creativity research often do not report all statistics related to experiment conducted. For instance, it is seen that studies do not report the mean and standard deviations for the results measured. At times, these results are published as bar charts and error bars which makes analysis difficult.

Recommendation: Guidelines for publishing studies in creativity can help reduce some of these variations. Reporting numeric values for means and standard deviations can reduce the risk of erroneous interpretation. This also helps during verification process if the experiment is reproduced to check its validity.

CHAPTER SEVEN

CONCLUSIONS AND RECOMMENDATIONS

Design problems are an important component of experiments in engineering design creativity. Experiment's requirements and researcher's objectives are two important factors which determine the nature of design problems used in experiments. As a result, different problems have been used in experiments which has resulted in problems being a source of difference between studies. The overall objective of this research is to address the need for using similar conceptual design problems in experiments in engineering design creativity. This objective is accomplished by addressing three sub-objectives i) to identify the pattern of design problem usage, ii) to enable comparison between two conceptual design problems based on their natural language representations and iii) to analyze the impact of design problems on effectiveness of example interventions used in user studies in engineering design creativity. For identifying the pattern of design problem usage, a graph between researchers and design problems is used. For comparing conceptual problems, two methods have been proposed and evaluated. The first method is based on identifying structural elements in a design problem and the second method is based on semantic similarity assessment of design problem statements. For analyzing the impact of design problem on effectiveness of examples as interventions, a meta-regression analysis is used.

7.1. Pattern of design problem usage (RQ1)

The first research question aims to identify the pattern of design problem usage in experiments in engineering design creativity. For this, two graphs based representation of

design problem usage has been used. The first graph shows the connectivity between researchers and design problems used by them. The second graph shows the connection between different experiments in design creativity published and the design problems used in them. The following conclusions can be made from this analysis:

- Design problem reuse is limited in engineering design creativity research possibly due to the difference in requirement of experiments. Therefore, the responsibility of justifying the appropriateness of design problem selected lies with the researcher. However, when a new problem is used in an experiment, it is difficult to prove its merit until the result from experiment is obtained.
- There are some design problems which have been used multiple times in different experiments and have been shared by researchers who have worked together on those experiments. Such problems may be useful as benchmark design problems since they have been tested and experimental evidence can be used to verify their effectiveness.

Recommendations

The following recommendations would be useful for enhancing problem reuse in engineering design creativity experiments:

- An opportunity exists for developing a repository of benchmarked design problems which can be shared and used by researchers according to their requirements. This will help researchers in selecting problems which have been used and tested while also reducing the differences between studies due to design problems.

- An opportunity also exists to develop methods to compare problems for similarity so that in case an existing problem cannot be reused, an alternative but similar problem may be used by researcher.

Contribution to overall research objective

Graph based analysis of design problems usage helped in highlighting the limited reuse of problems in creativity experiments. While this may be due to difference in study requirements, design problems are one source of difference between studies. Since problems are amongst the few variables which can be controlled by the researcher, it might be beneficial if this source of difference can be minimized. One way to achieve this could be to use existing problems which have similar characteristics with what is desired in a design problem for experiment. This would be beneficial in two ways:

- Reusing existing problems can help minimize one source of difference between studies
- Researchers can search and use problems which have been tested in experiments to minimize the need for developing a new design problem for experiments.

7.2. Design problem similarity assessment (RQ2 and RQ3)

The second and third research questions aim to identify methods to enable similarity comparison between design problems based on their natural language representation. For this, two methods have been proposed and evaluated. The first method is based on identifying the structural elements in a design problem statement. The second method is based on semantic similarity of design problem statements. The following conclusions can be made for these two research questions:

- The first method based on identifying the five characteristics in a design problem statement provides an element-wise comparison between two problems. The five elements include goals of a problem, functional requirements, non – functional requirements, information about end user and reference to an existing product. Comparison can be made based on the number of goals, functional and non-functional requirements or presence or absence of end user and reference to an existing product once these elements have been identified in the two problem statements.
- The second method for similarity assessment is based on computational technique known as Latent Semantic Analysis. This method provides a similarity score between the contextual meaning of two problem statements which can be useful for comparing problems.
- Both these methods may be used in conjugation with each other to address the requirement of comparing conceptual problems. LSA can be used to compare semantic similarity between problems while number of requirements can be used to compare structural similarity between problems. Thus, using the two methods together helps compare problems simultaneously for semantic and structural similarity.

Recommendations

The following recommendations may be useful for obtaining a similarity score for method based on structural element identification:

- Using information vectors generated by method based on identification of structural elements to obtain a structural similarity score between the two problems. A combination of both Euclidian distance between the two information vectors and

angular separation between them may be useful in comparing the two design problem for similarity. Cosine similarity scores alone may not be robust enough to correctly identify similar problems when the vector dimensions are reduced. Hence a combination of cosine similarity and Euclidean distance could provide a better estimate of similarity of the two information vectors.

Contribution to overall research objective

Enabling similarity comparison between conceptual design problems is useful primarily for two reasons. First, researchers need to use problems which are similar for different experimental requirements such as in pre-post test experiments. Since students are common subjects for experiments in engineering design creativity, researchers sometimes need to use different but similar problems in their experiments to prevent participant pool from being ‘tainted’ while ensuring that an appropriate problem is used in the experiment. Second, for the idea of design problem reuse to propagate, a repository of benchmarked design problems may be needed. This may require comparison between different benchmarked problems so that in case a problem cannot be used, a different but similar benchmark problem can be selected from the repository. The two methods for problem comparison may be seen as starting points for research in this area. By enabling researchers to compare problems structurally and semantically, the two methods can act as starting guides for researchers who need to use similar problems in their experiments.

7.3. Influence of design problems on effectiveness of example intervention (RQ4)

The fourth research question is aimed at understanding whether the choice of design problem used in user studies has a relationship with effectiveness of design method being

tested in the experiment. Specifically, relationship between design problem size defined as number of requirements contained in it and use of examples as interventions is assessed using meta-regression. The following conclusions can be made from this analysis:

- Design problem size may influence the effectiveness of examples as interventions as far as quantity of ideas generated is concerned. Using higher problem size in experiments may reverse the effect of example from positive to negative on the quantity of ideas generated by treatment groups. This hypothesis, however should be verified through experiments where examples are used as interventions and design problem size is used as an independent variable.
- Enough evidence did not exist to indicate a relationship between problem size and effectiveness of examples on quality and novelty of ideas generated by treatment group participants.

7.3.1. Influence of functional and non-functional requirements on effectiveness of examples

Additional meta-regression models were built to study the influence of functional and non-functional requirements separately on effectiveness of example interventions. The following conclusions can be made from this analysis:

- Both functional and non-functional requirements may have an influence on effectiveness of examples used as interventions for the quantity of ideas generated by treatment group participants. In other words, increasing the number of functional and non-functional requirements may reverse the positive effect of example intervention

into a negative one where participants in treatment groups may start generating less number of ideas as compared to control group.

- Enough evidence did not exist to identify a relationship between number of functional and non-functional requirements and effectiveness of example intervention on quality and novelty of ideas generated by treatment group participants.

Recommendations

Based on the results from meta-regression, the following recommendations are proposed:

- Significant residual variation was observed in most meta-regression models analyzed using problem size or number of functional and non-functional requirements. Other variables which might be a factor in explaining the residual variance need to be investigated further.
- In order to study the confounding effect produced by different number of functional and non-functional requirements on effectiveness of example interventions, an experiment with examples as interventions and number of functional and non-functional requirements as independent variables should be conducted. This would help verify the hypothesis that the number of functional or non-functional requirements influence effectiveness of example interventions in experiments.

7.3.2. Contribution to overall research objective

Meta-regression helped identify the influence that design problems may have on the conclusions that are drawn from results of user studies. For instance, the studies where

presenting examples produced improvement in quantity of ideas generated by treatment group participants may conclude that examples inspire participants to generate more ideas. Studies which report a reduction in quantity of ideas generated may conclude that examples cause fixation in participants. However, an important difference between the two studies is the design problem used. Regression results indicate that problem size may moderate the effectiveness of examples where using a higher problem might change the effect of example from inspiring to fixating. Thus, design problems may have the potential of convoluting research findings. One way to eliminate the concern could be to use same design problems across all studies. For instance, if all studies which analyzed the impact of examples on solution creativity used the same problem and reported different effects of example treatment, one can eliminate the choice of problem as a cause for observed difference in conclusions. In this case, other variables can be analyzed to identify the cause for disparity in observations for same treatment. Using the same design problem may not be feasible always. An alternative could be to use similar problems so that potential confounding due to choice of design problem can be reduced. Although methods to ensure problem similarity need to address other aspects such as solvability and complexity, the idea of using similar problems in experiments can help in reducing one source of variation which can possibly influence results of experiments.

7.4. Overall research conclusion

The overall conclusions that can be made from the research tasks accomplished in this research are as follows:

- Using problems which are similar to each other across creativity studies can help reduce one source of between-study variation. Although methods to assess similarity of conceptual design problems may not be capable of addressing all aspects for determining problem similarity, efforts in this area would be beneficial for the research community as a whole. Using similar problems can also help reduce the potential moderating effect that design problems may have on results of user studies and help develop more robust conclusions.
- Other sources of difference exist between studies whose influence on results of experiments may be difficult to identify. Therefore, a validation and verification framework for evaluating the results from experiments might be useful in development of a sound knowledge foundation upon which future work can be built.

7.5. Limitation of methods for comparing design problems

- Design problem solvability cannot be addressed by the two approaches proposed. Solvability of a problem is associated with the effort needed for solving the problem. This can be estimated by analyzing the coupling between problem requirements and design parameters present in an existing solution.
- The method based on structural element identification helps compare two problems element-wise but does not generate a similarity score between the two problems. This could be addressed by using vector representation of information contained in design problems and evaluating the cosine similarity and Euclidean distance between two vectors simultaneously. The challenge with this approach is in combining the two measures obtained.

- The robustness of protocol used for identification of structural elements in design problems should be improved. The current protocol used produced low inter-rater agreement for characteristic ‘reference to existing product’. Improving the protocol and testing its effectiveness can improve the reliability of the method.
- Although LSA enables to evaluate contextual similarity of design problems, obtaining a cut-off score above which problems can be considered as similar needs to be addressed.

7.6. Limitations of meta-regression results

Regression results from this study indicate the fact that design problems are an important covariate in user studies in design creativity and should be carefully selected during research design. Yet, the results presented here have some limitations which prevents their generalization. These limitations are discussed in following sections.

7.6.1. Publication bias

Publication bias germinates when research that appears is published literature is unrepresentative of the population of all conducted studies. Studies which have statistically significant findings are more likely to be published as compared to studies which did not observe a statistically significant effect. Publication bias has long been recognized as a problem in scientific research, and in meta – analyses in particular since the conclusions drawn from the meta – analysis depends on the results reported in studies included in the analysis. According to Rosenthal [164], for any research area, the extreme view believes that journals are filled with 5% of the studies which show Type I errors while “file drawers”

are filled with 95% of the studies which show no significant results. Publication bias is a potential threat to all research fields including quantitative studies and qualitative studies. This becomes more important when conducting meta – analyses, since it aims at combining the results of different studies to understand the common effect across different studies. As such, studies reporting only ‘positive’ results for treatments can inflate the conclusions which are drawn in a meta – analysis.

Assessment of publication bias is essential before drawing any generalized conclusions from a meta-regression analysis as well. One way of visually depicting the presence or absence of bias is by using funnel plot [133]. A funnel plot is a graphical representation in which the size of the study is plotted on the y – axis and the measure of effect size is plotted on the x – axis. The idea behind this approach is that studies with larger sample size would reduce sampling error and provide a more precise estimate of true treatment effect. In absence of bias, studies with smaller sample size should be spread around the base of the funnel and also, symmetrically distributed. In recent times, standard error of effect size estimates are increasingly been used for representing the y – axis of funnel plots [165]. In this case, if bias exists because smaller studies with non – significant results remain unpublished, the funnel plot would have an asymmetric appearance with a gap at the bottom corner of the graph. Such a representation provides a generic means of examining bias and other reasons may be associated with the asymmetry of funnel plot [165]. Since visual assessment of funnel plot asymmetry can be subjective, statistical tests have been developed to test the presence of asymmetry in funnel plots [166,167]. Egger [168] proposed a regression model where the effect size estimate (standardized mean

difference in this case) is used as a dependent variable with its precision (or standard error) as the independent variable. The assumptions behind this method is that in absence of any publication bias, the slope of the regression model is expected to be 0.

Funnel plots for standard errors versus effect size for all four creativity metrics are shown in Figure D-7 to Figure D-9 in Appendix D. Table D-5 also shows the results of regression test for asymmetry of the funnel plots. Visual analysis of the funnel plots indicate that most effect sizes observed are spread in the central region of the funnel with only a few studies populating the lower regions of the funnel. This indicates that studies with low sample size might be missing from this analysis since standard errors are expected to be higher for such studies. Results from regression test for asymmetry show that publication bias exists for the studies included in the analysis of metric ‘variety’ ($p < 0.01$). This can also be seen in the Figure D-9, the funnel plot for metric novelty which shows the absence of smaller studies on the left of the center line of funnel. Apart from this, asymmetry test for other metrics fails to reject the null hypothesis of no publication bias involved. However, the presence of bias in published literature cannot be ignored and consequently, the results from this study should be considered as hypothesis generating rather than proofs of hypotheses.

7.6.2. False positive conclusions

The quality of meta-regression depends on the quality of studies included in analysis. Often, heterogeneity in effect size can be explained by many possible experiment or participant characteristics rather than the one suspected. For instance, heterogeneity in effect size distribution for creativity user studies can be explained by other factors like

nature of participants, their academic standing, their experience level and host of other factors. Regression tests for moderators are thus based on subjective assessments which suggest that a certain covariate may be important. In this study, design problem is chosen as potential moderators based on prior studies conducted to test its impact on creativity results. However, it is possible that other such unexplored covariates might as well be significant moderators in explaining the existing heterogeneity.

7.6.3. Generalizability of meta-regression results

The relationships observed in meta-regression analyses are descriptive associations across studies and cannot be generalized as such. Significance of a moderator variable depends on the effect sizes and results reported in studies. It is difficult to include all possible user studies in a meta-analytic review due to resource limitations, hence it becomes difficult to justify the generalizability of such findings. It is possible that some studies which were not included in this analysis may change a significant moderator in to a non-significant one. Hence, the results from meta-regression should be considered as new experiment hypotheses which need to be verified through experiments.

CHAPTER EIGHT

FUTURE WORK

8.1. Assessing similarity of design problems based on their representation

The importance of using same or ‘similar’ design problems has been the central theme of this research. For enabling similarity assessment of problems, two methods were discussed which can help a researcher to compare problems based on their representation. However, the two methods in their present form are unable to assess all aspects of similarity which might be essential for researchers. These two methods can be used as starting guides when a researcher is assessing the design problem which should be used. Answers to the following research questions could empower design researchers in the problem selection process:

- a. **Research question:** How can researchers be allowed to choose problems based on their similarity?

Proposed research task: For answering this research question, a common database needs to be developed which contains various design problems used in the past in creativity research along with different similarity measures which can be used to compare them. A hierarchical similarity assessment can be used for this. At the first level, similar problems can be presented to a researcher based on semantic similarity of problem representations. As a next step, comparison between design problems remaining after the first stage can be made based on the five elements discussed in this research. After these two stages, problems which are similar, semantically and element wise would remain. The next step should

include analysis of design problems for their complexity as suggested by Summers and Shah [37] can be used for similarity assessment. This multi – level process would ensure that adequate justification can be given for problem selection and that at least one source of difference between design studies can be eliminated. This would also ensure that problem similarity is evaluated for both, their representation and effort needed for solving. Figure 8-1 shows the multi – level process which can be used.



Figure 8-1: Four level hierarchical approach proposed for problem similarity assessment

b. Research question: How can benchmarked design problems be developed?

Proposed research task: An essential requirement for development of a coherent methodology for design research is to prescribe benchmarked design problems which can be readily used by researchers for user studies. To accomplish this, the first requirement would be to assess whether the problems are similar to each other or not using the four level approach discussed above. Problems which are found to be similar can then be tested in experiments to see whether the results generated are significantly different or not as suggested by Durand and coauthors [15]. After following this two-way process for assessing problem similarity, a list of problems can then be published which were found to be similar, both analytically and experimentally. All such problems can then be treated as benchmarks and used by researchers for experiments.

8.2. Understanding the role of design problems in ideation experiments

Meta-regression analysis conducted on user studies published in the past showed that design problems are an important moderator of treatment effects in user studies in creativity and their importance needs to be emphasized to the research community. Answers to the following research questions can help elucidate the importance of design problems in creativity studies:

- a. **Research question:** Does the size of design problem reduce the quantity and variety of ideas generated in design tasks?

Proposed research task: For addressing this research questions, experimental evidence which supports or refutes this claim is needed. For this, user studies could be used to test whether the number of functional and non – functional requirements in a problem affect the quantity and variety aspects of solutions. However, replication of such an experiment would help gather conclusive evidence to either accept or reject this hypothesis. Some work in this regard has been done in the past [109], the results for which need to be further verified.

- b. **Research question:** Can other measures of design problem similarity be used to verify the influence of problem selection on treatment effects of user studies?

Proposed research task: A meta-regression analysis can be used again to see whether the choice of design problem has any influence on the outcome of interventions in user studies. However, a different metric for quantifying design problem for use in regression analysis can be used to verify the results from this research. For instance, the metric defined by Summers and Shah [37] to evaluate problem complexity can be used as a predictor variable

in regression model to verify whether design problem complexity serves as a moderator or not. This task will serve as a verification mechanism to check the claims made in this research.

8.3. Understanding the impact of other sources of difference between studies

The focus of meta-regression analysis used in this research was to understand whether design problem used in experiment have any influence on the treatment effect observed for various interventions. Meta-regression analysis can also be used to uncover some of the other moderators which might have an impact on the effect size. Answers to the following question could further help in enhancing the understanding of design research:

a. Research question: What is the influence of ideation time on effect size?

Proposed research task: The influence of ideation time used for experiments on results of creativity studies can be analyzed using meta-regression. For this, experiment time used in different studies can be used as a predictor variable to analyze if it moderates the effect size or not.

b. Research question: Does the education level of participant have any impact on effect size distribution in creativity studies?

Proposed research task: A large proportion of studies used for meta-regression analysis in this research used undergraduate student at different education levels in their experiments. Participant characteristics and their impact on outcome of user studies is an important issue which needs to be understood. It would not be an overstatement to say that majority of the

user studies in future will also rely on using student participants from academic institutions for studies in design creativity given the current predilection of researchers. Hence, it seems pragmatic to understand the impact of different levels of education on the measured outcome of user studies. Meta-regression analysis can be used to answer this task where education level of a participant can be used as a predictor variable to verify if it is important as far as moderating the effect sizes are concerned. For this, a dummy coded categorical independent variable can be included in the linear regression model similar to how categorical variables are tested in regular regression analysis. Each education level can be assigned a different code. This can help highlight whether some or all education levels are significant in predicting the effect size distribution for creativity scores.

8.4. Opportunities in design tool validation

The concept of validation is central to the idea of creating a global space for design practitioners and researchers to converse. At present, most research being done is in isolation and the research process is ambiguous. The process of validation can add a systematic approach to design research. It will act like a funnel, to channelize all new research towards at least one common ideology. This process, in no way intends to curb the individual inclinations and approaches of the researchers, but only gives them a framework to replicate and verify their ideas. With validation being accepted as a step in research process, it will bring the widespread design community together, since everyone will have at least one common point of discussion in their research. In the long run, a congenial environment can be developed, to propagate knowledge sharing and growth in this field.

Constructive criticism is an innate quality for nurturing research in any field. With a framework in place, it becomes easier to evaluate and verify the hypothesis put forward by others, and then to critique on their work. In an ideal scenario, such conversations between two researchers will significantly enhance knowledge transmission. Validation of design tools and methods provides an opportunity to attach a certain degree of scientific rigor to the design process. At present, the field is divergent, where different communities work on different beliefs and notions. As a result, several ideas and methods have evolved, which are neither established nor universally accepted or deployed. It is wise to adopt a diverse approach when working with engineering design, yet a more systematic and streamlined process will benefit the ultimate benefactors – industry and student community in the long run. At present, it is quite enigmatic to select a tool or process, since the process is ambiguous and the research community is in disharmony. Validation is one step towards the final aim of creating a scientific design process.

a. Research question: How can a collaborative framework for validation and verification of design tools/methods be established?

Proposed research task: A model framework for validating existing and new design tools and methods has been shown in Figure 8-2. This proposed framework is intended to serve as a starting guide for future research in this direction. Moreover, with time, more features and improvements can be included, to further enhance and encourage the validation process.

Since this framework is intended to cater to the needs and requirements of the design fraternity, it is essential to consider the requirements and needs of this group in

particular. Generating and reviewing the requirements for the framework should be a continuous process, to ensure usefulness of the framework for all users. The structure for the framework proposed here is based on an initial requirement list, which requires a framework which:

- a. Allows unrestricted access to design researchers, as and when needed.
- b. Allows a convenient way to implement various stages of the validation process, without any violation of the experimental requirements.
- c. Contains a wide range of experiments in its database, for testing several design tools and methods.

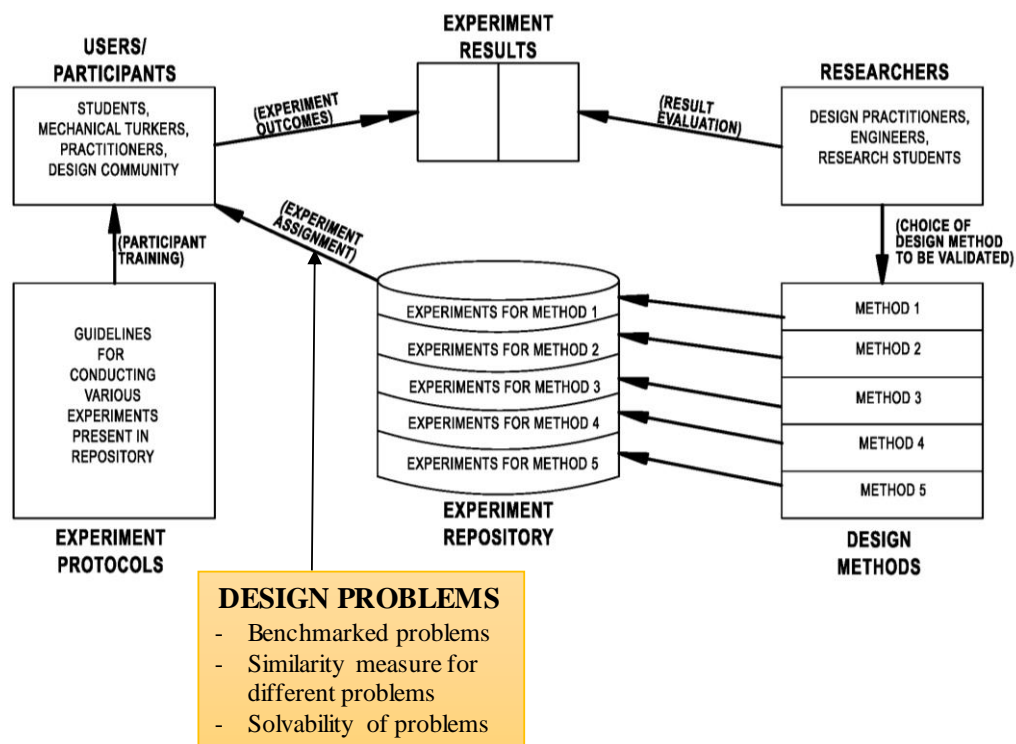


Figure 8-2: Proposed validation framework for design tools and methods

This framework should comprise of six primary elements:

- a. **Experiment repository:** To build on the works already done, it is essential to have a common repository of design experiments conducted over the years by various researchers and experimenters to test their hypothesis and tools. The repository would contain a pool of experiments which need to be verified and tested again through collaborative research.
- b. **Experiment protocol:** Every experiment has a set of rules, and procedure designed in order to render it useful for the study. Similar to an experiment repository, it is essential to keep a database for storing various protocols and procedures which need to be kept in mind if these experiments have to be replicated. Such a database, shall contain the guidelines and procedures to be followed for each experiment, which will help the researcher in educating the participants, and its smooth execution. This database shall also contain the hypothesis and performance metrics related to the different experiments stored in the repository, to help researchers recreate the experiments when desired.

Design problems: Design problems form an essential part of the protocol database. Benchmarked design problems can be stored in this repository with details of the study in which they were used, their size and solvability and possibly the average time required to solve them. Problems can be benchmarked by testing different problems for their impact on results of creativity experiments. This would enable a researcher to select a design problem based on the requirements of validation protocol and also reduce differences in design problem selection between studies. Similarly, the

similarity scores for different design problems based on its size (number of requirements and goals), contextual similarity and solvability can be stored in the database. This will enable researchers to compare and select appropriate design problems for validation studies. For example, if a researcher wants to test whether the robustness of the method is influenced by the choice of design problems, he/she can select different design problems based on their similarity and solvability scores. In this way, the researcher can ensure that the problems chosen by him were indeed different thereby reducing the threat to construct validity.

- c. **Design methods/tools:** A list of different design tools and methods that have or need to be tested shall be contained in this database. It can contain old, tested methods as well as new methods awaiting verification. Again as a startup, tools and methods related to creativity, function based design and decision support will be stored. Further additions, can of course be done as and when required. This database will enable the researcher to select the method to be validated, corresponding to which a list of experiments can be selected and thereby selected for trial.
- d. **Design researchers:** Members of the design community, with access to the framework need to be a part of the framework. It is paramount that a closed community is formed to monitor, verify and validate the experiments. All design researchers, research students and industrial practitioners will be a part of this, and will have access to experiments in the framework. This group will be like a team, whose job will be to select methods or tool to be validated, select the appropriate experiment from the repository and administer it to the participants. This group will also examine the results

obtained from the experiments conducted and evaluate it against an expected result, to ascertain the validity of the tool being used. Also, training and educating the participants about the experiment guidelines and process will be overseen by this group. Thus, researchers will get a chance to verify their new methods, as well as validate the existing ones.

- e. **Participants/users:** This group involves subjects, who will be a part of the experiment and validation process. This is the most diverse and large group in the framework, and requires extra efforts in order to ensure engagement in the main framework. Participants will involve students at various universities, design practitioners, mechanical turkers and others who wish to be a part of this framework. The diversity of this group is a key variable in this model. As with most design experiments, it is essential to determine the caliber and competency of these participants, since this will play a vital role in the results obtained from experiments conducted. A sound method for this is yet to be found, but preliminary tests like IQ, quantitative analysis and others may be useful in initial sorting. Additionally, background details for each participant needs to be stored in records, for the experimenter's reference before conducting any experiment. Thus, participation selection will be a major challenge for validation process, which needs to be done with utmost care and attention. Initial categorization can be done based on the participant's expertise, past experience with design tools, educational background and knowledge. It is, but essential for the researcher to check the quality and attributes desired in the participants for the experiment he/she is trying

to conduct. While recreating existing experiments, care is needed to select participants who represent the original case.

- f. **Experiment results:** Similar to all other elements, a repository to store the results from the experiments conducted is essential for the framework. Participants can report their outcomes, which can be stored for evaluation by the researcher. Researchers can then compare the results obtained from with the outcomes reported in original studies to verify the original claims.

APPENDICES

Appendix A: List of design problems and authors

Table A-1: List of design problems and their statements collected from 34 studies

DP	Problem statement	Problem name
DP1	Sketch Ideas for subway Improvement	Subway improvement
DP2	There is a need of designing a new drawing table that took up as little space as possible when not in use. Actually housing is becoming smaller, so one of the main goals of furniture design is to achieve the same functionality with the less space. The table is addressed to professionals or students in the field of design, architecture, engineering, etc., that need to draw at home.	New drawing table
DP3	Tubular map cases consist of a system for storage and transportation of maps. The new design requested must facilitate the introduction and extraction of one single plan. Currently tubular map cases require extracting all the plans in order to look for the desired one and separate it from the rest. After that, the other ones must be put again inside the tubular map case. The new system is addressed to professionals or students in the field of design, architecture, engineering, etc.	Tubular map case
DP4	Design a new system for gathering together and hiding the wires of the electronic equipment in an office table. Currently the work in the field of design, architecture and engineering needs of a personal computer, printers, and scanners. Each of these devices needs of electrical supply and the wires on table surface are annoying. Actually, there are simple solutions to gather them, but it is difficult to extract or introduce a wire, or they leave the wires hanging behind the table.	System to collect and hide electronic wires
DP5	It is asked to design a new table for offices that allows alternate sitting and stand up work. There are a lot of people who must work on sitting position the full day. The possibility to alternate positions during working time could drive to an improvement in health and productivity. The current tables that allow combining positions in work have limited surface, not enough for design, architecture and engineering needs.	New table for offices
DP6	Design a wearable binocular which satisfies the following constraints: 1. Both eyes should be used. 2. The product should be wearable using head or face of a user. 3. Manual adjustment is allowed for controlling lens and focus.	Wearable binocular
DP7	Design an urban (bi or tri) cycle for use by “white-collar workers”.	Bi/tri cycle
DP8	Design a water lifting device	Water lifting device

DP	Problem statement	Problem name
DP9	Design and build a low-cost, easy to manufacture peanut shelling machine that will increase the productivity of the African peanut farmers. Target throughput is approximately 50 Kg per hour. The goals include: a. Must remove the shell with minimal damage to peanuts b. Electrical outlets are not available as a power source c. A large quantity of peanuts must be quickly shelled.	Peanut shelling machine
DP10	Redesign of a traffic light that uses light-emitting diodes (LED) instead of incandescent bulbs leading to snow build-up on the lights in certain climates as LED's generate less heat to melt the snow	Traffic light using LED
DP11	Develop concepts for a new, innovative, product that can froth milk in a short amount of time. The product should be able to be used by the consumer with minimal instruction	Milk frothing device
DP12	Develop products that utilize sunlight for heating and cooking food. The products should be portable and made of inexpensive materials. It should be able to be used by individual families, and should be practical for adults to set up in a sunny spot. Specific materials requirements for a targeted temperature can be postponed to a later stage.	Solar device
DP13	Design a counter top stand to display and dispense candy and chocolate snacks at convenience stores. The requirements of this task are: a) the stand must be easy to use both by the final user to grab the product and by the shop attendant to refill the product, b) the stand must contain and visually identify one specific target brand and product presentation, c) the stand must be built in one single material to choose between cardboard or laminated plastic (PVC, PS or PETG), and d) the stand must be innovative, yet simple to manufacture and assemble.	Counter top stand
DP14	Design a device (i.e. to generate sketches) to transport a ping-pong ball the farthest distance powered only by a standard issue compression spring. The device is to be constructed with a limited set of given materials (e.g. balsa wood, wire and Styrofoam).	Ping pong ball transporter
DP15	Generate simple tools (i.e. no need for electricity, motors, computers, etc.) for an intelligent species in another planet. Draw, label and describe as many tools as possible.	Tool for alien species
DP16	In rural areas of developing countries, such as Kenya, cooking is done in the home with biomass type cooking systems. One of the adverse affects of these cooking systems is the emissions which cause respiratory illnesses for millions of children and women. The people in these developing countries are economically and culturally constrained by the types of cooking systems they use. Also, depending on the type of biomass used there can be unsustainable and detrimental effects on the environment. Develop several concepts for a cooking system that is culturally appropriate, sustainable and low cost to meet the needs of rural Kenya.	Biomass cooking device

DP	Problem statement	Problem name
DP17	<p>The Air Force wants to increase the functionality of their Unmanned Air Vehicle technology by using transformation to move between states that are suited to tagging, tracking, and performing an effect on targets of interest.</p> <p>Tag: Track:</p> <p>Physically mark or create a way to easily find and identify a target, even after direct contact has been lost. Possible means include, but are not limited to physically attaching, chemically marking, using energy or radiation, biological means, and visual or other sensory marking.</p> <p>Be able to continuously or sporadically determine certain characteristics of the target, such as position, velocity, size, strength, etc. Be able to lose direct visual contact but still reacquire and positively identify target when it reappears.</p> <p>Effect:</p> <p>Be able to remotely trigger a desired effect on the target when desired. Destruction or disablement are possible effects, but options for non-destructive actions specific to the mission are desirable.</p>	UAV
DP18	<p>Devise an innovative solution for a remote controller that meets the following specification: 1. Performs its functions using voice/sound-based commands. 2. Has intelligent yet easy-to-use functions, and uses maximum number of sensors of different kinds. The type of sensors may include altitude sensor, microphone, flow, force, pressure, proximity, stress and strain, temperature, vibration and wind speed. The device can allow the following functions: 1. Selection among multiple choices. 2. Set parameters. 3. Define sequence of selected choices. 4. Set time limits. 5. Associate responses to choices. 6. Set priority for multiple choices. The device should give the following responses to the user: 1. Indicate performed function. 2. Offer short demo of results. 3. Warn about impossible problems. 4. Offer different options of communication like audio, video, text, vibration.</p>	Innovative remote controller
DP19	Sketch a toaster with a side panel that you can draw on; the drawing can then be toasted onto the bread	Doodle toaster
DP20	Sketch a combination toaster/coffee maker.	Coffee maker
DP21	Sketch a horizontal toaster such that the bread is inserted and comes out horizontal to the ground.	Horizontal toaster
DP22	Sketch a toaster that has a removable crumb tray at the bottom.	Crumb tray toaster
DP23	Oars often propel boats that operate manually (human powered). However, oars can be difficult to maneuver. Inexperienced operators tire quickly, and if the oars are not used correctly, they rock the boat, and splash water on the deck where people are sitting. Your task is to develop designs for alternative means besides oars to manually propel boats.	Oars

DP	Problem statement	Problem name
DP24	One difficult sorting task is separating paper and plastic in curbside recycling systems, which is usually done by hand. Develop concepts that will enable removing paper or plastic from the mixed collection.	Waste sorting device
DP25	Lunar dust poses significant problems for space equipment and astronauts during operations on the Moon. These dust particles are very abrasive and have a tendency to stick to each other and other objects because of their rough surfaces. One essential device that must be protected from lunar dust is a Lydar. A Lydar is an optical device that produces laser for signaling purposes. It must be enclosed and protected while not in use. In past lunar operations, dust particles accumulated on the cover joints and lens during and after opening/closing of the lens cover. Develop concepts that effectively achieve protection from lunar dust. You should also consider the environment of the Moon, i.e., a low gravitational force, low atmospheric pressure, extreme low and high temperatures, etc.	Lunar device
DP26	Design a machine that registers a bottle to a capping station, caps it, and allows somebody to retrieve the capped bottle from the device.	Bottle capping device
DP27	Design a device that takes water, sodium bicarbonate (gas), and soda flavor syrup as input and mixes them into a soda drink. The device is targeted as a home type kitchen appliance. The inputting of the water can be accomplished through a standard kitchen faucet. Please assume that the soda flavor syrup is available in a separate container that can be poured into the device you are designing, and the sodium bicarbonate is contained in a canister that can safely transfer sodium bicarbonate into the system.	Soda maker
DP28	Redesign an electric toothbrush for increased portability.	Electric toothbrush
DP29	Design the next generation of breakthrough alarm clock.	Next Generation alarm clock
DP30	Design a litter collection system for use by student groups in volunteer "Adopt-A-highway" activities.	Litter collection device
DP31	People generally use their mobile phone or a traditional alarm clock in order to wake themselves up every morning. Yet, they are not always effective and can sometimes cause oversleeping. What would be other possible ways to wake up in the morning in a certain time without using any form of alarm clock? Generate as many ideas as possible.	Alternative alarm clock
DP32	Design a pumping unit to extract petroleum	Petroleum pumping unit

DP	Problem statement	Problem name
DP33	Since ancient times, transportation of people and goods has always been an essential human activity. Despite the rapid technological developments in the field of human transportation, it is still uncertain how this area will unfold in the future. Your task is to think about how human transportation will be like in 2050. You are kindly asked to draw as many different ideas as you can in 45 minutes	Human transportation system
DP34	Design a device to collect energy from human motion for use in developing and impoverished rural communities in places like India and many African countries. Our goal is to build a low-cost, easy to manufacture device targeted at individuals and small households to provide energy to be stored in a rechargeable battery with approximately 80% efficiency. The energy is intended to be used by small, low power draw electrical devices, such as a radio or lighting device, hopefully leading to an increase in the quality of life of the communities by increasing productivity, connection to the outside world, etc. The target energy production is 1 kW-h per day, roughly enough to power eight 25W compact florescent light bulbs for 5 h each per day, or enough to power a CB radio for the entire day. For reference, an average adult human can output about 200W with full body physical activity for short periods of time, with a significant reduction for sustained power output.	Device to collect energy from human motion
DP35	Due to potential foe leg injuries, MTREK is now requiring guides to carry additional supplies to treat leg and ankle injuries. In the design challenge, MTREK has hired you to design a device that can be used to immobilize a joint or limb in case of an extreme injury. This device must be (1) as light and small as possible when stored in the guide's packs but (2) rigid and large enough to immobilize the leg of an average sized male.	Leg immobilization device
DP36	Design a safety lock for a bicycle that is to be permanently fastened to it- not to be removed when used, i.e the lock function while attached to the bicycle frame. The lock is to be a lasting accessory, yet can still be removed or adjusted if necessary. Therefore, it should be small enough so as to be non-obstrusive to the bicyclist when riding and should, as well, be light weight, durable and relatively inexpensive	Bicycle lock
DP37	In order to help handicapped people (not able to stand up and grab the book) in wheel chairs catch books at the highest level of the bookshelf (at 6ft or above), a mechanism needs to be developed. the following performance requirements must be met: Convenient and safe to use Smooth operation without damaging the books Relatively simple assembly (can be installed on most existing bookshelves)	Mechanism to grab books

Appendix B: Design problem network analysis and representation

Table B-1: List of authors and design problems used by them for 34 studies collected

Reference no.	Author/s	Problem name
[44]	S. Kim (Faculty) Y. Kim (Faculty)	Subway improvement (DP1)
[24]	Sonseca (Faculty), Mulet (Faculty), Chakrabarti (Faculty)	1) New table (DP2) 2) Tubular map case (DP3) 3) System for gathering wire (DP4) 4) Table for offices (DP5)
[45]	S. Kim (Faculty) H. Kim (Student) Jin (Faculty)	Wearable binocular (DP6)
[64]	Rogers (Student) Salustri (Faculty)	Bi/tri cycle (DP7)
[67]	Naim (Student) Lewis (Faculty) S. Schmidt (Student) Viswanathan (Faculty) Linsey (Faculty) McAdams (Faculty) Campbell (Faculty) Poppa (Faculty) Robert (Faculty)	1) Water lifting device (DP8) 2) Peanut shelling machine (DP9)
[25]	Hernandez (Faculty) C. Schmidt (Faculty) Okudan (Faculty)	Traffic light using LED (DP10)
[68]	Toh (Student) Miller (Faculty)	Milk frothing device (DP11)
[69]	Linsey (Faculty), Markman (Faculty), Wood (Faculty)	1) Peanut shelling machine (DP9)
[23]	Daly (Faculty) Christian (Student)	Solar device (DP12)
[50]	Glier (Faculty) S. Schmidt (Student) Viswanathan (Faculty) Linsey (Faculty) McAdams (Faculty)	Peanut shelling machine (DP9)

Reference no.	Author/s	Problem name
[42]	Linsey (Faculty) Clauss (Faculty) Kurtoglu (Faculty) Murphy (Student) Wood (Faculty) Markman (Faculty)	Peanut Shelling machine (DP9)
[70]	Acuna (Student) Sosa (Faculty)	Counter top stand (DP13)
[22]	Hernandez (Faculty) Shah (Faculty) Smith (Faculty)	1) Ping pong ball transporter (DP14) 2) Tool for alien species (DP15)
[71]	C. Schmidt (Faculty) Hernandez (Faculty), Kremer (Faculty), Linsey (Faculty)	Biomass cooking device (DP16)
[72]	Weaver (Student), Wang (Student), Kuhr (Student), Crawford (Faculty)	UAV (DP17)
[73]	Doboli (Faculty) Umbarkar (Faculty)	Innovative remote controller (DP18)
[74]	Kudrowitz (Faculty), Te (Student), Wallace (Faculty)	1. Doodle toaster (DP19) 2. Coffee maker (DP20) 3. Horizontal toaster (DP21) 4. Crumb tray toaster (DP22)
[75]	Lopez (Student) Linsey (Faculty) Smith (Faculty)	Peanut Shelling machine (DP9)
[76]	Okudan (Faculty) Hernandez (Faculty), Jablokow (Faculty) C. Schmidt (Faculty) Lin (Faculty)	Traffic light using LED (DP10)
[77]	Jin (Faculty) Benami (Engineer)	Oars (DP23)
[78]	Cheong (Student), Chiu (Faculty) Shu (Faculty)	1) Waste sorting device (DP24) 2) Lunar device (DP25)
[79]	Kurtoglu (Faculty), Campbell (Faculty), Linsey (Faculty)	1) Bottle capping device (DP26) 2) Soda maker (DP27)

Reference no.	Author/s	Problem name
[80]	Tsenn (Faculty), Atilola (Student), McAdams (Faculty) Linsey (Faculty)	Peanut Shelling machine (DP9)
[81]	Toh (Student) Miller (Faculty) Okudan (Faculty)	Electric toothbrush (DP28)
[16]	Johnson (Student), Genco (Student), Paul Seepersad (Faculty) Otto (Faculty)	1) Next Generation alarm clock (DP29) 2) Litter collection device (DP30)
[82]	Shorachi (Faculty), Goncalves (Student)	Alternative alarm clock (DP31)
[83]	Lujun (Sole author)	Petroleum pumping unit (DP32)
[84]	Cardoso (Faculty), Goncalves (Student), Badke-Schaub (Faculty)	Human transportation system (DP33)
[85]	Chan (Faculty), Fu (Faculty), Schunn (Faculty), Cagan (Faculty), Wood (Faculty), Kotovskiy (Faculty)	Device to collect energy from human motion (DP34)
[86]	Cardoso (Faculty), Goncalves (Faculty), Badke-Schaub (Faculty)	Human transportation system (DP33)
[87]	Chan (Faculty) Fu (Faculty) Schunn (Faculty), Cagan (Faculty), Wood (Faculty), Kotovskiy (Faculty)	Device to collect energy from human motion (DP34)
[88]	Wilson (Student), Rosen (Faculty)	Leg immobilization device (DP35)
[89]	Vishwanathan (Faculty) Linsey (Faculty)	Peanut Shelling machine (DP9)
[90]	McKoy (Student), Hernandez (Faculty), Summers (Faculty), Shah (Faculty)	1) Bicycle lock (DP36) 2) Mechanism to grab books (DP37)

Table B-2: List of node numbers and type of nodes for researcher-problem

Vertex no	Type	Name	Vertex no	Type	Name
1	Researcher	S. Kim	41	Researcher	Cardoso
2	Researcher	Se Kim	42	Researcher	Goncalves
3	Researcher	Chulvi	43	Researcher	Badke-Schaub
4	Researcher	Sonseca	44	Researcher	Chan
5	Researcher	Mulet	45	Researcher	Fu
6	Researcher	Chakrabarti	46	Researcher	Schunn
7	Researcher	S. Jin	47	Researcher	Cagan
8	Researcher	Salustri	48	Researcher	Kotovskiy
9	Researcher	Lewis	49	Researcher	Rosen
10	Researcher	Linsey	50	Researcher	Summers
11	Researcher	McAdams	51	Problem	DP1
12	Researcher	Campbell	52	Problem	DP2
13	Researcher	Robert	53	Problem	DP3
14	Researcher	Vishwanathan	54	Problem	DP4
15	Researcher	Hernandez	55	Problem	DP5
16	Researcher	C. Schmidt	56	Problem	DP6
17	Researcher	Okudan	57	Problem	DP7
18	Researcher	S. Miller	58	Problem	DP8
19	Researcher	Markman	59	Problem	DP9
20	Researcher	Wood	60	Problem	DP10
21	Researcher	Daly	61	Problem	DP11
22	Researcher	Glier	62	Problem	DP12
23	Researcher	Acuna	63	Problem	DP13
24	Researcher	Shah	64	Problem	DP14
25	Researcher	Smith	65	Problem	DP15
26	Researcher	Crawford	66	Problem	DP16
27	Researcher	Doboli	67	Problem	DP17
28	Researcher	Kudrowitz	68	Problem	DP18
29	Researcher	Wallace	69	Problem	DP19
30	Researcher	Jablokow	70	Problem	DP20
31	Researcher	Y. Jin	71	Problem	DP21
32	Researcher	Cheong	72	Problem	DP22
33	Researcher	Chiu	73	Problem	DP23
34	Researcher	Shu	74	Problem	DP24
35	Researcher	Tsenn	75	Problem	DP25
36	Researcher	R. Miller	76	Problem	DP26
37	Researcher	Seepersad	77	Problem	DP27
38	Researcher	Otto	78	Problem	DP28
39	Researcher	Shorachi	79	Problem	DP29
40	Researcher	Lujun	80	Problem	DP30

Vertex no	Type	Name
81	Problem	DP31
82	Problem	DP32
83	Problem	DP33
84	Problem	DP34
85	Problem	DP35
86	Problem	DP36
87	Problem	DP37

Appendix C: Design problem similarity assessment

Table C-1: Protocol used for inter-rater reliability test

S.No.	Characteristic	Questions asked	Scoring system
1	Number of goals	What is the final objective of the problem statement? Does the problem statement ask only to design one object or more?	Count the number of goals or objectives mentioned in the problem statement.
2	Functional requirements (FR)	How many primary functions can you find? How many action verbs can you identify?	Count the number of functional requirements given in the problem statement. There can be 2 cases: 1: When a new product design is desired: In this case, FR should be specified in the problem statement. Else, give a score of 0.
3	Non-functional requirements	How many non-action and non-functional aspects can you identify? How many performance and usability aspects can you identify?	Count the number of Non-functional requirements given in the problem statement.
4	Information about end user	Can you identify who is going to use the product?	Check the problem statement to see if any information about the end user is provided or not. If yes, give a score of 1, else a 0
5	Reference to existing product.	Do you know if the product that needs to be designed exists?	If a reference to existing product exists in the problem statement, assign a score of 1, else 0

Task 1

GOAL

FR1

FR2

Design a new system for gathering together and hiding the wires of the electronic equipment in

an office table. Currently the work in the field of design, architecture and engineering needs of a

personal computer, printers, and scanners. Each of these devices needs of electrical supply and

the wires on table surface are annoying. Actually, there are simple solutions to gather them, but

it is difficult to extract or introduce a wire, or they leave the wires hanging behind the table.

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info (Y=1/N=0)
1	2	0	1	1

Figure C-1: Scores for evaluator 1 for problem 1

Task 2

GOAL

Goal

FRI

FRI

Design an automatic recycler device that can automatically sort plastic bottles, glass containers,

aluminum cans, and tin cans. The major differentiation between types of materials lay with the

given dimensions of the products: plastic bottles are the tallest, glass containers are very short

and heavy, and aluminum cans are lightweight. Devices are given strict requirements to adhere

NFR1

NFR2

NFR3

to such as volume and weight constraints, safety requirements, and most importantly, have to

NFR1

NFR2

NFR3

operate autonomously once a master shutoff switch is toggled.

NFR4

NFR4

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info (Y=1/N=0)
0	1	4	1	0

Figure C-2: Scores for evaluator 1 for problem 2

Task 3

GOAL

Goal

A mechanical system is required which, in the event of a fire, will enable people to escape from

a six-storey building by lowering themselves to the ground from windows. The system, which

might make use of a 5 mm diameter steel cable, must be capable of lowering either a small child

or a heavy adult at approximately the same constant speed.

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	1	2	1	1

Figure C-3: Scores for evaluator 1 for problem 3

Task 4

Goal (GOAL)

FRI (FRI)

(END USER)

Design and develop an artifact to facilitate grocery shopping in a typical French/Singapore Chinatown fresh market. The artifact should facilitate carrying of groceries from fresh market to home.

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	1	0	1	1

Figure C-4: Scores for evaluator 1 for problem 4

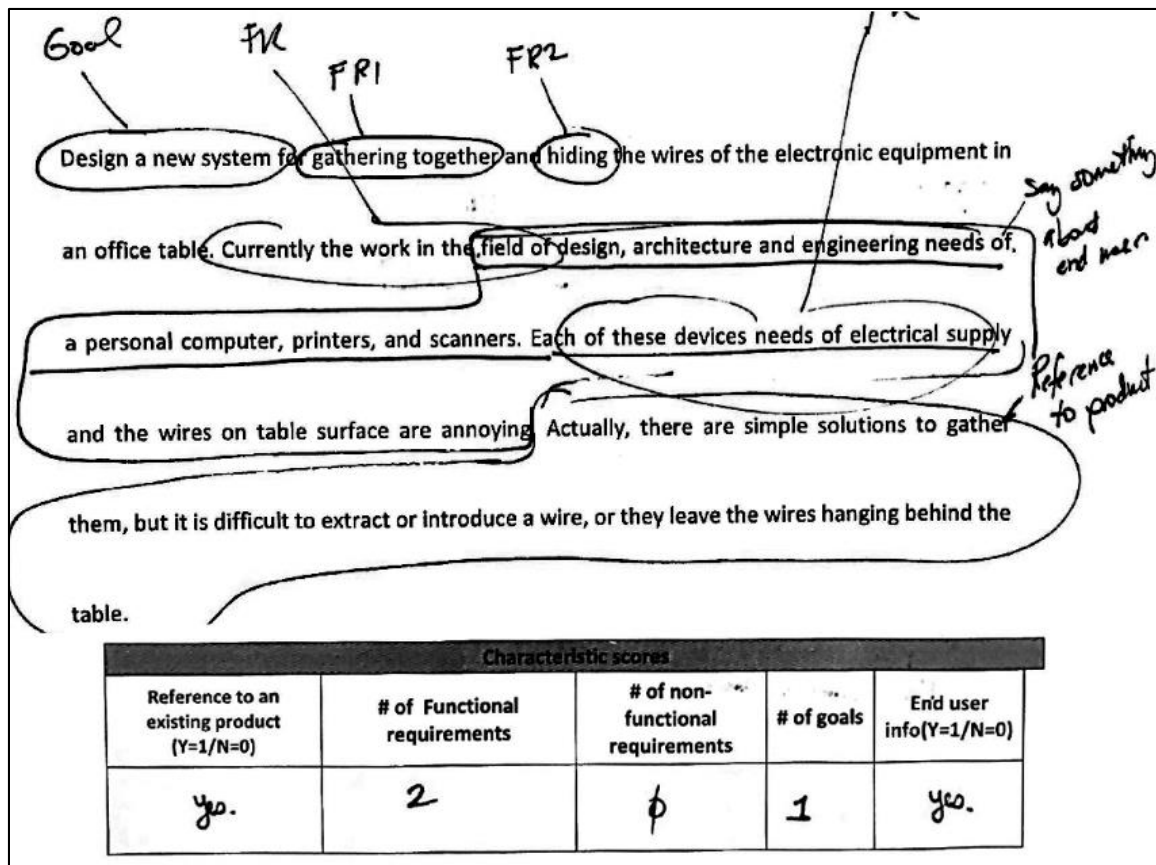


Figure C-5: Scores for evaluator 2 for problem 1

Task 2

Goal

FR₁

Design an automatic recycler device that can automatically sort plastic bottles, glass containers,

aluminum cans, and tin cans. The major differentiation between types of materials lay with the

given dimensions of the products: plastic bottles are the tallest, glass containers are very short

and heavy, and aluminum cans are lightweight. Devices are given strict requirements to adhere

to such as volume and weight constraints, safety requirements, and most importantly, have to

NF₁

NF₂

NF₃

operate autonomously once a master shutoff switch is toggled.

FR₁ - same as above

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info (Y=1/N=0)
N ₀	1	3	1	N ₀

Figure C-6: Scores for evaluator 2 for problem 2

Task 3

A mechanical system is required which, in the event of a fire, will enable people to escape from

a six-storey building by lowering themselves to the ground from windows. The system, which

might make use of a 5 mm diameter steel cable must be capable of lowering either a small

child or a heavy adult at approximately the same constant speed.

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
No	1	2	1	Yes.

Figure C-7: Scores for evaluator 2 for problem 3

Task 4

Design and develop an artifact to facilitate grocery shopping in a typical French/Singapore

Chinatown fresh market. The artifact should facilitate carrying of groceries from fresh market to

home.

Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	1	1 (Complete)	1	0

Figure C-8: Scores for evaluator 2 for problem 4

Goal GOAL FR1 FR2
 F.R. 1 F.R. 2

Design a new system for gathering together and hiding the wires of the electronic equipment in an office table. Currently the work in the field of design, architecture and engineering needs of a personal computer, printers, and scanners. Each of these devices needs of electrical supply and the wires on table surface are annoying. Actually, there are simple solutions to gather them, but it is difficult to extract or introduce a wire, or they leave the wires hanging behind the table.

(END USER)

(REF TO EXISTING)

FR 1

NFR 1

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info (Y=1/N=0)
1	2	1	1	1

Figure C-9: Scores for evaluator 3 for problem 1

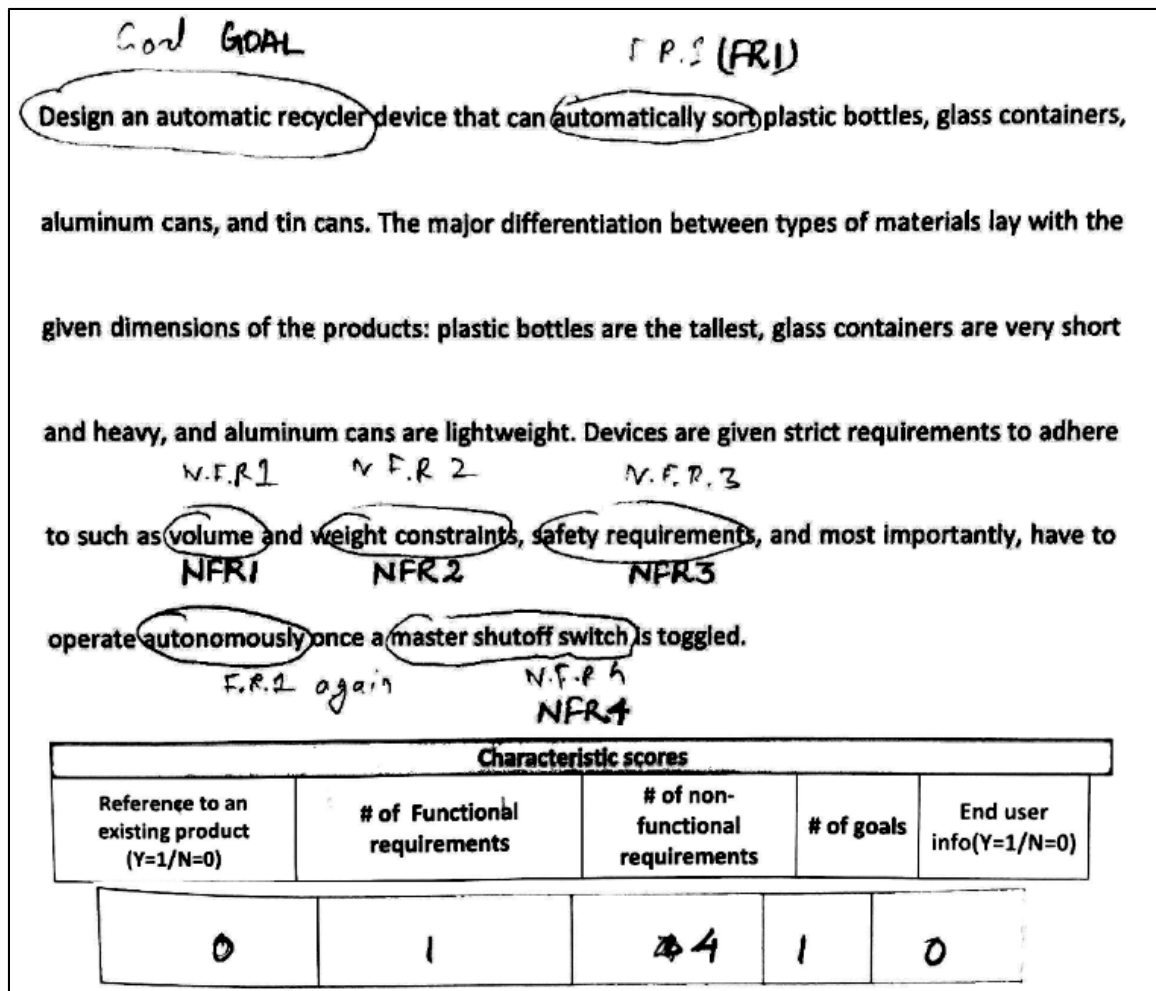


Figure C-10: Scores for evaluator 3 for problem 2

Goal GOAL

A mechanical system is required which, in the event of a fire, will enable people to escape from a six-storey building by lowering themselves to the ground from windows. The system, which might make use of a 5 mm diameter steel cable, must be capable of lowering either a small child or a heavy adult at approximately the same constant speed.

F.R. 1 FRI

N.F.R. 2 NFR1

N.F.R. 2 NFR2

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	1	2	1	0

Figure C-11: Scores for evaluator 3 for problem 3

Goal GOAL

End user (END USER)

Design and develop an artifact to facilitate grocery shopping in a typical French/Singapore Chinatown fresh market. The artifact should facilitate carrying of groceries from fresh market to home.

F.R. 1 (FRI)

Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	1	0	1	1

Figure C-12: Scores for evaluator 3 for problem 4

GOAL

FR1

FR2

Design a new system for gathering together and hiding the wires of the electronic equipment in an office table. Currently the work in the field of design, architecture and engineering needs of a personal computer, printers, and scanners. Each of these devices needs of electrical supply and the wires on table surface are annoying. Actually, there are simple solutions to gather them, but it is difficult to extract or introduce a wire, or they leave the wires hanging behind the table.

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	2	0	1	0

Figure C-13: Scores for evaluator 4 for problem 1

(GOAL) FR1				
Design an automatic recycler device that can automatically sort plastic bottles, glass containers, aluminum cans, and tin cans. The major differentiation between types of materials lay with the given dimensions of the products: plastic bottles are the tallest, glass containers are very short and heavy, and aluminum cans are lightweight. Devices are given strict requirements to adhere to such as volume and weight constraints, safety requirements, and most importantly, have to operate autonomously once a master shutoff switch is toggled.				
NFR1 NFR2 NFR3 1 NFR 2 NFR 3 NFR				
4 NFR NFR4				
Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	1	4	1	0

Figure C-14: Scores for evaluator 4 for problem 2

A mechanical system is required which, in the event of a fire, will enable people to escape from a six-storey building by lowering themselves to the ground from windows. The system, which might make use of a 5 mm diameter steel cable, must be capable of lowering either a small child or a heavy adult at approximately the same constant speed.

NFR1

NFR2

NFR3

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	1	3	1	1

Figure C-15: Scores for evaluator 4 for problem 3

GOAL

Design and develop an artifact to facilitate grocery shopping in a typical French/Singapore Chinatown fresh market. The artifact should facilitate carrying of groceries from fresh market to home.

FR1

Characteristic scores				
Reference to an existing product (Y=1/N=0)	# of Functional requirements	# of non-functional requirements	# of goals	End user info(Y=1/N=0)
0	1	0	1	0

Figure C-16: Scores for evaluator 4 for problem 4

Table C-2: Scores for element 'No. of goal' for all problems

	Score			
Problem number	E1	E2	E3	E4
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1

Table C-3: Scores for element 'No. of functional requirements' for all problems

	Score			
Problem number	E1	E2	E3	E4
1	2	2	2	2
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1

Table C-4: Scores for element 'No. of non - functional requirements' for all problems

	Score			
Problem number	E1	E2	E3	E4
1	0	0	1	0
2	4	3	4	4
3	2	2	2	3
4	0	1	0	0

Table C-5: Scores for element 'Information about end user' for all problems

	Score (Yes = 1, No = 0)			
Problem number	E1	E2	E3	E4
1	1	1	1	0
2	0	0	0	0
3	1	1	0	1
4	1	0	1	0

Table C-6: Scores for element 'Reference to existing product' for all problems

	Score (Yes = 1, No = 0)			
Problem number	E1	E2	E3	E4
1	1	1	1	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0

Table C-7: LSA scores for design problem statements. Top row and first column represent the different problem statements.

	D P 1	D P 2	D P 3	D P 4	D P 5	D P 6	D P 7	D P 8	D P 9	D P 10	D P 11	D P 12	D P 13	D P 14	D P 15	D P 16	D P 17	D P 18	D P 19	D P 20	D P 21	D P 22	D P 23	D P 24	D P 25	D P 26	D P 27	D P 28	D P 29	D P 30	D P 31	D P 32	D P 33	D P 34	D P 35	D P 36	D P 37
D P 1	1 . 1 3	0 . 0 9	0 . 0 6	0 . 0 3	0 . 1 5	0 . 0 5	0 . 0 7	0 . 0 3	0 . 0 6	0 . 1 2	0 . 0 5	0 . 1 1	- 0 4	0 . 0 4	0 . 0 1	0 . 0 2	0 . 0 8	0 . 0 4	0 . 0 1	0 . 2 0	0 . 0 4	0 . 0 7	0 . 0 1	0 . 0 9	- 0 3	0 . 0 3	0 . 0 8	0 . 0 7	0 . 0 5	0 . 0 4	0 . 0 2	0 . 0 2	0 . 0 2	0 . 0 2	0 . 0 1	0 . 0 2	
D P 2	0 . 1 3	1 . 3 5	0 . 4 5	0 . 4 1	0 . 4 6	0 . 1 2	0 . 1 9	0 . 1 3	0 . 0 6	0 . 1 7	0 . 1 3	0 . 2 1	0 . 2 4	0 . 2 7	0 . 1 6	0 . 2 1	0 . 2 6	0 . 2 2	0 . 1 2	0 . 3 6	0 . 2 2	0 . 2 7	0 . 1 2	0 . 0 7	0 . 2 4	0 . 1 5	0 . 2 1	0 . 2 3	0 . 1 4	0 . 2 6	0 . 1 9	0 . 2 5	0 . 1 6	0 . 1 5	0 . 1 9	0 . 2 1	
D P 3	0 . 0 9	0 . 3 5	1 . 3 1	0 . 2 4	0 . 0 9	0 . 1 1	0 . 1 3	0 . 1 1	0 . 0 1	0 . 1 9	0 . 0 2	0 . 1 5	0 . 1 3	0 . 0 5	0 . 1 5	0 . 0 8	0 . 1 4	0 . 0 2	0 . 1 8	0 . 0 4	0 . 0 6	0 . 1 9	0 . 1 1	0 . 0 1	0 . 1 1	0 . 2 5	0 . 1 5	0 . 0 4	0 . 1 9	0 . 0 7	0 . 1 2	0 . 1 1	0 . 1 8	0 . 1 5	0 . 1 8	0 . 1 5	
D P 4	0 . 0 6	0 . 4 5	0 . 3 1	1 . 5 4	0 . 0 9	0 . 1 6	0 . 2 2	0 . 2 2	0 . 1 1	0 . 0 4	0 . 0 9	0 . 1 6	0 . 2 7	0 . 0 5	0 . 4 1	0 . 3 5	0 . 1 5	0 . 2 6	0 . 1 4	0 . 0 9	0 . 1 1	0 . 1 8	0 . 0 7	0 . 1 9	0 . 2 2	0 . 3 7	0 . 1 1	0 . 2 4	0 . 1 5	0 . 2 4	0 . 1 4	0 . 2 2	0 . 1 4	0 . 2 1	0 . 1 3	0 . 2 3	
D P 5	0 . 1 3	0 . 4 1	0 . 2 4	0 . 5 4	1 . 0 8	0 . 2 2	0 . 2 3	0 . 1 8	0 . 0 5	0 . 0 7	0 . 1 1	0 . 1 9	0 . 2 1	0 . 0 7	0 . 1 7	0 . 2 4	0 . 1 7	0 . 2 3	0 . 0 9	0 . 1 7	0 . 0 6	0 . 1 8	0 . 0 5	0 . 2 6	0 . 1 2	0 . 2 1	0 . 1 6	0 . 0 8	0 . 1 4	0 . 0 3	0 . 1 4	0 . 1 9	0 . 0 4	0 . 1 9	0 . 2 8	0 . 1 9	
D P 6	0 . 0 5	0 . 1 6	0 . 0 9	0 . 0 8	1 . 0 1	0 . 1 1	0 . 1 1	0 . 1 6	0 . 0 3	0 . 0 6	0 . 3 3	0 . 2 6	0 . 1 5	0 . 1 4	0 . 1 6	0 . 1 4	0 . 0 9	0 . 0 7	0 . 0 6	- 0 2	0 . 1 1	0 . 0 1	0 . 0 7	0 . 0 4	0 . 1 2	0 . 0 4	0 . 0 9	0 . 0 8	0 . 0 1	0 . 1 4	0 . 0 4	0 . 0 2	0 . 0 4	0 . 2 6	0 . 3 6	0 . 1 4	
D P 7	0 . 0 5	0 . 2 2	0 . 1 1	0 . 1 6	0 . 2 2	0 . 1 1	0 . 1 5	0 . 0 3	0 . 0 3	0 . 0 4	0 . 1 1	0 . 2 2	0 . 0 8	0 . 0 2	0 . 0 8	0 . 0 7	0 . 0 3	0 . 0 9	0 . 0 9	- 0 3	0 . 1 7	0 . 0 4	0 . 0 7	0 . 0 4	0 . 1 2	0 . 0 1	0 . 0 9	0 . 0 2	0 . 0 8	0 . 0 8	0 . 1 4	0 . 0 2	0 . 1 2	0 . 1 6	0 . 0 9	0 . 0 9	
D P 8	0 . . 1	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	1 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .	0 . . .

	D P 1	D P 2	D P 3	D P 4	D P 5	D P 6	D P 7	D P 8	D P 9	D P 1 0	D P 1 1	D P 1 2	D P 1 3	D P 1 4	D P 1 5	D P 1 6	D P 1 7	D P 1 8	D P 1 9	D P 2 0	D P 2 1	D P 2 2	D P 2 3	D P 2 4	D P 2 5	D P 2 6	D P 2 7	D P 2 8	D P 2 9	D P 3 0	D P 3 1	D P 3 2	D P 3 3	D P 3 4	D P 3 5	D P 3 6	D P 3 7
	0 7	1 9		2 2	2 3	1 1	1 5		0 9	0 2	0 8	0 7	1 3	1 5	1 2	0 9	0 8	0 9	1 5	1 3	0 6	1 4	4 4	0 5	0 8	2 8	5 3	0 3	1 5	1 1	0 5	2 1	0 2			1 4	1 1
D P 9	0 0 3	0 1 3	0 1 3	0 2	0 8	0 1 6	0 1 3	0 0 9	1 0	0 9	0 1 7	0 2	0 4	0 1 9	0 2 2	0 2 3	0 2 9	0 1 1	0 2 1	0 0 2	0 0 8	0 0 6	0 1 9	0 1 1	0 2 2	0 2 9	0 1 6	0 2 3	0 0 6	0 0 6	0 0 7	0 1 9	0 1 7	0 5	0 2 6	0 2 7	0 2 6
D P 1 0	0 0 6	0 0 6	0 0 1	0 1	0 5	0 0 3	0 0 3	0 0 2	0 0 9	1 0	0 5	0 1 4	0 0 9	0 0 7	0 0 5	0 0 8	0 0 8	0 0 8	0 0 5	0 0 2	0 0 3	- 0 3	0 0	0 1	0 0 4	0 0 7	0 2 2	- 0 2	0 0 4	0 0 4	0 0 2	- 0 2	0 2 2	0 2 3	0 2 9	0 0 2	
D P 1 1	0 1 2	0 1 7	0 1 1	0 0 4	0 0 7	0 3 6	0 0 4	0 0 8	0 1 7	0 0 5	1 0	0 4 3	0 4 5	0 1 2	0 1 4	0 1 8	0 1 1	0 1 1	0 1 2	0 0 8	0 0 5	0 0 1	0 1 1	0 1 5	0 1 6	0 1 8	0 0 4	0 0 9	0 1 1	0 1 6	0 1 2	0 0 9	0 1 1	0 1 6	0 2 1	0 1 1	
D P 1 2	0 0 5	0 1 3	0 0 9	0 0 9	0 1 1	0 3 1	0 1 2	0 0 7	0 0 2	0 1 4	0 4 3	0 3 8	0 3 3	0 1 1	0 2 1	0 1 9	0 1 8	0 0 4	0 0 2	0 0 7	0 0 8	0 1 2	0 1 2	0 2 4	0 0 4	0 1 1	0 0 2	0 0 8	0 1 3	0 1 1	0 1 8	0 1 6	0 2 1	0 2 3	0 3 2	0 1 3	
D P 1 3	0 1 1	0 2 1	0 2 2	0 1 6	0 1 9	0 2 6	0 1 2	0 1 3	0 4 9	0 4 5	0 3 8	1 0	0 1 5	0 1 2	0 2 4	0 1 4	0 1 9	0 1 9	0 1 2	0 1 7	0 1 6	0 1 9	0 1 4	0 2 2	0 1 8	- 0 3	0 1 4	0 1 6	0 1 3	0 2 2	0 0 8	0 1 2	0 1 7	0 2 9	0 2 3		
D P 1 4	- 0 4	0 2 4	0 1 5	0 2 7	0 2 1	0 1 5	0 0 8	0 1 5	0 1 9	0 0 7	0 1 2	0 3 3	0 2 8	1 0	0 1 9	0 1 3	0 2 4	0 2 3	0 0 7	0 1 1	0 0 1	0 0 9	0 1 2	0 1 8	0 2 3	0 1 9	0 2 1	0 1 6	0 0 7	0 2 1	0 0 6	0 2 2	0 2 3	0 3 3	0 1 2		
D P 1 5	0 0 4	0 2 7	0 1 3	0 4 5	0 1 7	0 1 4	0 1 2	0 2 2	0 0 5	0 1 2	0 1 1	0 1 5	0 1 9	1 0	0 1 7	0 1 4	0 2 4	0 2 3	0 1 3	0 0 5	0 1 2	0 0 2	0 1 5	0 1 3	0 1 1	0 4 5	0 1 3	0 1 5	0 0 2	0 1 7	0 0 6	0 2 1	0 2 1	0 1 6	0 1 7	0 1 5	
D P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	D P 1	D P 2	D P 3	D P 4	D P 5	D P 6	D P 7	D P 8	D P 9	D P 10	D P 11	D P 12	D P 13	D P 14	D P 15	D P 16	D P 17	D P 18	D P 19	D P 20	D P 21	D P 22	D P 23	D P 24	D P 25	D P 26	D P 27	D P 28	D P 29	D P 30	D P 31	D P 32	D P 33	D P 34	D P 35	D P 36	D P 37	
I 6		1 6	2 5	3 1	1 7	1 6	0 8	0 9	2 2	0 7	1 4	2 1	1 2	1 3	1 7		2 9	2 5		0 8	0 7	0 3			2 8	0 8	3 2	0 6	0 4	3 4	2 2	1 1	1 4	2 5	2 2	1 6	2 8	
D P 1 7	- 0 0 1	0 2 1	0 1	0 1 5	0 2 4	0 1 4	0 0 7	0 0 8	0 2 3	0 0 8	0 1 8	0 1 9	0 2 4	0 1 4	0 2 9		1 0 9	0 0 8	0 0 8	0 1 7	0 0 7	0 2 2	0 1 5	0 3 2	0 1 2	0 0 8	0 0 7	0 1 4	0 2 9	0 0 2	1 4	2 9		2 4	0 3 4	0 2 8	0 2 4	
D P 1 8	0 1 2	0 1 8	0 0	0 2	0 0 9	0 4 9	0 0 3	0 0 9	0 1 9	0 0 8	0 1 8	0 1 4	0 2 3	0 1 4	0 2 5	0 2 9	1 0	0 0	0 0	0 0	0 0	0 1	0 1 8	0 2 3	0 1 1	0 2 1	0 0 9	0 0 3	0 1 3	0 1 5	0 0 9	0 1 1	0 3 1	0 3 6	0 2 6	0 2 1	0 1 3	
D P 1 9	0 2 6	0 3 4	0 1 6	0 1 6	0 1 7	0 0 7	0 0 9	0 1 5	0 1 1	0 0 8	0 1 1	0 0 4	0 0 9	0 2 2	0 0 8	0 0 4	1 0	0 0	0 0	0 0	0 0	0 1	0 1 2	0 1 5	0 0 4	0 2 2	0 0 9	0 0 9	0 0 9	0 0 9	0 2 5	0 1 6	0 0 7	0 1 6	0 0 6	0 0 5		
D P 2 0	0 0 8	0 2 7	0 0 4	0 1 5	0 1 7	0 0 6	0 0 9	0 1 3	0 0 2	0 0 5	0 1 2	0 0 2	0 1 7	0 0 2	0 0 8	0 0 4	0 0	0 0	0 0	0 0	0 0	0 1	0 0	0 2	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	- 0 1	- 0 4		
D P 2 1	0 0 4	0 2 2	0 1 2	0 1 9	0 2 3	0 0 4	0 0 7	0 0 6	0 0 8	0 0 2	0 0 8	0 0 7	0 1 2	0 1 3	0 1 7	0 0 2	0 0	0 0	0 0	0 0	0 0	0 1	0 2	0 0	0 1	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
D P 2 2	0 1 1	0 2 2	0 0 4	0 1 1	0 1 1	- 0 2	0 0 3	0 1 4	0 0 6	0 0 3	0 0 5	0 0 8	0 1 7	0 0 5	0 0 3	0 0 8	0 0	0 0	0 0	0 0	0 0	0 1	0 0	0 1	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0
D P 2 3	0 1 2	0 1 6	0 0 4	0 1 1	0 1 8	0 0 1	0 0 1	0 0 4	0 0 9	- 0 3	0 0 1	0 0 2	0 0 6	0 0 9	0 0 2	0 0 2	0 0	0 0	0 0	0 0	0 0	0 1	0 0	0 1	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0	0 0

	D P 1	D P 2	D P 3	D P 4	D P 5	D P 6	D P 7	D P 8	D P 9	D P 10	D P 11	D P 12	D P 13	D P 14	D P 15	D P 16	D P 17	D P 18	D P 19	D P 20	D P 21	D P 22	D P 23	D P 24	D P 25	D P 26	D P 27	D P 28	D P 29	D P 30	D P 31	D P 32	D P 33	D P 34	D P 35	D P 36	D P 37
D P 2 4	0 . 0 7	0 . 0 7	0 . 1 9	0 . 1 8	0 . 0 5	0 . 1 7	0 . 0 5	0 . 0 1	0 . 1 1	0 . . .	0 . 1 2	0 . . 2	0 . 1 9	0 . 1 9	0 . . 3	0 . 1 5	0 . 1 8	0 . 0 5	0 . 1 8	0 . 0 7	0 . 1 7	0 . 1 3	1 . . .	0 . 1 4	0 . 0 2	0 . 1 7	0 . 0 7	0 . 2 6	0 . 1 1	0 . 0 4	0 . 1 6	0 . 0 9	0 . 1 8	0 . 2 1	0 . 0 9		
D P 2 5	0 . 0 1	0 . 2 4	0 . 1 1	0 . 1 9	0 . 2 .	0 . 3 .	0 . 0 4	0 . 0 8	0 . 2 2	0 . 1 .	0 . 1 5	0 . 2 4	0 . 1 4	0 . 1 2	0 . 1 5	0 . 2 8	0 . 3 2	0 . 0 3	- . 0 2	0 . 0 3	- . 0 3	0 . 1 2	0 . 1 4	1 . . .	0 . 1 1	0 . 0 1	0 . 0 3	0 . 0 5	0 . 1 2	0 . 0 8	0 . 1 1	0 . 2 2	0 . 2 7	0 . 3 .	0 . 2 1		
D P 2 6	0 . 0 9	0 . 1 5	0 . 1 1	0 . 2 7	0 . 1 6	0 . 0 7	0 . 1 8	0 . 2 9	0 . 2 4	0 . 0 6	0 . 1 4	0 . 0 2	0 . 2 8	0 . 1 3	0 . 0 8	0 . 1 1	0 . 2 2	0 . 0 1	0 . 1 5	0 . 1 4	0 . 1 6	0 . 1 1	0 . 1 2	0 . 1 1	1 . . .	0 . 3 2	0 . 1 7	0 . 1 3	0 . 0 7	0 . 1 3	0 . 0 4	0 . 1 3	0 . 0 8	0 . 1 8	0 . 1 8		
D P 2 7	- . 0 3	0 . 2 4	0 . 2 .	0 . 3 2	0 . 2 .	0 . 1 2	0 . 1 2	0 . 5 3	0 . 1 6	0 . 0 7	0 . 1 8	0 . 1 .	0 . 1 8	0 . 1 9	0 . 1 1	0 . 3 2	0 . 1 2	0 . 2 1	0 . 2 2	0 . 1 8	0 . 1 3	0 . 2 7	0 . 1 7	0 . 1 3	0 . 3 2	1 . . .	0 . 0 8	0 . 0 8	0 . 2 1	0 . 1 5	0 . 2 6	0 . 1 3	0 . 2 2	0 . 2 5			
D P 2 8	0 . 0 3	0 . 1 .	0 . . .	0 . 3 7	0 . 1 1	- . 0 4	0 . 0 1	0 . 0 3	0 . 2 3	0 . 2 2	0 . 0 4	0 . 0 2	- . 0 3	0 . 2 3	0 . 4 5	0 . 0 6	0 . 0 8	0 . 0 9	0 . 1 3	0 . 0 7	0 . 0 4	0 . 0 2	0 . 0 1	0 . 1 2	0 . 0 8	1 . 0 3	- . 0 3	0 . . .	0 . 0 2	- . 0 2	0 . 0 7	0 . 2 7	0 . 0 9	0 . 0 6	- . 0 2		
D P 2 9	0 . 0 8	0 . 2 1	0 . 1 5	0 . 1 1	0 . 1 6	0 . 0 9	0 . 0 9	0 . 1 5	0 . 0 6	- . 0 2	0 . 0 9	0 . 0 8	0 . 1 4	0 . 1 3	0 . 0 4	0 . 0 7	0 . 0 3	0 . 2 7	0 . 0 7	0 . 0 4	0 . 0 2	0 . 0 7	0 . 0 3	0 . 1 7	0 . 0 8	- . 0 3	1 . 0 6	0 . . .	0 . 5 3	0 . 1 5	0 . 0 8	0 . 0 8	0 . 1 8	0 . 0 8	0 . 0 5		
D P 3 0	0 . 0 7	0 . 2 3	0 . 4 .	0 . 2 4	0 . 0 8	0 . 0 8	0 . 1 2	0 . 1 1	0 . 0 6	0 . 0 4	0 . 1 1	0 . 1 3	0 . 1 6	0 . 1 6	0 . 3 4	0 . 1 4	0 . . .	0 . 0 9	0 . 0 4	- . 0 2	0 . 0 5	0 . 2 6	0 . 0 5	0 . 1 3	0 . 2 1	0 . . .	0 . 0 6	1 . . .	0 . 1 2	0 . 0 5	0 . 0 8	0 . 0 8	0 . 1 1	0 . . .			
D P	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	1 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .	0 . .

	D P 1	D P 2	D P 3	D P 4	D P 5	D P 6	D P 7	D P 8	D P 9	D P 10	D P 11	D P 12	D P 13	D P 14	D P 15	D P 16	D P 17	D P 18	D P 19	D P 20	D P 21	D P 22	D P 23	D P 24	D P 25	D P 26	D P 27	D P 28	D P 29	D P 30	D P 31	D P 32	D P 33	D P 34	D P 35	D P 36	D P 37		
3 1	25	16	09	15	14	11	08	05	07	04	16	11	13	07	17	22	29	15	09	21	08	02	23	11	12	07		02	53	12		05	24	18	17	18	22		
D P 3 2	00 4	00 9	00 7	00 2	00 3	00 4	00 8	00 1	00 9	00 2	00 2	00 8	00 2	00 1	00 6	00 1	00 9	00 2	00 5	00 2	00 1	00 9	00 4	00 8	00 2	00 1	-0 2	00 5	00 1	00 5	00 1	00 5	10 5	00 5	00 4	00 3	00 7	00 1	
D P 3 3	00 2	00 1	00 1	00 1	00 1	00 0	00 4	00 0	00 2	-0 7	00 0	00 1	00 0	00 0	00 2	00 1	00 2	00 1	00 5	00 0	00 1	00 7	00 1	00 3	00 1	00 1	00 0	00 0	00 0	00 2	00 0	00 5	10 3	00 2	00 1	00 7	00 0	00 1	
D P 3 4	00 2	00 1	00 1	00 2	00 2	00 2	00 2	00 1	00 5	00 2	00 1	00 1	00 2	00 2	00 1	00 4	00 2	00 6	00 7	00 0	00 1	00 6	00 2	00 9	00 2	00 3	00 1	00 2	00 8	00 0	00 8	00 1	00 1	00 4	10 3	00 2	00 6	00 1	
D P 3 5	00 2	00 1	00 1	00 2	00 1	00 4	00 1	00 2	00 6	00 2	00 1	00 2	00 3	00 6	00 2	00 4	00 3	00 7	00 3	00 1	00 0	00 1	00 2	00 3	00 8	00 7	00 0	00 1	00 0	00 8	00 1	00 7	00 3	00 1	00 7	00 4	00 5	00 1	00 9
D P 3 6	00 1	00 1	00 1	00 1	00 3	00 6	00 6	00 4	00 7	00 2	00 6	00 3	00 2	00 3	00 1	00 6	00 2	00 6	00 0	-0 1	00 0	00 7	00 1	00 2	00 1	00 3	00 2	00 0	00 8	00 1	00 1	00 7	00 0	00 8	00 2	00 5	00 2	00 8	
D P 3 7	00 2	00 2	00 2	00 2	00 1	00 4	00 9	00 1	00 6	00 2	00 1	00 3	00 3	00 2	00 5	00 8	00 4	00 3	00 5	-0 4	00 0	00 6	00 1	00 9	00 1	00 8	00 5	-0 2	00 5	00 2	00 2	00 1	00 1	00 2	00 1	00 9	00 8	1	

Appendix D: Data for meta-regression analysis

Table D-1: Standardized mean difference and within - study errors for metric ‘quantity’

ID	Author	Year	Treatment group			Control group			t	F	SMD'	SE
			N1	Mean1	SD1	N2	Mean2	SD2				
9b.	Tsenn [80]	2014	9	12.9	3.3	18	19.2	11.9			-0.61	0.42
12a	Smith,S [130]	1993	50	2.9	1.31	44	2.82	1.06		0.11	0.07	0.21
12b	Smith,S	1993	50	3.08	0.94	44	2.95	1.18		0.33	0.12	0.21
12c	Smith,S	1993	66	NA	NA	25	NA	NA		0.22	0.11	0.23
12d	Smith,S	1993	66	NA	NA	25	NA	NA		0.46	0.16	0.24
12e	Smith,S	1993	20	3.65	0.88	20	3.55	1.1			0.10	0.32
12f	Smith,S	1993	20	3.9	1.29	20	3.55	1.1			0.29	0.32
13a	Shorachi [82]	2015	15	3.01	1.14	15	2.13	0.83			0.86	0.38
13b	Shorachi	2015	15	2.51	0.81	15	2.13	0.83			0.45	0.37
14a	Lujun [83]	2011	19	3.58	1.84	19	4	2			-0.21	0.33
14b	Lujun	2011	19	3.58	1.3	19	4	2			-0.24	0.33
14c	Lujun	2011	19	4.32	1.7	19	4	2			0.17	0.33
15a	Cardoso [84]	2012	20	5.7	1.52	18	5.39	2.57			0.15	0.33
15b	Cardoso	2012	20	5.05	2.21	18	5.39	2.57			-0.14	0.33
16a	Agogue [150]	2013	27	NA	NA	78	NA	NA	1.74		-0.39	0.22
17a	Cardoso [151]	2011	21	10.3	NA	19	8.7	NA			-0.08	0.32
17b	Cardoso	2011	20	10	NA	19	8.7	NA			-0.08	0.32
18a	Chan [85]	2011	64	3.45	2.26	24	4.38	1.02			-0.46	0.24
18b	Chan	2011	63	4.48	4.3	24	4.38	1.02			0.03	0.24
18c	Chan	2011	66	2.85	3.14	24	4.38	1.02			-0.55	0.24
18d	Chan	2011	61	5.16	3.53	24	4.38	1.02			0.25	0.24
19	Pertulla [153]	2016	8	9.38	4.07	8	8.75	2.31			0.18	0.50
20a.	Goncalves [86]	2012	19	5.37	2.14	18	5.39	2.57			-0.01	0.33
20b.	Goncalves	2012	20	7.7	5.99	18	5.39	2.57			0.48	0.33
20c.	Goncalves	2012	19	5.89	4	18	5.39	2.57			0.14	0.33
21a	Chrysikou [154]	2005	30	2.73	1.48	30	2.67	1.27			0.04	0.26
21b	Chrysikou	2005	29	2.55	1.35	30	2.67	1.27			-0.09	0.26
21c	Chrysikou	2005	30	2.93	1.17	30	2.87	1.22			0.05	0.26
21d	Chrysikou	2005	29	2.45	1.3	30	2.87	1.22			-0.33	0.26
21e	Chrysikou	2005	20	1.8	1	18	1.83	0.92			-0.03	0.32
21f	Chrysikou	2005	20	1.65	0.67	18	1.83	0.92			-0.22	0.33
21g	Chrysikou	2005	20	2.3	1.26	18	2.35	1.7			-0.03	0.32

ID	Author	Year	Treatment group			Control group			t	F	SMD'	SE
			N1	Mean1	SD1	N2	Mean2	SD2				
21h	Chrysikou	2005	20	1.7	0.73	18	2.35	1.7			-0.50	0.33
22a	Siangliulue [155]	2015	19	10.9	5.83	25	10.9	5.43			0.00	0.30
22b	Siangliulue	2015	28	13.8	6.8	25	10.9	5.43			0.46	0.28
22c	Siangliulue	2015	25	8.8	3.44	25	10.9	5.43			-0.45	0.29
23a	Atilola [128]	2014	20	13.7	9.52	20	19.6	9.52			-0.61	0.32
23b	Atilola	2014	20	10.43	9.8	20	19.6	9.52			-0.93	0.33
23c	Atilola	2014	20	11.76	10.8	20	19.6	9.52			-0.75	0.33
25a	Linsey [157]	2015	15	8.11	3.9	15	10.81	3.8			-0.68	0.38
25b	Linsey	2015	15	10.45	4.32	15	10.81	3.8			-0.09	0.37
25c	Linsey	2015	15	9.84	4.96	15	10.81	3.8			-0.21	0.37
27a	Tseng [158]	2008	17	6.83	3.41	18	5.67	1.99			0.41	0.34
27b	Tseng	2008	18	7.67	3.15	18	5.67	1.99			0.74	0.34
27c	Tseng	2008	18	7.96	3.98	18	5.67	1.99			0.71	0.34

Table D-2: Standardized mean difference and within - study errors for metric 'quality'

ID	Author	Year	Treatment group			Control group			t	F	SMD'	SE
			N1	Mean1	SD1	N2	Mean2	SD2				
14a	Lujun	2011	19	2.24	0.58	19	1.51	0.5			1.32	0.36
14b	Lujun	2011	19	2.19	0.55	19	1.51	0.5			1.27	0.36
14c	Lujun	2011	19	2.12	0.55	19	1.51	0.5			1.14	0.35
17a	Cardoso	2011	21	3.56	1.21	19	3.7	1.19			-0.11	0.32
17b	Cardoso	2011	20	3.56	1.1	19	3.7	1.19			-0.12	0.32
23a	Atilola	2014	20	0.93	0.25	20	0.829	0.27			0.38	0.32
23b	Atilola	2014	20	1.115	0.32	20	0.829	0.27			0.95	0.33
23c	Atilola	2014	20	1.07	0.21	20	0.829	0.27			0.98	0.33
25a	Linsey	2015	15	0.87	0.42	15	0.86	0.37			0.02	0.37
25b	Linsey	2015	15	1.14	0.53	15	0.86	0.37			0.6	0.37
25c	Linsey	2015	15	0.8	0.31	15	0.86	0.37			-0.17	0.37
26a	Chan [169]	2013	24	0.475	0.142	24	0.45	0.097			0.2	0.29
26b	Chan	2013	24	0.38	0.114	24	0.45	0.097			-0.65	0.3

Table D-3: Standardized mean difference and within - study errors for metric ‘novelty’

ID	Author	Year	Treatment group			Control group			t	F	SMD'	SE
			N1	Mean1	SD1	N2	Mean2	SD2				
12a	Smith,S	1993	47	NA	NA	47	NA	NA		8.13	-0.58	0.21
12b	Smith,S	1993	47	NA	NA	47	NA	NA		12.5	-0.72	0.21
12c	Smith,S	1993	66	NA	NA	25	NA	NA			-0.64	0.24
12d	Smith,S	1993	66	NA	NA	25	NA	NA		12.53	-0.82	0.24
12e	Smith,S	1993	20	NA	NA	20	NA	NA	1.56		-0.48	0.32
12f	Smith,S	1993	20	NA	NA	20	NA	NA	2.89		-0.9	0.33
13a	Shorachi	2015	15	2.4	0.32	15	2.26	0.47			0.34	0.37
13b	Shorachi	2015	15	2.55	0.44	15	2.26	0.47			0.62	0.37
15a	Cardoso	2012	20	2.86	0.23	18	2.62	0.23			1.02	0.35
15b	Cardoso	2012	20	2.71	0.37	18	2.62	0.23			0.28	0.33
16a	Agogue	2013	27	0.43	0.27	78	0.569	0.134			-0.77	0.23
16b	Agogue	2013	26	0.648	0.126	78	0.569	0.134			0.59	0.23
17a	Cardoso	2011	21	NA	NA	19	NA	NA			-0.63	0.32
17b	Cardoso	2011	20	NA	NA	19	NA	NA			-1.12	0.34
18a	Chan	2011	64	0.91	0.098	24	0.88	0.1			0.3	0.24
18b	Chan	2011	63	0.86	0.113	24	0.88	0.1			-0.18	0.24
20a.	Goncalves	2012	19	14.06	4.3	18	15.32	3.73			-0.31	0.33
20b.	Goncalves	2012	20	12.1	3.15	18	15.32	3.73			-0.92	0.34
20c.	Goncalves	2012	19	13.76	4.41	18	15.32	3.73			-0.37	0.33
22a	Siangliulue	2015	19	0.176	0.22	25	-0.177	0.48			0.89	0.32
22b	Siangliulue	2015	28	-0.013	0.24	25	-0.177	0.48			0.43	0.28
22c	Siangliulue	2015	25	0.051	0.209	25	-0.177	0.48			0.61	0.29
24a	Wilson [88]	2010	9	0.76	0.24	8	0.41	0.33			1.16	0.53
24b	Wilson	2010	9	0.94	0.11	8	0.41	0.33			2.1	0.61
25a	Linsey	2015	15	0.5	0.33	15	0.56	0.26			-0.2	0.37
25b	Linsey	2015	15	0.64	0.135	15	0.56	0.26			0.38	0.37
25c	Linsey	2015	15	0.56	0.25	15	0.56	0.26			0	0.37
26a	Chan	2013	24	0.927	0.042	24	0.933	0.044			-0.14	0.29
26b	Chan	2013	24	0.88	0.077	24	0.933	0.044			-0.83	0.3
27a	Tseng	2008	17	0.887	0.016	18	0.866	0.028			0.89	0.35
27b	Tseng	2008	18	0.858	0.03	18	0.866	0.028			-0.27	0.33
27c	Tseng	2008	18	0.883	0.024	18	0.866	0.028			0.64	0.34

Table D-4: Problem size as used in meta-regression analysis. NFR represents a non – functional requirement and FR represents a functional requirement

ID	Problem name	Problem statement	FR	NFR	Size
9, 23, 25	Peanut shelling machine	Design and build a <u>low-cost <NFR1>, easy to manufacture <NFR2> peanut shelling machine <FR1></u> that will <u>increase the productivity <NFR3></u> of the African peanut farmers. Target throughput is approximately 50 Kg per hour. The goals include: a. <u>Must remove the shell with minimal damage <NFR4></u> to peanuts b. <u>Electrical outlets are not available <NFR5></u> as a power source c. A large quantity of peanuts must be quickly shelled <u><FR2></u> .	2	5	7
12a	New toy	Imagine that you are employed by a toy company that is in need of new ideas for toys. Your task is to design some new <u>toys <FR1></u> for the company. Within the allotted 20 minutes draw as many new and different toys of your own creative design as you are able. <u>Duplication of toys that currently exist or have already existed is not permitted. <NFR1></u>	1	1	2
12b	Alien creature	Imagine a planet just like Earth existing somewhere in the universe. It is currently uninhabited. Your task is to design new creatures to <u>inhabit the planet. <FR1></u> Within the allotted 20 minutes draw as many new and different creatures of your own creative design as you are able. <u>Duplication of creatures now extinct or living on the planet Earth is not permitted. <NFR1></u>	1	1	2
13	Pseudo alarm clock	People generally use their mobile phone or a traditional alarm clock in order to wake themselves up every morning. Yet, they are not always effective and can sometimes cause oversleeping. What would be other possible ways to <u>wake up in the morning <FR1></u> in a certain time <u>without using any form of alarm clock? <NFR1></u> Generate as many ideas as possible.	1	1	2
14	Pumping unit	Design a pumping unit <u>to extract petroleum <FR1></u>	1	0	1
15, 20	Future transport -ation solution	Since ancient times, transportation of people and goods has always been an essential human activity. Despite the rapid technological developments in the field of human transportation, it is still uncertain how this area will unfold in the future. Your task is to think about how <u>human transportation will be like in 2050. <FR1></u> You are kindly asked to draw as many different ideas as you can in 45 minutes	1	0	1
16	Hen's egg	Ensure that a hen's <u>egg dropped from a height of 10 m does not break. <FR1></u>	1	0	1

ID	Problem name	Problem statement	FR	NFR	Size
17	Device to pick books	Design a device that allows people <u>to pick up a book from a shelf</u> <FR1> (e.g. in a library) that is out of their reach, for instance, above their head. It should be: 1. <u>Ease of use</u> — solution ideas' <NFR1> level of usability in terms of how easy/comfortable would be for an user to retrieve a book using it. 2. <u>Manufacture — feasibility</u> <NFR2> of building such a device using existing technologies. 3. <u>Minimized book damage</u> — <NFR3> likelihood of a book being damaged during retrieval.	1	3	4
18, 26	Device to harness human power	Design a device to <u>collect energy from human motion</u> <FR1> for use in developing and impoverished rural communities in places like India and many African countries. Our goal is to build a <u>low-cost</u> , <NFR1> <u>easy to manufacture device</u> <NFR2> targeted at individuals and small households to <u>provide energy to be stored in a rechargeable battery</u> <FR2> with approximately 80% efficiency. The energy is intended to be used by small, low power draw electrical devices, such as a radio or lighting device, hopefully leading to an increase in the quality of life of the communities by increasing productivity, connection to the outside world, etc. The target energy production is 1 kW-h per day, roughly enough to power eight 25W compact florescent light bulbs for 5 h each per day, or enough to power a CB radio for the entire day. For reference, an average adult human can output about 200W with full body physical activity for short periods of time, with a significant reduction for sustained power output.	2	2	4
19	Automatic watering device	Watering of house-plants is an easy task. However, when people leave on holiday or business, this task is often left to other persons. Your assignment is to generate as many different ideas as possible for an <u>automatic watering device</u> <FR1> for house-plants. The <u>device should provide a plant with about a deciliter of water</u> <FR2> per week - no more or less. The <u>device should be able to water the plant for a minimum</u> <FR3> of one month.	3	0	3
21a	Bike rack	Suppose you are asked to construct a new <u>bike rack for cars</u> . <FR1> You should construct as many designs as possible, write comments with each design, and number each individual design. There are no constraints in the materials you may want to use. The problems to be addressed are: 1. <u>Easy mounting of the bicycle</u> <NFR1> 2. <u>Easy mounting of the rack</u> <NFR2> 3. <u>Cannot harm bike or car</u> <NFR3> 4. <u>Must be versatile for all bikes and cars</u> <NFR4>	1	4	5

ID	Problem name	Problem statement	FR	NFR	Size
21c	Coffee cup	Suppose you are asked to construct an <u>inexpensive</u> , <NFR1> <u>disposable</u> , <NFR2> spill-proof <u>coffee cup</u> . <FR1> You should construct as many designs as possible, write comments with each design, and number each individual design. There are no constraints in the materials you may want to use. The problems to be addressed are: 1. Leaking of the cup if it tips over <FR2> 2. Leaking of the cup when squeezed <FR3> 3. Hot liquid burning the user's mouth <FR4>	4	2	6
22	Fabric display	Generate product ideas for an imaginary technology—a <u>touch-sensitive</u> , <FR1> “fabric display” that could <u>render high resolution images</u> <FR2> and <u>videos on any fabric</u> <FR3> through a <u>penny-sized connector</u> . <NFR1>	3	1	4
24	Device to immobilize joint	Due to potential leg injuries, MTREK is now requiring guides to carry additional supplies to treat leg and ankle injuries. In the design challenge, MTREK has hired you to design a device that can be used to <u>immobilize a joint or limb in case of an extreme injury</u> . <FR1> This device must be (1) <u>as light</u> <NFR1> and <u>small</u> <NFR2> as possible when stored in the guide's packs but (2) <u>rigid</u> <NFR3> and <u>large</u> <NFR4> enough to immobilize the leg of an average sized male.	1	4	5
27	Alternative clock	The clock is one of the oldest human inventions, requiring a physical process that will proceed at a known rate and a way to gauge how long that process has run. As the seasons and the phases of the moon can be used to measure the passage of longer periods of time, shorter processes had to be used to measure off hours, minutes, and seconds. You need to come up with as many of these <u>shorter processes to measure the passage of hours</u> , <FR1> minutes, and seconds as you can in ten minutes. The time measurement does not have to be in any known unit so long as it is repeatable <FR2> so that you can repeat it with a clock at a later time. You are alone in a large featureless room with no windows, a door with doorknob, a hanging light fixture on the 10-foot ceiling, and a sink and drain with working tap. Please draw or describe the concept of your solutions in order in the boxes provided and mark the time as projected by the laptop in the front of the classroom to the second (hh:mm:ss) in the space provided when you finish each solution.	2	0	2

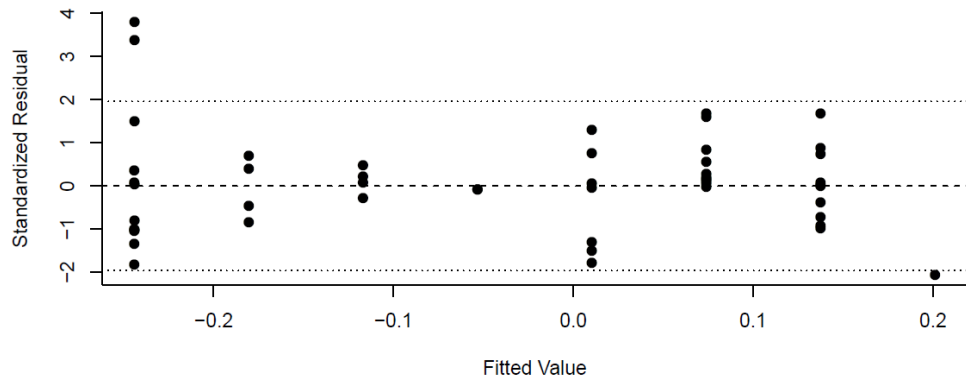


Figure D-1: Standardized residual vs predicted value of effect size for metric 'quantity' for design problem regression model

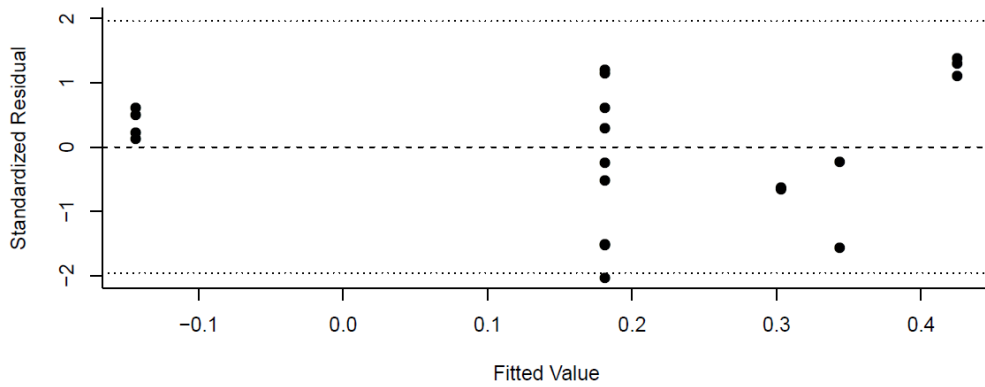


Figure D-2: Standardized residual vs predicted value of effect size for metric 'quality' for design problem regression model

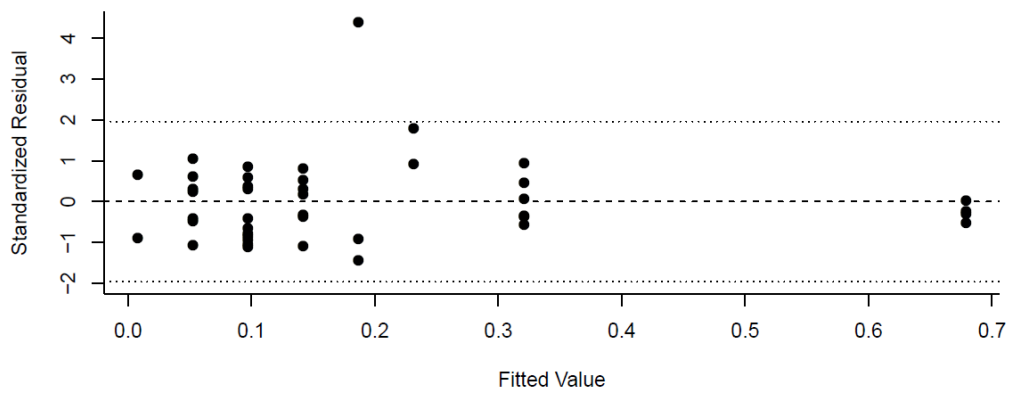


Figure D-3: Standardized residual vs predicted value of effect size for metric 'novelty' for design problem regression model

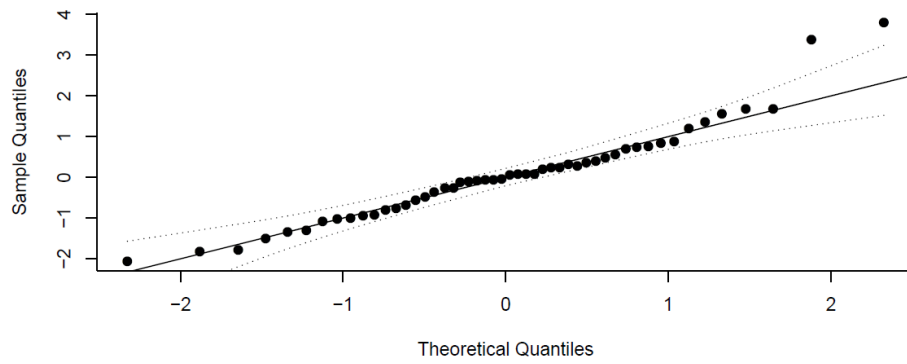


Figure D-4: Normal Q-Q plot for metric 'quantity' for design problem regression model

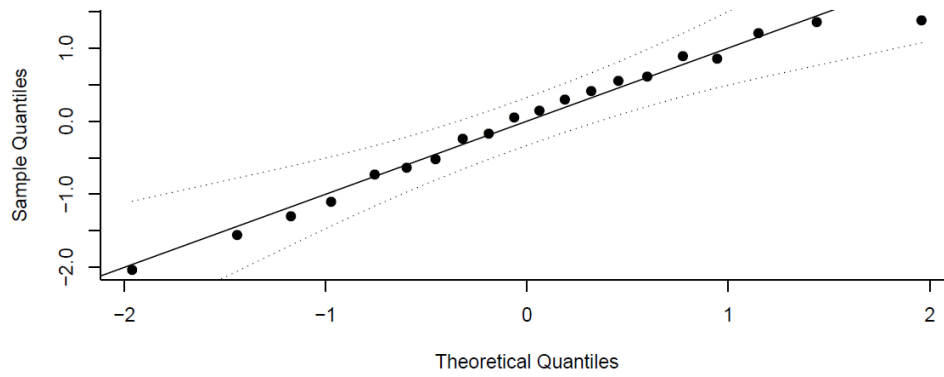


Figure D-5: Normal Q-Q plot for metric 'quality' for design problem regression model

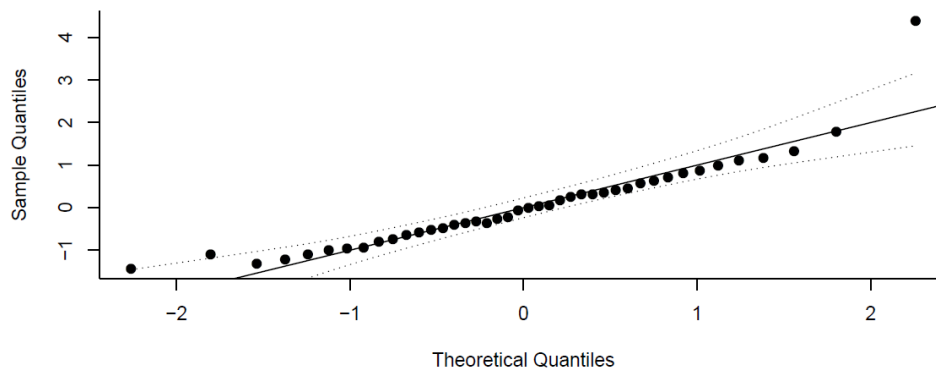


Figure D-6: Normal Q-Q plot for metric 'novelty' for design problem regression model

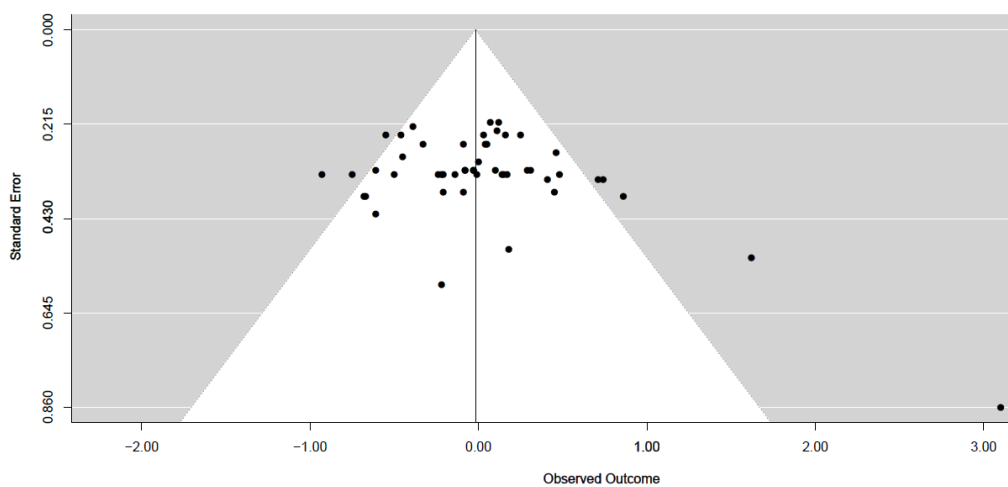


Figure D-7: Funnel plot for metric 'quantity'

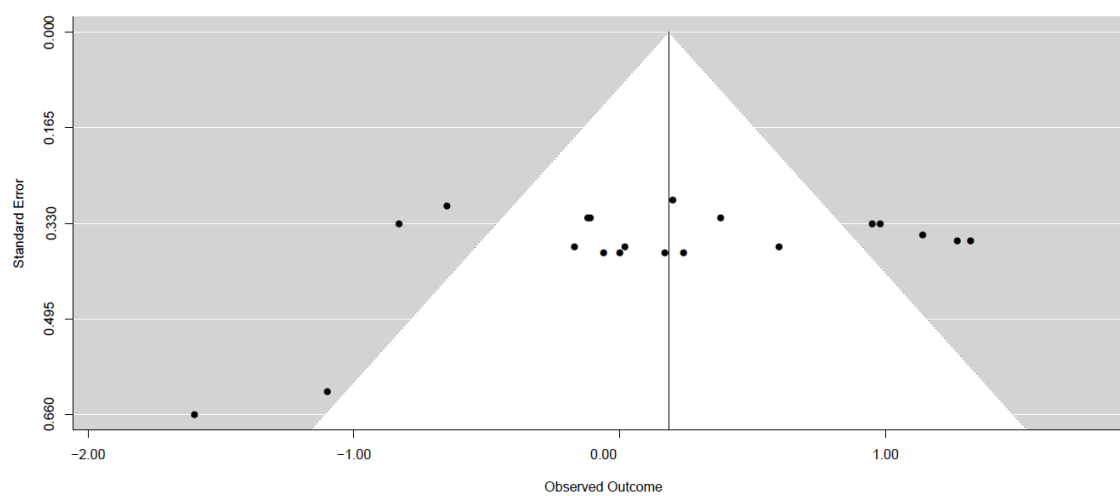


Figure D-8: Funnel plot for metric 'quality'

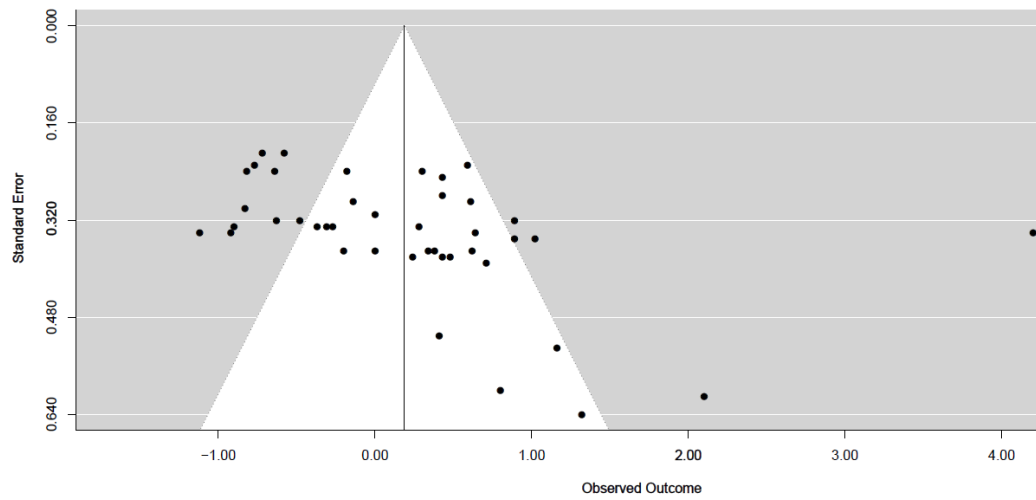


Figure D-9: Funnel plot for metric ‘novelty’

Table D-5: Results of asymmetry test of funnel plots

	Metric		
	Quantity	Quality	Novelty
t - value	1.79	-1.14	2.77
Degree of freedom	45	13	32
p value	0.08	0.27	<0.01

REFERENCES

- [1] Cross, N., 2008, "The nature of design," *Engineering design methods: strategies for product design*, John Wiley & Sons, Chippenham, Wiltshire, pp. 11–12.
- [2] Cross, N., and Roy, R., 1989, "Nature of design," *Engineering design methods*, Wiley New York, pp. 11–12.
- [3] Hull, E., Jackson, K., and Dick, J., 2011, "Introduction," *Requirements engineering*, Springer London, London, pp. 2–3.
- [4] Lamar, C., 2009, "Linguistic Analysis of Natural Language Engineering Requirements."
- [5] Oxman, R., 2004, "Think-maps: teaching design thinking in design education," *Des. Stud.*, **25**(1), pp. 63–91.
- [6] Lamar, C., and Mocko, G., 2010, "Linguistic analysis of natural language engineering requirement statements," 8th International Symposium on Tools and Methods of Competitive Engineering, TMCE 2010.
- [7] Carroll, J. M., Thomas, J. C., and Malhotra, A., 1980, "Presentation and representation in design problem-solving," *Br. J. Psychol.*, **71**(1), pp. 143–153.
- [8] Simon, H. A., 1977, "The structure of ill-structured problems," *Models of discovery*, Springer, pp. 304–325.
- [9] Dorst, K., 2003, "The Problem of Design Problems," *Design*, **4**(Cross), pp. 135–147.
- [10] Dorst, C. H., 2006, "Design Problems and Design Paradoxes," *Des. Issues*, **22**(3), pp. 4–17.
- [11] Dorst, K., and Cross, N., 2001, "Creativity in the design process: Co-evolution of problem-solution," *Des. Stud.*, **22**(5), pp. 425–437.
- [12] Dinar, M., and Langley, P., 2011, "Towards a Formal Representation Model of Problem Formulation in Design," *Proc. ASME 2011 Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, pp. 1–10.
- [13] Maher, M. Lou, 2000, "A model of co-evolutionary design," *Eng. Comput.*, **16**(3-4), pp. 195–208.
- [14] Lawson, B. R., 1979, "Cognitive Strategies in Architectural Design," *Ergonomics*, **22**(1), pp. 59–68.
- [15] Durand, F., Helms, M. E., Tsenn, J., McAdams, D. A., and Linsey, J. S., 2015, "In search of effective design problems for design research," *IDETC & CIE*, American Society of Mechanical Engineers, Boston, pp. 1–10.
- [16] Johnson, D. G., Genco, N., Saunders, M. N., Williams, P., Seepersad, C. C., and Hölttä-Otto, K., 2014, "An Experimental Investigation of the Effectiveness of Empathic Experience Design for Innovative Concept Generation," *J. Mech. Des.*, **136**(May 2014), p. 051009.

- [17] Pahl, G., and Beitz, W., 1996, "Conceptual Design," *Engineering design: a systematic approach*, K. Wallace, ed., Springer London, London, p. 147.
- [18] Pahl, G., and Beitz, W., 2013, *Engineering design: a systematic approach*, Springer Science & Business Media.
- [19] Dieter, G. E., and Schmidt, L. C., 2013, "Problem definition and need identification," *Engineering design*, McGraw-Hill New York, New York, pp. 111–112.
- [20] Brunetti, G., and Golob, B., 2000, "A feature-based approach towards an integrated product model including conceptual design information," *Comput. Des.*, **32**(14), pp. 877–887.
- [21] Schmidt, L. A., 2010, "Crowdsourcing for Human Subjects Research."
- [22] Hernandez, N. V., Shah, J. J., and Smith, S. M., 2010, "Understanding design ideation mechanisms through multilevel aligned empirical studies," *Des. Stud.*, **31**(4), pp. 382–410.
- [23] Daly, S., and Christian, J., 2012, "Assessing design heuristics for idea generation in an introductory engineering course," *Int. J. Eng. Educ.*, **Vol. 28**(2), pp. 1–11.
- [24] Chulvi, V., Sonseca, A., Mulet, E., and Chakrabarti, A., 2012, "Assessment of the Relationships Among Design Methods, Design Activities, and Creativity," *J. Mech. Des.*, **134**(11), p. 111004.
- [25] Hernandez, N. V., Schmidt, L. C., and Okudan, G. E., 2013, "Systematic Ideation Effectiveness Study of TRIZ," *J. Mech. Des.*, **135**(10), p. 101009.
- [26] Worinkeng, E., Joshi, S., and Summers, J. D., 2015, "An experimental study: analyzing requirement type influence on novelty and variety of generated solutions," *Int. J. Des. Creat. Innov.*, **3**(2), pp. 61–77.
- [27] Hannah, R., Joshi, S., and Summers, J. D., 2012, "A user study of interpretability of engineering design representations," *J. Eng. Des.*, **23**(6), pp. 443–468.
- [28] Patel, A., Kramer, W., Summers, J. D., and Shuffler-Porter, M., 2016, "Function Modeling: A Study of Model Sequential Completion Based on Count and Chaining of Functions," *International Design Engineering Conferences and Computers in Engineering Conference (ASME IDETC/CIE, ASME-AMER SOC MECHANICAL ENG, Charlotte, NC*, pp. DETC2016–59860.
- [29] Eastman, C. M., 1969, "Cognitive processes and ill-defined problems: A case study from design," *Proc. Int. Jt. Conf. Artif. Intell. IJCAI*, **69**, pp. 669–690.
- [30] Rittel, H. W. J., and Webber, M. M., 1973, "Dilemmas in a General Theory of Planning," *Policy Sci.*, **4**(December 1969), pp. 155–169.

- [31] Rush, M., Wallace, D., and Newman, D., 2008, "Creative thinking in a first year mechanical engineering design course at the massachusetts institute of technology: a community of practice model," pp. 1–7.
- [32] Dixon, J. R., Duffey, M. R., Irani, R. K., Meunier, K. L., Orelup, M. F., 1988, "A Proposed Taxonomy of Mechanical Design Problems," *Comput. Eng.*, **1**, pp. 41–46.
- [33] Ullman, D. G., 1992, "A taxonomy for mechanical design," *Res. Eng. Des.*, **3**(3), pp. 179–189.
- [34] Goel, V., and Pirolli, P., 1989, "Motivating the notion of generic design within information-processing theory: the design problem space," *AI Mag.*, **10**(1), pp. 18–36.
- [35] Goel, V., and Pirolli, P., 1992, "The structure of design problem spaces," *Cogn. Sci.*, **16**(3), pp. 395–429.
- [36] Frost, R. B., 1994, "A Suggested Taxonomy for Engineering Design Problems," *J. Eng. Des.*, **5**(4), pp. 399–410.
- [37] Summers, J. D., and Shah, J. J., 2010, "Mechanical Engineering Design Complexity Metrics: Size, Coupling, and Solvability," *J. Mech. Des.*, **132**(2), p. 021004.
- [38] Thoe, S., and Summers, J. D., 2013, "Correlating Problem / Process Exam Question Complexity To Anticipated Effort : a Modeling Protocol," *ASME Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, pp. 1–9.
- [39] Wilson, T. D., 1981, "On user studies and information needs," *J. Doc.*, **37**(1), pp. 3–15.
- [40] Siatri, R., 1999, "The evolution of user studies," *Libri*, **49**(3), pp. 132–141.
- [41] Oman, S., Gilchrist, B., Rebhuhn, C., Tumer, I. Y., and Nix, A., 2014, "TOWARDS A REPOSITORY OF INNOVATIVE PRODUCTS TO ENHANCE ENGINEERING CREATIVITY EDUCATION," pp. 1–12.
- [42] Linsey, J. S., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L., and Markman, a. B., 2011, "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods," *J. Mech. Des.*, **133**(3), p. 031008.
- [43] Kurtoglu, T., Campbell, M. I., and Linsey, J. S., 2009, "An experimental study on the effects of a computational design tool on concept generation," *Des. Stud.*, **30**(6), pp. 676–703.
- [44] Kim, M. S., 2007, "Analysis of team interaction and team creativity of student design teams based on personal creativity modes," *Proc. ASME 2007 Int. Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, pp. 1–13.
- [45] Kim, Y., Kim, M., and Jin, S., 2005, "Cognitive characteristics and design creativity : An experimental study," *Des. Eng.*, pp. 1–9.

- [46] Linsey, J. S., Green, M. G., Murphy, J. T., Wood, K. L., and Markman, A. B., 2005, "Collaborating to Success: An Experimental Study of Group Idea Generation Techniques," ASME Design Theory and Methodology Conference, p. 14.
- [47] Yilmaz, S., and Christian, J., 2011, "Collaborative idea generation using design heuristics," DS 68-10 ..., (August), pp. 1–11.
- [48] Rogers, D., and Salustri, F. A., 2014, "Comparative evaluation of a novel concept design method," pp. 1–10.
- [49] Shalley, C. E., 1995, "Effects of Coaction, Expected Evaluation, and Goal Setting on Creativity and Productivity.," *Acad. Manag. J.*, **38**(2), pp. 483–503.
- [50] Glier, M., and Schmidt, S., 2011, "Distributed ideation: Idea generation in distributed capstone engineering design teams," *Int. J. Eng. Educ.*, **27**(6), pp. 1281–1294.
- [51] Parmee, I. C., and Bonham, C. R., 2000, "Towards the support of innovative conceptual design through interactive designer/evolutionary computing strategies," *Ai Edam*, **14**(01), pp. 3–16.
- [52] Toh, C. a., and Miller, S. R., 2013, "Visual Inspection or Product Dissection? The Impact of Designer-Product Interactions on Engineering Design Creativity," Vol. 5 25th Int. Conf. Des. Theory Methodol. ASME 2013 Power Transm. Gearing Conf., p. V005T06A011.
- [53] Sternberg, R. J., 1999, "Experimental studies of creativity," *Handbook of creativity*, R.J. Sternberg, ed., Cambridge University Press, Cambridge, pp. 62–63.
- [54] Frey, D. D., and Dym, C. L., 2006, "Validation of design methods: lessons from medicine," *Res. Eng. Des.*, **17**(1), pp. 45–57.
- [55] Blessing, L. T. M., and Chakrabarti, A., 2009, *DRM: A Design Reseach Methodology*, Springer.
- [56] Osborn, A., 2013, "Creative collaboration by teams," *Applied Imagination-Principles and Procedures of Creative Writing*, Read Books Ltd, pp. 110–115.
- [57] Lamm, H., and Trommsdorff, G., 1973, "Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review," *Eur. J. Soc. Psychol.*, **3**(4), pp. 361–388.
- [58] Lang, A., 1996, "The Logic of Using Inferential Statistics with Experimental Data from," *J. Broadcast. Electron. Media*, **40**, pp. 422–430.
- [59] Meltzer, C. E., Naab, T., and Daschmann, G., 2012, "All Student Samples Differ: On Participant Selection in Communication Science," *Commun. Methods Meas.*, **6**(April), pp. 251–262.
- [60] Blessing, L. T. M., and Chakrabarti, A., 2009, "Descriptive Study Methods," *DRM: A Design Reseach Methodology*, Springer, London, pp. 260–262.

- [61] Takai, S., Midha, A., and Esterman, M., 2012, "An approach toward developing metrics to predict performance and creativity of final products in project-based design class," ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 103–110.
- [62] Shah, J. J., Millsap, R. E., Woodward, J., and Smith, S. M., 2012, "Applied Tests of Design Skills—Part 1: Divergent Thinking," *J. Mech. Des.*, **134**(2), p. 021005.
- [63] Oman, S., and Tumer, I., 2010, "Assessing creativity and innovation at the concept generation stage in engineering design: a classroom experiment," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 1–9.
- [64] Rogers, D., and Salustri, F. A., 2013, "Comparative evaluation of a novel concept design method," ASME 2013 International Mechanical Engineering Congress and Exposition, pp. V012T13A028–V012T13A028.
- [65] Woodman, R. W., Sawyer, J. E., and Griffin, R. W., 1993, "TOWARD A THEORY OF CREATIVITY," **18**(2), pp. 293–321.
- [66] Walton, A. P., 2003, "The impact of interpersonal factors on creativity," *Int. J. Entrep. Behav. Res.*, **9**(4), pp. 146–162.
- [67] English, K., Naim, A., Lewis, K., Schmidt, S., Viswanathan, V., Linsey, J., McAdams, D. a., Bishop, B., Campbell, M. I., Poppa, K., Stone, R. B., and Orsborn, S., 2010, "Impacting Designer Creativity Through IT-Enabled Concept Generation," *J. Comput. Inf. Sci. Eng.*, **10**(3), p. 031007.
- [68] Toh, C. A., and Miller, S. R., 2013, "Exploring the utility of product dissection for early-phase idea generation," *Proc. ASME Design Engineering Technical Conferences*, August, pp. 4–7.
- [69] Linsey, J. S., Markman, a. B., and Wood, K. L., 2012, "Design by Analogy: A Study of the WordTree Method for Problem Re-Representation," *J. Mech. Des.*, **134**(4), p. 041009.
- [70] Acuna, A., and Sosa, R., 2011, "The complementary role of representations in Design creativity: sketches and models," *Des. Creat.* 2010.
- [71] Schmidt, L. C., and Hernandez, N. V., 2014, "PILOT OF SYSTEMATIC IDEATION STUDY WITH LESSONS LEARNED," pp. 1–9.
- [72] Weaver, J., and Kuhr, R., 2009, "Increasing innovation in multi-function systems: evaluation and experimentation of two ideation methods for design," ASME 2009 ..., pp. 1–19.
- [73] Doboli, A., and Umbarkar, A., 2014, "The role of precedents in increasing creativity during iterative design of electronic embedded systems," *Des. Stud.*, **35**(3), pp. 298–326.
- [74] Kudrowitz, B., Te, P., and Wallace, D., 2012, "The influence of sketch quality on perception of product-idea creativity," *Artif. Intell. Eng. Des. Anal. Manuf.*, **26**(03), pp. 267–279.

- [75] Lopez, R., Linsey, J. S., and Smith, S. M., 2011, "Characterizing the effect of domain distance in design-by-analogy," ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 141–151.
- [76] Okudan, G. E., Schmidt, L. C., Hernandez, N. V., Jablokow, K., and Lin, C., 2012, "Questioning the Impact of Personality on Design Outcomes: Should We Take It Into Account?," ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 95–102.
- [77] Jin, Y., and Benami, O., 2010, "Creative patterns and stimulation in conceptual design," *Artif. Intell. Eng. Des. Anal. Manuf.*, **24**(02), pp. 191–209.
- [78] Cheong, H., Chiu, I., and Shu, L., 2010, "Extraction and transfer of biological analogies for creative concept generation," ASME 2010 ..., pp. 1–16.
- [79] Linsey, J. S., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L., and Markman, a. B., 2011, "An Experimental Study of Group Idea Generation Techniques: Understanding the Roles of Idea Representation and Viewing Methods," *J. Mech. Des.*, **133**(3), p. 031008.
- [80] Tsenn, J., Atilola, O., McAdams, D. a., and Linsey, J. S., 2014, "The effects of time and incubation on design concept generation," *Des. Stud.*, **35**(5), pp. 500–526.
- [81] Toh, C. a., Miller, S. R., and Okudan Kremer, G. E., 2014, "The Impact of Team-Based Product Dissection on Design Novelty," *J. Mech. Des.*, **136**(4), p. 041004.
- [82] Al-Shorachi, E., Sasasmit, K., and Gonçalves, M., 2015, "Creativity intervention: Using storytelling and math problems as intervening tasks for inducing incubation," *Int. Conf. Eng. Des. ICED15*, (July), pp. 1–10.
- [83] Zhang, L., 2011, "Design fixation and solution quality under exposure to example solution," *International Conference on Computing, Control and Industrial Engineering*, pp. 129–132.
- [84] Cardoso, C., and Gonçalves, M., 2012, "Searching for inspiration during idea generation: pictures or words?," *International Design Conference, The design society, Dubrovnik, Croatia*, pp. 1–10.
- [85] Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K., 2011, "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples," *J. Mech. Des.*, **133**(8), p. 081004.
- [86] Guerreiro Goncalves, M., Cardoso, C., and Badke-Schaub, P., 2012, "Find your inspiration: exploring different levels of abstraction in textual stimuli," *International conference on design creativity, The design society, Glasgow, UK*, pp. 1–8.

- [87] Fu, K., Chan, J., Cagan, J., and Kotovsky, K., 2013, "The Meaning of 'Near' and 'Far': The Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output," *J. Mech. Des.*, **135**(February), pp. 1–12.
- [88] Wilson, J. O., Rosen, D., Nelson, B. a., and Yen, J., 2010, "The effects of biological examples in idea generation," *Des. Stud.*, **31**(2), pp. 169–186.
- [89] Viswanathan, V., and Linsey, J., 2011, "Understanding fixation: A study on the role of expertise," *ICED 11 - 18th Int. Conf. Eng. Des. - Impacting Soc. Through Eng. Des.*, **7**(August), pp. 309–319.
- [90] McKoy, F. L., Vargas-Hernández, N., Summers, J. D., and Shah, J. J., 2001, "Influence of Design Representation on Effectiveness of Idea Generation," *Des. Eng. Tech. Conf. Comput. Inf. Eng. Conf.*, pp. 1–10.
- [91] Kumar, V., and Mocko, G., 2016, "Similarity of engineering design problems to enable reuse in design research experiments," *IDETC & CIE*, American Society of Mechanical Engineers, Charlotte, NC, p. (in print).
- [92] Pahl, G., Beitz, W., Feldhusen, J., and Grote, K.-H., 1996, "Suggestions for using this book in design teaching," *Engineering design: a systematic approach*, K. Wallace, ed., Springer London, London, p. xxv.
- [93] Landauer, T. K., Foltz, P. W., and Laham, D., 1998, "An introduction to latent semantic analysis," *Discourse Process.*, **25**(2-3), pp. 259–284.
- [94] Lipsey, M. W., and Wilson, D. B., 2001, "Selecting, Computing, and Coding the Effect Size Statistic," *Practical meta-analysis*, Sage Publications, Inc, Thousand Oaks, CA, US, pp. 48–50.
- [95] Glass, G. V, MacGaw, B., and Smith, M. L., 1984, *Meta-analysis in social research*, Sage Beverly Hills, CA.
- [96] Walker, E., Hernandez, A. V., and Kattan, M. W., 2008, "Meta-analysis: Its strengths and limitations," *Cleve. Clin. J. Med.*, **75**(6), pp. 431–439.
- [97] Borenstein, M., Hedges, L. V, Higgins, J., and Rothstein, H. R., 2009, "Heterogeneity," *Introduction to meta analysis*, Wiley Online Library, Chichester, England, p. 106.
- [98] Kreyszig, E., 2010, "Graphs," *Advanced engineering mathematics*, John Wiley & Sons, pp. 954–960.
- [99] Newman, M., 2010, "Metrics and measures," *Networks: an introduction*, Oxford University Press, New York, pp. 168–200.
- [100] Fortunato, S., 2010, "Community detection in graphs," *Phys. Rep.*, **486**(3-5), pp. 75–174.

- [101] NASA, "Requirement engineering overview" [Online]. Available: http://www.hq.nasa.gov/office/codeq/software/ComplexElectronics/1_requirements2.htm. [Accessed: 02-Nov-2015].
- [102] Hull, E., Jackson, K., and Dick, J., 2010, "Defining requirement engineering," Requirements engineering, Springer Science & Business Media, London, pp. 6–8.
- [103] IEEE, 1990, "Standard Glossary of SoftwareEngineering Terminology," IEEE Softw. Eng. Stand. Collect. IEEE, pp. 610–612.
- [104] Summers, J. D., and Morkos, B., 2013, "Requirements Evolution: Impact of Functional and Non-functional Change on Project Success," International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, Portland, pp. 35–45.
- [105] Glinz, M., 2007, "On non-functional requirements," Requirements Engineering Conference, 2007. RE'07. 15th IEEE International, pp. 21–26.
- [106] Chung, L., and Nixon, B. A., 1995, "Dealing with non-functional requirements: three experimental studies of a process-oriented approach," Proceedings of the 17th international conference on Software engineering, pp. 25–37.
- [107] Mylopoulos, J., Chung, L., and Nixon, B., 1992, "Representing and using nonfunctional requirements: A process-oriented approach," Softw. Eng. IEEE Trans., **18**(6), pp. 483–497.
- [108] Chung, L., and do Prado Leite, J. C. S., 2009, "On non-functional requirements in software engineering," Conceptual modeling: Foundations and applications, Springer, pp. 363–379.
- [109] Worinkeng, E., 2013, "Analyzing Requirement Type Influence on Generated Solutions," Clemson.
- [110] Kotonya, G., and Sommerville, I., 1998, "Introduction," Requirements engineering: processes and techniques, J. Wiley, pp. 5–11.
- [111] Shankar, P., Morkos, B., and Summers, J. D., 2010, "A hierarchical modeling scheme with non functional requirements," ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 283–295.
- [112] Hirtz, J., Stone, R., McAdams, D., Szykman, S., and Wood, K., 2002, "A functional basis for engineering design: Reconciling and evolving previous efforts," Res. Eng. Des., **13**(2), pp. 65–82.
- [113] Krippendorff, K., 2011, "Computing Krippendorff' s Alpha Reliability," Dep. Pap., p. 12.
- [114] Krippendorff, K., 2012, Content analysis: An introduction to its methodology, Sage Publications, Inc.

- [115] Hayes, A. F., and Krippendorff, K., 2007, "Answering the call for a standard reliability measure for coding data," *Commun. Methods Meas.*, **1**(1), pp. 77–89.
- [116] Harispe, S., Ranwez, S., Janaqi, S., and Montmain, J., 2015, "Semantic similarity from natural language and ontology analysis," *Synth. Lect. Hum. Lang. Technol.*, **8**(1), pp. 1–254.
- [117] Gomaa, W. H., and Fahmy, A. A., 2013, "A survey of text similarity approaches," *Int. J. Comput. Appl.*, **68**(13).
- [118] Liu, H., and Wang, P., 2014, "Assessing text semantic similarity using ontology," *J. Softw.*, **9**(2), pp. 490–497.
- [119] Rajaraman, A., and Ullman, J. D., "Data Mining," *Mining of Massive Datasets*, Cambridge University Press, Cambridge, pp. 1–17.
- [120] Bao, J., Lyon, C., Lane, P. C. R., Ji, W., and Malcolm, J., 2007, "Comparing different text similarity methods."
- [121] University of Colorado, "Latent semantic analysis" [Online]. Available: <http://lsa.colorado.edu/>. [Accessed: 17-May-2016].
- [122] Dumais, S. T., 2004, "Latent semantic analysis," *Annu. Rev. Inf. Sci. Technol.*, **38**(1), pp. 188–230.
- [123] Martin, D. I., and Berry, M. W., 2007, "Mathematical Foundations Behind Latent Semantic Analysis," *Handb. Latent Semant. Anal.*, pp. 35–55.
- [124] Stanley, T. D., and Jarrell, S. B., 1989, "Meta-Regression analysis: A quantitative method of literature surveys," *J. Econ. Surv.*, **3**(2), pp. 161–170.
- [125] Johnson, E. S., Lanes, S. F., Wentworth, C. E., Satterfield, M. H., Abebe, B. L., and Dicker, L. W., 1999, "A meta regression analysis of dose response effect of aspirin on stroke," *Arch. Intern. Med.*, **159**(11), pp. 1248–1253.
- [126] Hey, J., Linsey, J., Agogino, A. M., and Wood, K. L., 2008, "Analogies and metaphors in creative design," *Int. J. Eng. Educ.*, **24**(2), p. 283.
- [127] Goel, A. K., 1997, "Design, analogy, and creativity," *IEEE Expert*, **12**(3), pp. 62–70.
- [128] Atilola, O., and Linsey, J., 2015, "Representing analogies to influence fixation and creativity: A study comparing computer-aided design, photographs, and sketches," *Artif. Intell. Eng. Des. Anal. Manuf.*, **29**(02), pp. 161–171.
- [129] Jansson, D. G., and Smith, S. M., 1991, "Design fixation," *Des. Stud.*, **12**(1), pp. 3–11.
- [130] Smith, S. M., Ward, T. B., and Schumacher, J. S., 1993, "Constraining effects of examples in a creative generation task," *Mem. Cognit.*, **21**(6), pp. 837–845.

- [131] Cooper, H., Hedges, L. V, and Valentine, J. C., 2009, *The handbook of research synthesis and meta-analysis*, Russell Sage Foundation.
- [132] Russo, M. W., 2007, "How to Review a Meta-analysis," *Gastroenterol. Hepatol. (N. Y.)*, **3**(8), pp. 637–642.
- [133] Lipsey, M. W., and Wilson, D. B., 2001, *Practical meta-analysis.*, Sage Publications, Inc, Thousand Oaks, CA, US.
- [134] Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., and Marín-Martínez, F., 2015, "A comparison of procedures to test for moderators in mixed-effects meta-regression models," *Psychol. Methods*, **20**(3), pp. 360–74.
- [135] Hedges, L. V., and Vevea, J. L., 1998, "Fixed- and random-effects models in meta-analysis," *Psychol. Methods*, **3**(4), pp. 486–504.
- [136] DerSimonian, R., and Kacker, R., 2007, "Random-effects model for meta-analysis of clinical trials: an update," *Contemp. Clin. Trials*, **28**(2), pp. 105–114.
- [137] Thompson, S. G., and Sharp, S. J., 1999, "Explaining heterogeneity in meta-analysis: A comparison of methods," *Stat. Med.*, **18**(20), pp. 2693–2708.
- [138] Higgins, J., and Thompson, S. G., 2002, "Quantifying heterogeneity in a meta-analysis," *Stat. Med.*, **21**(11), pp. 1539–1558.
- [139] Thompson, S. G., 1994, "Why sources of heterogeneity in meta-analysis should be investigated," *BMJ*, **309**(6965), pp. 1351–5.
- [140] Thompson, S. G., and Pocock, S. J., 1991, "Can meta-analyses be trusted?," *Lancet*, **338**(8775), pp. 1127–1130.
- [141] Berlin, J. A., Laird, N. M., Sacks, H. S., and Chalmers, T. C., 1989, "A comparison of statistical methods for combining event rates from clinical trials," *Stat. Med.*, **8**(2), pp. 141–151.
- [142] Borenstein, M., Hedges, L. V, Higgins, J., and Rothstein, H. R., 2009, *Meta regression*, Wiley Online Library, Chichester, England.
- [143] Baron, R. M., and Kenny, D. a, 1986, "The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations," *J. Pers. Soc. Psychol.*, **51**(6), pp. 1173–1182.
- [144] Raudenbush, S. W., 2009, "Analyzing effect sizes: Random-effects models," *Handb. Res. Synth. meta-analysis*, **2**, pp. 295–316.
- [145] Thompson, S. G., and Higgins, J. P. T., 2002, "How should meta-regression analyses be undertaken and interpreted?," *Stat. Med.*, **21**(11), pp. 1559–1573.

- [146] Berkey, C. S., Hoaglin, D. C., Mosteller, F., and Colditz, G. A., 1995, "A random-effects regression model for meta-analysis," *Stat. Med.*, **14**(4), pp. 395–411.
- [147] Hocevar, D., and Bachelor, P., 1989, "A taxonomy and critique of measurements used in the study of creativity," *Handbook of creativity*, Springer, pp. 53–75.
- [148] Oman, S. K., Tumer, I. Y., Wood, K., and Seepersad, C., 2012, "A comparison of creativity and innovation metrics and sample validation through in-class design projects," *Res. Eng. Des.*, **24**(1), pp. 65–92.
- [149] Shah, J. J., Smith, S. M., and Vargas-Hernandez, N., 2003, "Metrics for measuring ideation effectiveness," *Des. Stud.*, **24**(2), pp. 111–134.
- [150] Agogu , M., Kazak i, A., Hatchuel, A., Masson, P., Weil, B., Poir el, N., and Cassotti, M., 2014, "The impact of type of examples on originality: Explaining fixation and stimulation effects," *J. Creat. Behav.*, **48**(1), pp. 1–12.
- [151] Cardoso, C., and Badke-Schaub, P., 2011, "The influence of different pictorial representations during idea generation," *J. Creat. Behav.*, **45**(2), pp. 130–146.
- [152] Chan, J., Fu, K., Schunn, C., Cagan, J., Wood, K., and Kotovsky, K., 2011, "On the Benefits and Pitfalls of Analogies for Innovative Design: Ideation Performance Based on Analogical Distance, Commonness, and Modality of Examples," *J. Mech. Des.*, **133**(8), p. 081004.
- [153] Perttula, M. K., Liikkanen, L. A., and others, 2006, "Exposure effects in design idea generation: Unconscious conformity or a product of sampling probability?," *Dev. Process From Idea to World's First Bionic Prosthet. Foot*.
- [154] Chrysikou, E. G., and Weisberg, R. W., 2005, "Following the wrong footsteps: fixation effects of pictorial examples in a design problem-solving task.," *J. Exp. Psychol. Learn. Mem. Cogn.*, **31**(5), pp. 1134–1148.
- [155] Siangliulue, P., Chan, J., Gajos, K. Z., and Dow, S. P., 2015, "Providing timely examples improves the quantity and quality of generated ideas," *Proc. 2015 ACM SIGCHI Conf. Creat. Cogn. - C&C '15*, pp. 83–92.
- [156] Atilola, O., and Linsey, J., 2015, "Representing analogies to influence fixation and creativity: A study comparing computer-aided design, photographs, and sketches," *Artif. Intell. Eng. Des. Anal. Manuf.*, **29**(02), pp. 161–171.
- [157] Atilola, O., Tomko, M., and Linsey, J. S., 2016, "The effects of representation on idea generation and design fixation: A study comparing sketches and function trees," *Des. Stud.*, **42**, pp. 110–136.
- [158] Tseng, I., Moss, J., Cagan, J., and Kotovsky, K., 2008, "The role of timing and analogical similarity in the stimulation of idea generation in design," *Des. Stud.*, **29**(3), pp. 203–221.

- [159] Higgins JPT, 2006, "Cochrane Handbook for Systematic Reviews of Interventions," Text, (September), p. 265.
- [160] Hedges, L. V, 1981, "Distribution theory for Glass's estimator of effect size and related estimators," J. Educ. Stat., **6**(2), pp. 107–128.
- [161] Hardy, R. J., and Thompson, S. G., 1998, "Detecting and describing heterogeneity in meta-analysis," Stat. Med., **17**(8), pp. 841–856.
- [162] Hardy, R. J., and Thompson, S. G., 1996, "A likelihood approach to meta-analysis with random effects," Stat. Med., **15**(6), pp. 619–629.
- [163] Viechtbauer, W., and others, 2010, "Conducting meta-analyses in R with the metafor package," J Stat Softw, **36**(3), pp. 1–48.
- [164] Rosenthal, R., 1979, "The file drawer problem and tolerance for null results.," Psychol. Bull., **86**(3), p. 638.
- [165] Higgins, J. P. T., Green, S., and others, 2005, "Special topics," Cochrane handbook for systematic reviews of interventions, Wiley Online Library, pp. 152–153.
- [166] Idris, N. R. N., 2012, "A comparison of methods to detect publication bias for meta-analysis of continuous data," J. Appl. Sci., **12**(13), p. 1413.
- [167] Macaskill, P., Walter, S. D., and Irwig, L., 2001, "A comparison of methods to detect publication bias in meta-analysis," Stat. Med., **20**(4), pp. 641–654.
- [168] Egger, M., Smith, G. D., Schneider, M., and Minder, C., 1997, "Bias in meta-analysis detected by a simple, graphical test," Bmj, **315**(7109), pp. 629–634.
- [169] Fu, K., Chan, J., Cagan, J., and Kotovsky, K., 2013, "The Meaning of 'Near' and 'Far': The Impact of Structuring Design Databases and the Effect of Distance of Analogy on Design Output," J. Mech. Des., **135**(February), pp. 1–12.