

A SYSTEMATIC APPROACH TO EVALUATING DESIGN PROMPTS IN SUPPORTING EXPERIMENTAL DESIGN RESEARCH

Patel, Apurva; Elena, Maria-Vittoria; Summers, Joshua

Clemson University

ABSTRACT

Experiments that study engineering behavior in design often rely on participants responding to a given design prompt or a problem statement. Moreover, researchers often find themselves testing multiple variables with a relatively small participant pool. In such situations multiple design prompts may be used to boost replication by giving each participant an equivalent problem with a different experimental condition. This paper presents a systematic approach to compare given design prompts using a two-step process that allows an initial comparison of the prompts and a post-experiment verification of the similarity of the given prompts. Comparison metrics are provided which can be used to evaluate a level of similarity of existing prompts as well as develop similar problems. These metrics include complexity (size, coupling, and solvability), familiarity, and prompt structure. Statistical methods are discussed for post-experiment verification. Guidelines are provided for a post-experiment survey which may be used for an additional perspective of prompt similarity. The proposed approach is demonstrated using an experiment where two design prompts were used for within-subject replication.

Keywords: Design theory, Research methodologies and methods, Early design phases, Design problem

Contact:

Summers, Joshua
Clemson University
Mechanical Engineering
United States of America
jsummer@clemson.edu

Cite this article: Patel, A., Elena, M.-V., Summers, J. (2019) 'A Systematic Approach to Evaluating Design Prompts in Supporting Experimental Design Research', in *Proceedings of the 22nd International Conference on Engineering Design (ICED19)*, Delft, The Netherlands, 5-8 August 2019. DOI:10.1017/dsi.2019.282

1 MOTIVATION: WHY COMPARE DESIGN PROMPTS

The engineering design process is seen as a key element in idea generation and development of a product (Pahl *et al.*, 2013; Shah *et al.*, 2003; Ullman, 1992; Ulrich and Eppinger, 2008). This process is generally iterative; including problem definition, conceptual, embodiment, and detailed design. The first step in the design process is to define the problem. Thus, the degree of success of the product is directly related to how the problem statement is defined and understood by the designer.

Understanding how engineers behave within the framework of solving engineering problems is of interest if new tools and methods are to be developed. A popular method of conducting research on designer behaviour is by giving a problem to participants to identify requirements or generate solutions. Some of these studies have been conducted on how design solutions change according to the engineer's level of expertise when given a design problem (Cross, 2004). Other research using design problems focuses on the generation of requirements based on different problem statements given to students (Worinkeng *et al.*, 2015). More work on design problems focuses on design thinking and representation mapping (Mckoy *et al.*, 2001). Solutions to these design problems are generally assessed according to four different ideation metrics: quality, quantity, variety, and novelty (Shah *et al.*, 2003). These studies can include both controlled experiments, where the outcomes are the focus of the study, or protocol studies, where the steps taken by the participants and their thinking is the focus of the investigation. In both tools, design problems are often used to either provide context to the design element being investigated or to prompt the participant to perform design activities. Most design prompts are taken from previous literature and, in some cases, adapted or changed accordingly. Design prompts are what are provided to participants in experimental studies, while design problems define the motivation and goal of engineering design projects. Often, the design problems will be used as design prompts.

1.1 Design prompts in literature

Design prompts are important when generating and assessing solutions. Researchers use the way design prompts are presented to restrict the solution space by elements such as its format, purpose, and content. By limiting the solution space, the solution is more focused, and therefore it is easier to explore a solution within this space (Goel and Pirolli, 1992). Design problems are also used to expand the design space. Thus, prompts focus on a specific solution space that addresses a deeper and extended design space. However, other approaches view the problem and the solution as co-evolving (Dorst and Cross, 2001). This means that the design prompt and the solution progress in combination with one another.

Existing literature addresses how the design problem affects the solution. A variety of design problems have been used to study requirements, functions, and conceptual design. When using more than one prompt in an experiment, problem similarity becomes crucial. The equivalency of more than one design problem is determined by the similarity of the solutions produced. Creating equivalent design prompts remains a challenge since they must encourage quality, quantity, novelty, and variety while limiting fixation (Linsey *et al.*, 2005). Work has been done to identify defining elements of design problems (Durand *et al.*, 2015; Kumar and Mocko, 2016) as well as investigating problem similarity (Kumar and Mocko, 2016; Levy *et al.*, 2017). The degree of similarity of many design problems was assessed based on their representation using Latent Semantic Analysis (Kumar & Mocko, 2016). The complexity of design problems was divided into three categories: size, solvability, and coupling (Summers and Shah, 2010). The equivalency of four design problems was also studied based on Shah's quantity, variety, novelty, completeness, and quality (Kumar & Mocko, 2016; Levy, Hilton, Tomko, & Linsey, 2017). On the other hand, it has been argued that the design problem is not what drives the solution, but rather the experience level of the designer (Dorst, 2003). These works are an effort to better understand design problems and their purpose within engineering design.

1.2 Why use multiple prompts?

A non-exhaustive review of design problems used in engineering design literature is provided in Table 1. In the table, the cells are coded as P (problem) or S (solution) depending on what was analysed by the study. The column on the far right indicates how many problems were used in each experiment. The shaded row at the bottom shows the frequency of each design problem used throughout the studies. The rows on the left are numbered 1-14 and show the studies with their respective university affiliations. The top row

displays the design problems used in the studies, lettered A-N. At the very bottom of the table, a legend is shown to describe the university abbreviations.

Table 1: Design prompts in literature

Number	Study Reference	University Affiliation	Design Problem														Number of Problems
			A	B	C	D	E	F	G	H	I	J	K	L	M	N	
			Chocolate Packaging	Drinking Fountain	Peanut Shelling	Corn Husking	Coconut Harvesting	Personal Alarm Clock	Bike Safety Lock	Book Picking (a)	Blind Cup	Clothes Ironing	Recycling Sorter (b)	Spill proof coffee cup	Bike Rack	Burrito Folding	
1	(Chawla and Summers, 2018)	CU	-	-	-	-	-	-	-	-	S	S	-	-	-	2	
2	(Levy et al., 2017)	GT	-	-	P	P	P	P	-	-	-	-	-	-	-	4	
3	(Patel et al., 2016)	CU	-	-	-	-	-	-	-	-	S	S	-	-	-	2	
4	(Durand et al., 2015)	GT	-	-	S	S	-	S	-	S	-	-	-	-	-	4	
5	(Worinkeng et al., 2015)	CU	-	-	-	-	-	-	-	-	-	-	-	-	S	1	
6	(Richardson III et al., 2011)	CU	-	-	-	-	-	-	-	-	-	-	-	-	S	1	
7	(Cardoso and Badke-Schaub, 2011)	DU	-	-	-	-	-	-	S	-	-	-	-	-	-	1	
8	(Linsey et al., 2011)	GT	-	-	S	-	-	-	-	-	-	-	-	-	-	1	
9	(Fu et al., 2010)	GT	-	-	S	-	-	-	-	-	-	-	-	-	-	1	
10	(Linsey et al., 2008)	GT	-	-	S	-	-	-	-	-	-	-	-	-	-	1	
11	(Goldschmidt and Smolkov, 2006)	IT	S	S	-	-	-	-	-	-	-	-	-	-	-	2	
12	(Smith et al., 2006)	CU	-	-	-	-	-	-	-	-	-	-	-	-	S	1	
13	(Mckoy et al., 2001)	ASU	-	-	-	-	-	-	S	S	-	-	-	-	-	2	
14	(Jansson and Smith, 1991)	TA	-	-	-	-	-	-	-	S	-	-	S	S	-	3	
Frequency			1	1	5	2	1	2	1	2	2	2	2	1	1	3	
University Affiliation: CU - Clemson University; GT - Georgia Tech; DU - Delft University of Technology; IT- Israel Tech; ASU – Arizona State University TA – Texas A&M Analysis Type: P – Problem focused; S – Solution focused;																	
(a) Similar to peach picking (Joshi and Summers, 2014)																	
(b) Similar to trash disposal system (Dorst and Cross, 2001)																	

Table 1 provides insight into the design problems used in literature. Study 2 is the only one that focuses on the problem space. This study discusses problem similarity by comparing problems C, D, E, and F and evaluating them using the ideation metrics of quality, quantity, variety, and novelty (Shah et al., 2003). All other studies listed use design problems for idea generation methods or design fixation and analyse the solutions generated by the participants. Data presented in Table 1 also suggests that problems are reused within the same university but not across different ones. For example, Georgia Tech reuses C, D, E, and F, while Clemson University reuses problems J, K, and N. Additionally, it is important to note that the recycling sorter prompt (K) was derived from the litter disposal system design problem (Dorst and Cross, 2001). Similarly, a peach picking problem was created from the book picking prompt (H) (Joshi and Summers, 2014). Overall, the most frequently used problem statement is C, used repeatedly from studies conducted at Georgia Tech. This is not an exhaustive review of the problems used in engineering design research, but representative for the purposes of this paper.

Another important takeaway is that not all studies use two or more design problems in the same experiment. Study 14 reports on three distinct experiments, each employing a different design problem. Likewise, study 11 details two experiments each with a unique design problem. On the other hand, studies 1, 3, and 13 use two problems for one experiment. Studies 2 and 4 also use four problems to support each experiment reported. The rest of the listed literature, however, uses only one design prompt for one experiment.

This work presents an integrated approach to help researchers develop similar design problems to boost repetition and increase the statistical power of resulting observations. The number of design problems needed depends on the number of available participants. For example, when a significantly large sample size can be achieved with the available participants, a single design problem may be sufficient. However, when fewer participants are available, multiple design problems can be used to increase the sample size by extracting more than one response from each participant. Therefore, using multiple design problems, can help researchers increase replication and subsequently the statistical power of tests conducted. The following section details the proposed procedure of developing similar design prompts and validating the similarity based on the observations as well as participant perceptions.

2 PROPOSED METHOD

We propose a two-step process that includes deliberately developing similar design problems which can be administered to participants for within-subject replication. After the experiment, the participant responses are compared to validate the problem similarity based on predefined test measures. It should be noted that this approach has been developed mainly for design prompts represented in a textual format. However, the approach can be modified to be more applicable for a different representation.

2.1 Developing similar design prompts

When designing an experiment, researchers may be trying to understand the effects of multiple factors with varying levels of treatments. In such cases, high amount of replication is necessary to draw statistically significant conclusions. Unfortunately, experiments with human participants generally have a low amount of replication, resulting in limited generalizability of the conclusions. This is especially true in cases where researchers are forced to divide the participant pool into multiple sections in order to isolate the effects of different factors and levels. Thus, introducing a new but similar problem can help with maintaining a higher level of replication.

Developing similar design prompts will require a systematic and reliable comparison between given design problems. Prior work investigating design problems has identified characteristics of a design problem (Durand *et al.*, 2015; Kumar and Mocko, 2016) which can serve as a guideline to develop a systematic comparison between given design problems. The problem can be compared based on the complexity of the design prompt, the familiarity of the content, or semantics. These metrics are shown in Table 2. Ultimately, the researcher should justify the set of metrics selected for comparison, provide a discussion of the evaluation methods, and finally identify what level of similarity is acceptable.

Table 2: Metrics for prompt comparison

		Metric	Evaluation Methods
Complexity	Size	Words	Count
		Characters	Count
		Sentences	Count
		Paragraphs	Count
		Embedded Requirements or Functions	Count
	Coupling		Delphi Method/Survey
	Solvability	Difficulty level	Delphi Method/Survey
Time to solution		Delphi Method/Survey	
Familiarity	Domain Familiarity	Delphi Method/Survey	
	Cultural Consistency	Delphi Method/Survey	
	Reading level	Flesch Reading Level	
End-user info			Identify
Existing artefacts			Identify
Prompt Structure			Storyboarding
Semantic Similarity			LSA/LDA, Survey

Complexity of the design prompt can be further divided into three branches: size, coupling, and solvability (Summers and Shah, 2010). Size of the prompt can be evaluated using various measures such as the length of the prompt in words, the number of characters, the number of sentences, or the number of paragraphs. These are objective measures of the design prompt and therefore provide a more objective comparison of given prompts. Additionally, the size of the prompt can be measured in terms of the number of requirements embedded or the number of functions explicitly identified within the prompt. Although the embedded requirements and functions can be identified with the design prompt, comparisons on these metrics between two or more prompts may be challenging since an objective comparison of two requirements or high-level functions may not be possible (especially at the design problem level). When comparing design prompts using size-based metrics, a similarity level of 90% or higher can be considered sufficient for most cases.

Another aspect of complexity is coupling. This can refer to the coupling of energy, material, and signal flows through the system, the coupling within the desired functionality, or requirements coupling. The researcher may be able to evaluate this metrics based on past experiences; however, a panel of experts should be consulted, similar to the Delphi method (Ozkan and Dogan, 2013), to identify the level of coupling. For simplicity, it may be beneficial to only identify whether the problem has low, medium, or high coupling and use that for comparing design prompts. Finally, the complexity of design prompts can also be measured based on solvability. In this case, solvability has two aspects: difficulty level of the prompt, and time to completion. Both aspects can either be measured by surveying a representative population, or by consulting a panel of experts such as relevant faculty (Thoe and Summers, 2013). Additionally, time to solution can also be determined by performing a pilot study with a small number of participants and using the time needed to complete as an estimate.

Next, the design prompts can be evaluated using the participants' familiarity with the given prompt. Familiarity includes domain familiarity, cultural consistency, and reading level. Domain familiarity refers to the subject matter of the design prompt. This can be evaluated either on the broad scale, such as mechanical or biology, or on a narrow scale, such as tire design or the control system for a vacuum robot. However, the evaluation scale must be consistent between prompts. While the researcher may be able to accurately identify the domain for the given prompts, it may be beneficial to use a panel of experts or a survey to be more objective. Another aspect of familiarity is cultural consistency, which refers to participants from different cultural backgrounds being able to interpret the given design prompt in a similar manner. This metric can be evaluated using a survey of culturally diverse populations regarding the topics discussed within the prompts. Finally, the reading level of the prompts should be similar. This metric can be evaluated using either the Flesch reading-ease test, or another comparable test. The goal of this metric is to ensure that the reading level is appropriate for the participants, and consistent between design prompts.

Following familiarity, the additional information provided in the prompt should also be used for comparison. This includes information about the end-user of the artefact, or the system requested by the design prompt, as well as any existing artefacts explicitly identified in the prompt. Both metrics can be evaluated by reviewing the design prompt and identifying explicit references of the end-user and existing artefacts. The design prompts should also be compared based on the structure of the prompt. This include the order in which the different parts of the design prompt are presented, and the emphasis on different aspects of the prompt. This metric can be evaluated by creating a storyboard or "diagramming" the prompt, which can then be used to compare prompts based on an edit distance or percent similarity. Finally, the semantic similarity of the prompts can be compared using Latent Semantic Analysis.

Using these metrics, a level of similarity can be defined for a set of design prompts. The design prompts can be modified until an acceptable level of similarity is achieved. This may depend on the type of research being conducted, or the specific subject being invested. Additionally, the selection of comparison metrics presented in this paper are a general guideline, and the specific metrics for comparing the design prompts should be determined and justified on a case by case basis. When the design prompts are deemed to be similar, they can be administered to the participants. Data collected from the experiment can then be used to validate the similarity of the design prompts.

2.2 Validating prompt similarly

The design prompts given to participants are inputs to the experiments; correspondingly, the participant responses are the outputs from the experiment. As such, the similarity of the design

problems can be validated by comparing the experiment outputs, or participant responses, for similarity. Assuming discrete data points are collected as part of the experiment, a statistical approach for comparison can be used to identify whether the design problems produced similar results (Levy *et al.*, 2017). As this paper is not intended to be a review of statistical methods, an exhaustive discussion of the various statistical tests is not provided; instead, a brief overview of common methods is offered.

An analysis of variances (ANOVA) is commonly used to identify differences between group means in a given dataset. Among the studies presented in Table 1, studies 2, 4, 5, 7, 8, 9, 11, and 13 used ANOVA to some extent in their statistical analysis. The Kruskal-Wallis test is a similar analysis used to test whether samples belong to the same distribution (used by study 2 and 9). Another commonly used method is hypothesis testing using the t-statistic, or commonly known as a t-test. This type of hypothesis testing has been used by studies 2, 3, 4, and 11. Similar to the t-test, a Mann-Whitney U test (or Wilcoxon rank-sum test) can also be used to determine whether two samples are from the same distribution (used by studies 2 and 4). Unlike the t-test, the Mann-Whitney U test does not require an assumption of normality. As certain statistical analyses require an assumption of normality, it may be beneficial to test whether the data collected is from a normal distribution. This can be done using the Shapiro-Wilks test or the Pearson Chi-squared test. Ultimately, the specific statistical analysis needed will depend on the type of data collected, the research questions being investigated, and the participant pool being studied.

2.3 Post-experiment survey

In addition to the deliberate design of similar prompts and the validation of prompt similarity based on experiment outputs, how the design prompts are perceived by the participants is an important factor in ensuring that the results collected from the two different design prompts are truly comparable. Participant perceptions can be measured using methods such as surveys, interviews, and focus groups. The specific method of comparison may depend on the resources available to the researcher as well as the scale of the experiment; however, certain guidelines should be considered in any case.

As the goal of the survey is to measure the participants' perception of the design prompts, especially the similarity between prompts, a cross-sectional survey design will generally be used. A cross-sectional survey is intended to measure opinions, beliefs, and practices at a particular instance in time (Creswell, 2002). Within the survey, different types of questions may be used, including but not limited to personal or demographic questions, behavioural questions, fact-finding questions, and closed/open-ended questions. A mix of different types of questions should be used to collect information from different perspectives and triangulate responses. Additionally, questions regarding participants' individual characteristics should be included as these will provide information needed to identify relationships between individual characteristics and attitudes about problem similarity.

The survey should be introduced with easy questions, such as considered demographics, and should be divided into sections to avoid overwhelming the participants. One should use graphics where necessary and a rating scale where possible. The rating scale should be consistent throughout the survey, unless otherwise necessary. Additionally, binary questions should be used for soliciting basic facts. Finally, the survey should be structured such that the most important information is collected first, in case the participants are unable to complete the entire survey. If possible, the survey should be tested in a pilot study with a representative population to estimate time needed and identify any issues with the survey.

3 APPLICATION EXAMPLE

In this section, application of the proposed approach will be demonstrated using data from an experiment studying the effects of model chaining in function structure models.

3.1 Experiment overview

An experiment was conducted to investigate the effects of seeding function structure models with different types of seeds (Patel *et al.*, 2016). A mixed factorial experiment design was used with within-subject and between-subject replication. The two independent variables, chaining method and level of completion, each had three levels of treatment. As a result, the participant pool had to be divided into nine groups in order to avoid learning effects from reuse of the design prompt. To maintain a reasonable amount of replication for each independent variable and each treatment, a second design prompt was used. This allowed each participant to provide two independent responses, effectively doubling the number of participants. Fourth year mechanical engineering students participated in the study.

3.2 Design prompts

The design prompts used in this experiment are shown in Table 3. Prompt A was adopted from previous work investigating how designers create function structure models (Sen and Summers, 2012), whereas prompt B was inspired from an existing design problem for a trash disposal system (Dorst and Cross, 2001).

Table 3: Design prompts used in experiment

Prompt A: Automatic Clothes Ironing Device
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.
Prompt B: Automatic Recycling Sorter
Design an automatic recycling machine for household use. The device should sort plastic bottles, glass containers, aluminium 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.

The design prompts were then compared using the comparison metrics. Table 4 shows the comparison of the design problems based on complexity. A similarity score is provided.

Table 4: Comparing prompts based on size

Metric	Prompt A	Prompt B	Similarity	
Words	70	61	85.20%	✘
Characters	415	391	93.90%	✓
Sentences	5	5	100%	✓
Paragraphs	1	1	100%	✓
Embedded Functions	3 (sort, press, fold)	3 (sort, compress, store)	100%	✓
Coupling	Low	Low	100%	✓
Difficulty level	Moderate	Moderate	100%	✓
Time to solution	15-20 minutes	15-20 minutes	100%	✓

While the prompts are not of identical size, the level of similarity in size was considered sufficient for the purpose of this experiment. Additionally, the level of coupling and the difficulty level were identified to be similar by a faculty familiar with the student participants. A pilot study was conducted to identify the time needed for completion for each design prompt.

As the participants were enrolled in a pre-capstone design course, a similar level domain familiarity was expected for both design prompts. Additionally, the concepts of ironing and recycling were assumed to be culturally consistent for participants, most of whom were united states residents between the ages of 18 and 25. Finally, a Flesch reading ease score of 60 and 47 was calculated for prompt A and prompt B respectively. Both suggested a recommended reading level below that of college students, ensuring that the participants will be able to understand the prompts. Each prompt provided limited information regarding the end user and neither prompt alluded to an existing artefact. The structure of the prompts was designed to be similar, starting with background, followed by intended functions, resource restrictions, and finally time requirements. A Latent Semantic Analysis¹ of the design prompts found a term-to-term similarity of 89% and a document-to-document similarity of 52%. This discrepancy in semantic similarity suggests that the terminology used in the two prompts was similar; however, the overall meaning of the prompt was different. This was intended when creating the design prompts.

¹ This was computed using the pairwise comparison tool at <http://lsa.colorado.edu/> (2018.12.05)

3.3 Comparison of responses

The responses collected from the experiment included the number of functions and flows added by each participant. In order to mitigate bias from the given number of functions and flows, a percent increase in the number of functions and flows was calculated. Next, these responses were separated into prompt A and prompt B, and each set of samples was tested for normality using the Shapiro-Wilks test. A significance level of 0.05 was used, and the p-values are shown in Table 5.

Table 5: Normality test results (p-values)

	Prompt A	Prompt B
Added Functions	1.94E-09	9.72E-07
Added Flows	5.48E-07	4.72E-07

The small p-values suggest that the Shapiro-Wilks tests determined the data was significantly different from a sample normal distribution. Statistical tests that rely on the data being normal are not advised. Next, the two prompts were compared using a two-sample t-test to investigate the similarity between responses. The means were hypothesized to be equal, and an assumption of unequal variances was used. A total of 86 samples were collected for each prompt, which were then separated based on chaining method used in the seed model. A significance level of 0.15 was used in order to allow more flexibility for the problems to be different (Wetmore *et al.*, 2010). Means, Variances, and results from the t-test are shown in Table 6.

Table 6: Percent increase in functions based on chaining methods

		Prompt A	Prompt B	p-Value	Result
Backward	Mean	93.3%	83.5%	0.690	Failed to Reject
	Variance	87.3%	81.1%		
Forward	Mean	60.1%	50.5%	0.447	Failed to Reject
	Variance	25.9%	20.9%		
Nucleation	Mean	98.2%	81.8%	0.509	Failed to Reject
	Variance	138.5%	21.4%		

Prompt A generally yielded a larger increase in function compared to prompt B; however, the p-values show that sufficient evidence was not found to determine a significant difference between the two prompts. Similar results were seen for the increase in number of flows. For instance, output for t-test comparing the increase in flows when using backward chaining are shown in Table 7.

Table 7: t-Test results for comparison of increase in flows based on backward chaining

	Prompt A	Prompt B
Mean	1.13	1.08
Variance	1.14	1.58
Observations	31	26
Hypothesized Mean Difference	0	
df	49	
t Stat	0.16	
P(T<=t) one-tail	0.44	
t Critical one-tail	1.05	
P(T<=t) two-tail	0.88	
t Critical two-tail	1.46	

This t-test was performed assuming unequal variances, and the result suggests that the two prompts are not significantly different ($P = 0.44$) based on the increase in flows for backward chaining. Therefore, the prompts were not found to be significantly different with respect to the increase in functions and flows. This supports the initial assumption that the design prompts are comparable, and the data collected from both prompts can be combined to obtain more replication within the experiment. Finally, the participating students reported that they were able to understand both prompts to a similar level and perceive them of a similar difficulty on average (survey responses on a Likert scale).

Participants were also equally familiar with both concepts of ironing and recycling; they reported participating in each of these activities. Details of the survey responses collected have been omitted from this paper for brevity.

4 CONCLUSIONS

In conclusion, a review of design prompts in engineering literature showed that most studies rely on a single design prompt and that design prompts themselves are not a frequent topic of research. Existing work on design prompts demonstrates tools and methods that can be used to systematically compare design prompts. A two-step approach is presented in this paper to use two or more design prompts in an experiment. Metrics are provided to compare given prompts and identify a level of similarity. These can also be used as guidelines to develop similar design prompts. A method for validating similarity of the prompts is also discussed that includes participant responses to the prompt as well as their subjective perception. Finally, an example demonstrating the proposed approach is provided. While this approach provides an alternative method for increasing replication with inadequate number of participants, following limitations have been identified.

- Additional design prompts need to be developed, which requires more time and effort from the researcher. Also, in specific cases, developing problems with an acceptable level of similarity may be challenging.
- More work is required from the participants, which may lead to participants losing interest or burning out. For example, if two prompts are provided, the participants may perform poorly on the second prompt. This should be considered during the experiment design and data analysis.
- A learning effect may occur, resulting in the within-subject replication losing independency. In other words, if two prompts are provided, responses to the second prompt may be affected by the participants' experience with the first prompt.

REFERENCES

- Cardoso, C. and Badke-Schaub, P. (2011), "The influence of different pictorial representations during idea generation", *Journal of Creative Behavior*, Vol. 45 No. 2, pp. 130–146.
- Chawla, A. and Summers, J.D. (2018), "Function Ordering Within Morphological Charts: An Experimental Study", *ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, pp. V007T06A012–V007T06A012.
- Creswell, J. (2002), "Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, Fourth", *Pearson Education*.
- Cross, N. (2004), "Expertise in design: An overview", *Design Studies*, Vol. 25 No. 5, pp. 427–441.
- Dorst, K. (2003), "The problem of design problems", *Expertise in Design*, pp. 135–147.
- Dorst, K. and Cross, N. (2001), "Creativity in the design process: Co-evolution of problem-solution", *Design Studies*, Vol. 22 No. 5, pp. 425–437.
- Durand, F., Helms, M.E., Tsenn, J., McAdams, D.A. and Linsey, J.S. (2015), "In Search of Effective Design Problems for Design Research", Vol. 7: *27th International Conference on Design Theory and Methodology*, ASME, p. V007T06A011.
- Fu, K., Cagan, J. and Kotovsky, K. (2010), "Design Team Convergence: The Influence of Example Solution Quality", *Journal of Mechanical Design*, Vol. 132 No. 11, p. 111005.
- Goel, V. and Pirolli, P. (1992), "The structure of Design Problem Spaces", *Cognitive Science*, Vol. 16 No. 3, pp. 395–429.
- Goldschmidt, G. and Smolkov, M. (2006), "Variances in the impact of visual stimuli on design problem solving performance", *Design Studies*, Vol. 27 No. 5, pp. 549–569.
- Jansson, D.G. and Smith, S.M. (1991), "Design fixation", *Design Studies*.
- Joshi, S. and Summers, J.D. (2014), "Impact of Requirements Elicitation Activity on Idea Generation: A Designer Study", *Detc*, No. November, pp. 1–10.
- Kumar, V. and Mocko, G. (2016), "Similarity of Engineering Design Problems to Enable Reuse in Design Research Experiments", Vol. 7: *28th International Conference on Design Theory and Methodology*, ASME, p. V007T06A042.
- Levy, B., Hilton, E., Tomko, M. and Linsey, J. (2017), "Investigating Problem Similarity Through Study of Between-Subject and Within-Subject Experiments", Vol. 7: *29th International Conference on Design Theory and Methodology*, ASME, p. V007T06A012.

- 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”, *Journal of Mechanical Design*, Vol. 133 No. 3, p. 031008.
- 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”, Vol. 5a: *17th International Conference on Design Theory and Methodology*, Vol. 2005, ASME, pp. 277–290.
- Linsey, J.S., Wood, K.L. and Markman, A.B. (2008), “Increasing Innovation: Presentation and Evaluation of the Wordtree Design-by-Analogy Method”, Vol 4: *20th International Conference on Design Theory and Methodology; Second International Conference on Micro- and Nanosystems*, ASME, pp. 21–32.
- Mckoy, F.L., Vargas-hernández, N., Summers, J.D. and Shah, J.J. (2001), “Influence of design representation on effectiveness of idea generation”, *Proceedings of the ASME Design Engineering Technical Conference*, Pittsburgh, PA, United States, pp. 39–48.
- Ozkan, O. and Dogan, F. (2013), “Cognitive strategies of analogical reasoning in design: Differences between expert and novice designers”, *Design Studies*, Elsevier, Vol. 34 No. 2, pp. 161–192.
- Pahl, G., Beitz, W., Blessing, L., Feldhusen, J., Grote, K.-H.H. and Wallace, K. (2013), *Engineering Design: A Systematic Approach*, edited by Second, 3rd ed., Vol. 11, Springer-Verlag London Limited, London.
- Patel, A.R., Summers, J.D., Kramer, W.S. and Shuffler-Porter, M.L. (2016), “Function modeling: A study of sequential model completion based on count and chaining of functions”, *Proceedings of the ASME Design Engineering Technical Conference*, Vol. 7, available at: <https://doi.org/10.1115/DETC201659860>.
- Richardson III, J.L., Summers, J.D. and Mocko, G.M. (2011), “Function Representations in Morphological Charts: An Experimental Study on Variety and Novelty on Means Generated”, *Interdisciplinary Design: Proceedings of the 21st CIRP Design Conference*, p. 76.
- Sen, C. and Summers, J.D. (2012), “A Pilot Protocol Study on How Designers Construct Function Structures in Novel Design”, edited by Gero, J. *5th International Conference on Design Computing and Cognition*, Springer-Verlag, College Station, TX, No. 37.
- Shah, J.J., Smith, S.M. and Vargas-Hernández, N. (2003), “Metrics for Measuring Ideation Effectiveness”, *Design Studies*, Vol. 24, pp. 111–134.
- Smith, G., Troy, T.J. and Summers, J.D. (2006), “Concept Exploration Through Morphological Charts: An Experimental Study”, Vol. 4a: *18th International Conference on Design Theory and Methodology*, Vol. 2006, ASME, pp. 495–504.
- Summers, J.D. and Shah, J.J. (2010), “Mechanical Engineering Design Complexity Metrics: Size, Coupling, and Solvability”, *Journal of Mechanical Design, American Society of Mechanical Engineers*, Vol. 132 No. 2, p. 021004.
- Thoe, S. and Summers, J.D. (2013), “Correlating Problem/Process Exam Question Complexity to Anticipated Effort: A Modeling Protocol”, Vol. 1: *15th International Conference on Advanced Vehicle Technologies; 10th International Conference on Design Education; 7th International Conference on Micro- and Nanosystems*, ASME, p. V001T04A012.
- Ullman, D.G. (1992), “The Mechanical Design Process”, *Maelabsucsdedu*, Vol. 3. ed, McGraw-Hill, New York, NY.
- Ulrich, K.T. and Eppinger, S.D. (2008), “Product Design and Development”, *Product Design and Development*, 4th ed., Vol. 384, McGraw-Hill, New York, USA, available at: <https://doi.org/10.1016/B978-0-7506-8985-4.00002-4>.
- Wetmore, W.R., Summers, J.D. and Greenstein, J.S. (2010), “Experimental study of influence of group familiarity and information sharing on design review effectiveness”, *Journal of Engineering Design*, Vol. 21 No. 1, pp. 111–126.
- Worinkeng, E., Joshi, S. and Summers, J.D. (2015), “An experimental study: analyzing requirement type influence on novelty and variety of generated solutions”, *International Journal of Design Creativity and Innovation*, Taylor & Francis, Vol. 3 No. 2, pp. 61–77.