



THE IMPACT OF DESIGN BRIEFS ON CREATIVITY: A STUDY ON MEASURING STUDENT DESIGNERS OUTCOMES

G. Koronis, A. Silva and J. Kang

Abstract

This study is based on an interdisciplinary project aimed at ways to improve creativity among student designers. We examine the influence of different kinds of stimuli and relationships between ideas generation in product design creative outcomes. This entails a design of experiments approach to measure and determine whether factors as quantitative requirements, visual and physical stimuli can affect creativity scores. The statistical analysis suggests that briefs of no quantitative data without additional stimuli produce high scored ideas and minimize the variability of all three factors.

Keywords: idea generation, conceptual design, design creativity

1. Introduction

The scope of this study encompasses the delivery of an in-class exercise to first-year undergrad student designers in "Introduction to Design", an essential foundation course for all Singapore University of Technology and Design (SUTD) freshmen students. One of the key objectives of this teaching exercise is to enable design teams to explore the problem and design space and generate creative and innovative outputs at the time of ideation. It has been observed that a designer's performance is dependent on the type of information provided in the brief, meaning that different types of stimuli in the design brief have different effects on the design ideas (Howard et al., 2010). Therefore, we seek to investigate how to craft a design brief with the most significant qualitative and quantitative information in order to foster creative concept generation. To the best of the authors' knowledge, little effort has been devoted coherently and purposefully to extract guidelines or rules for writing design briefs and applying them in design practice and research. Most of the existing effort has been in the construction sector; for example, Ryd (2004) studied client-contractor interactions for a construction project where the client experimented with an unconventional construction brief. Studies in the vein of engineering design have explored individual creativity (Storme et al., 2014) or creativity in small groups of students (Demirkan and Afacan, 2012; Toh and Miller, 2015; Chen, 2016) and professionals (Binyamin and Carmeli, 2017). However, these studies have limited sample sizes, and this limits the generalizability of results from the past literature. While there is debate in the literature on the effect of constraints on creativity, there are studies which suggest that including quantitative requirements in design briefs adds a higher level of constraints that improves the creative processes, as validated in industry-related studies such as in the design of healthcare devices (Onarheim, 2012) and applied art and product design by professionals and student designers (Bonnardel, 2000). In contrast, suggested that the type of design brief used has no relationship with cognitive styles in the creative process (i.e. the "exploration, generation and development of ideas") (Kim and Kim, 2015). Through the evaluation of the results of a Collaborative Sketching (C-Sketch) activity, we seek to explore if these claims concerning idea generation are valid and if the evidence

supports the claim that creative outcomes are influenced by briefs with different sets of stimuli and/or given information. The findings will be useful for educators and professionals from different backgrounds (practicing engineers, senior managers, etc.) to deliver briefs that are tailored to promote creative outcomes.

With the abovementioned objectives, we examine the following research questions specific to this paper:

1. Which are the main factors or interaction effects which affect creativity among student designers?
2. Which main effects or interaction effects influence variability in creativity?
3. What is the best factor setting for maximizing creativity and minimizing variability among responses?

2. Materials and methods

A design of experiments (DoE) approach was implemented to determine whether quantitative brief guidelines, physical or visual examples, or a combination of stimuli can yield higher scores for all creativity metrics. Having a set of metrics to evaluate the creativity of outcomes from the C-Sketch activity is critical for analyzing our results. However, to our knowledge, there is no well-established or standardized methodology concerning the amount or type of metrics for idea evaluation, and as seen by the following list of some of the creativity evaluation metrics in the literature. A comprehensive study on metrics for creativity and innovation using in-class design projects has been done by Oman et al. (2012). Shah et al. (2003) suggested an innovation assessment framework which employs the metrics of novelty, variety, quality, and quantity, measuring different aspects of ideation effectiveness. They concluded that since each of metrics measures something different, summing up the metrics to create an overall ideation effectiveness measure makes no sense. Kamylyis and Valtanen (2010) suggest that creative product(s) must be novel (original, unconventional) and appropriate (valuable, useful). In a study by Bourgeois-Bougrine et al. (2017), creativity involves the production of original, potentially workable, ideas to solve a problem.

We chose to employ a combination of metrics from the design science literature as a framework for evaluating the creativity of our experiment's ideation outcomes. We thereby employed the metrics novelty, appropriateness and usability for use in our study. We consider novel ideas as the ones that are unconventional, unusual and/or original, regardless of whether the ideas are aligned with the brief. Measuring novelty in idea generation is of fundamental importance (Shah et al., 2003), as such, the more "original" the concept ideas for the group, the higher the group scores in that metric. We additionally judge appropriateness based on how the designs are aligned with the brief's specifications. On a pilot study we conducted using these creativity metrics we realised that our metrics did not capture the effectiveness of the design outcome; an important element in functional engineering creativity (Cropley and Cropley, 2005). Thus, to make our assessment more comprehensive added a third metric, usability, to describe workability and effectiveness of the product in terms of extracting the most amount of juice with minimal effort. This is aligned to the definition of usability as a quality attribute, as described by Nielsen (1993) and in ISO (2013) as the extent to which a system, product, or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. In that sense, usability is a quality metric for efficiency and success rate. We acknowledge that unless a usability test is run, it is difficult to assess user-product interactions (Sauer and Sonderegger, 2009). However, the sketches give a flavor of the performance of the product, which we assume can be quantified by our judges.

2.1. Full factorial DoE

To test the effect of stimuli in design briefs on the creativity of student outcomes, we carried out a 2³ two-level, full-factorial experiment and studied all possible combinations and their respective interactions. The three factors were: quantitative requirements (Q), meaning information on specific costs, maximum number of manufacturing processes, durability cycles and product volumes; visual stimulus (V) comes in the form of a video showing a user utilizing a conventional orange squeezer, which consists of a simple, single-part device; physical stimulus (P), which is a sample of a simple, single-part orange squeezer that is handed to selected teams. Table 1 presents the factor levels and the response values collected for each condition. There are three main effects to be evaluated (Q, V, and P)

and their second-order interactions $Q \times V$, $Q \times P$ and $V \times P$. Third-order interactions are generally ignored in experiments in industrial settings, which is not the case here. Thus, the third-order interaction ($Q \times V \times P$) was retained to observe possible interactions of the main effects related to their variability. For each of the factor combinations, we collected 4-5 observations to obtain a reasonable estimate of error variance (or experimental error) and provide sufficient degrees of freedom to characterize our sample distribution. In total, 35 experimental data points were collected. Each of the data points represents the average score for all sketches per group of students (see Section 2.3 for details). The statistical software Minitab® 17 was used for the statistical analysis of the experimental data.

Table 1. List of factors and factor levels

Factors	Labels	Low Level (-)	High Level (+)
Quantitative Description	Q	No	Yes
Visual Stimulus	V	No	Yes
Physical Stimulus	P	No	Yes

2.2. Participants

The participants in this study were 181 first-year undergraduate engineering/architecture student designers in an engineering course. The students were 18-21 years old and were allocated to 8 classes with an even mix of gender, nationality, educational background, academic grades, subject preference for majors, and Myer-Briggs personality scores. Therefore, the classes were reasonably homogeneous and with a uniform standard of student performance, which allows for fair comparison between groups.

2.3. Experiment and analysis procedure

All student designers from the 8 different classes were divided into teams of 4-6 students, and each student was given an identical set of materials that included an A3 sheet of paper and a colored pen. The C-Sketch activity, a variation of the 6-3-5 concept development team methodology, was conducted by presenting 8 different design briefs to the 8 different classes. Student designers were asked to "Design a device to extract juice from fresh oranges at home." The design brief description was displayed on projectors during the entire period of the exercise. Subjects were asked to sketch and elaborate on their ideas on paper as concept sketching is open-ended enough to permit flexibility and novelty in responses (Amabile, 1982). The exercise starts with each student drawing three concept drawings in response to the design brief. Upon a given time limit per session, the drawings are circulated among all team members so that everyone adds features and annotations to each of his or her teammates' concept ideas. Verbal communication and distractions were kept to a minimum. We did not expect any of the participants to have extensive prior experience in designing analogous juice extracting devices, but we assumed that squeezing oranges is a task that most students must have experienced before. This exercise yielded 522 concept drawings (some students did not complete all 3 drawings). As the C-Sketch exercise was performed within teams, we also chose to evaluate the results at team-level and therefore have 35 data points (not all classes had the same number of teams). Descriptions of the briefs provided to each class are listed below, with the number in parentheses indicating the number of students ($n=181$) per brief/class.

- (26) Brief A: Simple problem description (control).
- (23) Brief B: Control + More detailed quantitative requirements.
- (24) Brief C: Control + Physical example of an orange squeezer.
- (19) Brief D: Control + More detailed quantitative requirements + Physical example of an orange squeezer.
- (21) Brief E: Control + Video of a manual orange squeezer + Physical example of an orange squeezer.
- (20) Brief F: Control + More detailed quantitative requirements + Physical example of an orange squeezer + Video of a manual orange squeezer.
- (25) Brief G: Control + Video of a manual orange squeezer.
- (23) Brief H: Control + More detailed quantitative requirements + Video of a manual orange squeezer.

Figure 1 shows the control design brief. Figure 2 shows the physical orange squeezer. Figure 3 depicts a video that shows the use of a manual orange squeezer. Figure 4 shows the extended design brief including more detailed quantitative requirements. Figure 5 includes selected sketches that received high scores from all judges. Contrastively, Figure 6 portrays sketches that received low scores and thus represent poor examples of the response values.

Design Problem	Design a device to extract orange juice from fresh oranges at home.
Sample Customer Needs	Low cost. Easy to manufacture. Machine washable. Small footprint.
Other Relevant Data	Do not use blenders or blender-type machines as a base for designs.

Figure 1. The control brief A



Figure 2. A sample of orange squeezer provided with briefs C, D, E and F



Figure 3. Video showing the use of a manual orange squeezer in briefs E, F, G and H

Design problem	Design a device to extract orange juice from fresh oranges at home.
Sample Customer Needs	Target cost - \$20. Easy to manufacture – do not use more than 2 manufacturing processes. Machine washable – no significant damage after 100 washes. Small footprint – fit within a volume of 15×15×15 cm ³ .
Other relevant data	Do not use blenders or blender-type machines as a base for designs.

Figure 4. Detailed quantitative requirements provided in briefs B, D, F and H

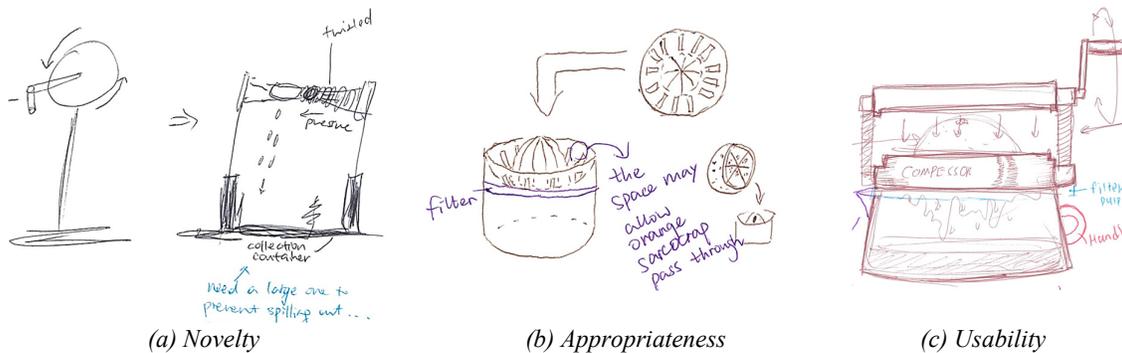


Figure 5. Examples of sketches with high scores for each of the three creativity metrics

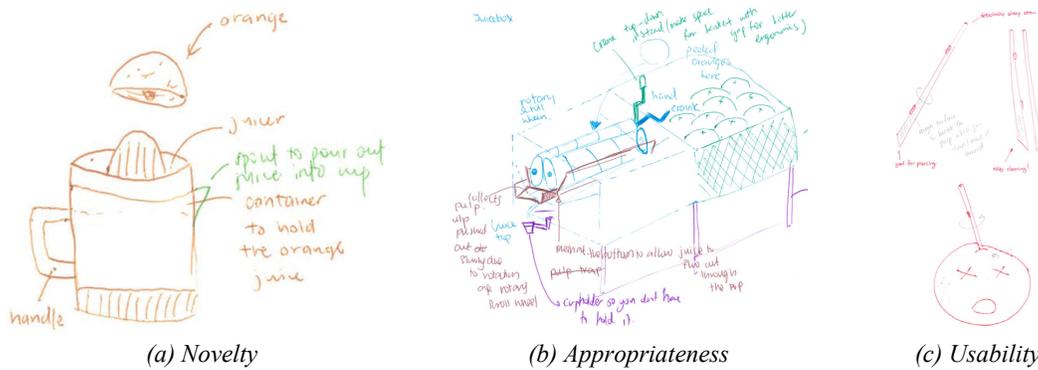


Figure 6. Examples of sketches with low scores for each of the three creativity metrics

2.4. Creativity assessment

Our experimental data was evaluated in consideration of Teresa M. Amabile’s consensual technique of creativity assessment (Amabile, 1996). Based on several studies of a group of “expert judges,” the evaluation of creativity is reliable and consistent with expert judges (Amabile, 1996; Baer, 2008). Therefore, we recruited judges who are engineering design or product development experts who have at least 3-4 years of design education and involvement in product development projects. The three judges independently evaluated all concept drawings using a rubric-based system with 1 to 5 Likert-type items (seen Table 2). These guidelines help the judges to evaluate drawings for the three metrics of novelty, appropriateness, and usability according to a scale from 1 to 5. We further assume that the judges share similar understandings of all three creativity metrics to a reasonable degree. We tested out the evaluation rubric during our pilot study and based on judges feedback we amended it so as to contain more detailed guidelines and descriptions. This assisted the judges during the creativity assessment process and improved our inter-rater reliability.

Table 2. Scoring rubric for the evaluation phase

Novelty	Score	Level	Example
The extent to which the design is different from the usual way of extracting juice.	1	Entirely similar	Copy of existing product
	2	Quite similar	Minimal differences
	3	Some similarities/differences	Average, no surprises
	4	Quite different	Minimal similarities
	5	Entirely different	The idea is a real surprise
Appropriateness	Score	Level	Example
The extent to which the design is aligned to the brief guidelines.	1	Not aligned	Does not meet the guidelines
	2	Slightly aligned	Falls in some guidelines
	3	Somewhat aligned	50/50 aligned to guidelines
	4	Mostly aligned	Meets most of the design guidelines
	5	Completely aligned	Fully aligned with the brief
Usability	Score	Level	Example
The ability of the design to efficiently extract the most amount of juice with minimal effort.	1	Poor	Cannot extract juice at all
	2	Fair	Little amount can be extracted
	3	Good	Extracts moderate amount of juice with moderate effort
	4	Very good	Easily squeezes out juice/ comfortable handling
	5	Excellent	Effortless juice extraction and handling

2.5. Inter-rater reliability and consistency

Since we used Likert-type items in the evaluation, the Cronbach's alpha measures the correlation between judges' scores. The split-half, inter-judge reliability is described by Equation 1 below:

$$\alpha = \frac{p}{p-1} \left[1 - \frac{\sum_{i=1}^p \sigma_i^2}{\sigma_T^2} \right] \quad (1)$$

The α coefficient of reliability ranges from 0 to 1. Should items be entirely independent of one another, i.e. they are not correlated nor share covariance, then $\alpha = 0$. The Cronbach's alpha based on standardized items ($n=522$) was 0.66 for novelty, 0.69 for appropriateness and 0.49 for usability. In general, values over 0.5 and close to 0.6 are considered acceptable with moderate reliability (Hinton et al., 2004).

3. Results and discussion

The experimental layout along with the average scores for each factor are summarized in Table 3. The coded format means that low levels of factors are denoted by “-” and high levels of factors are represented by “+”. The table also shows the mean score and standard deviation (Std) for each response. We can notice at this point that for novelty, none of the other briefs had a higher mean score than the control brief. The mean appropriateness score, on the other hand, improves by 20% when quantitative requirements and visual stimulus are provided together (Brief H). With the same combination of factors, the mean usability score can be improved by 16%. The row at the bottom summarizes the overall mean across all responses. Overall mean scores for novelty, appropriateness, and usability from all 8 experimental runs were 2.72, 2.81 and 2.76 respectively. Looking at Table 3, the highest novelty scores were found in control Brief A (3.02) followed by Brief B (2.98), highest appropriateness scores in Brief H (3.13) and highest usability scores in Brief H (3.02). The lowest scoring brief for both novelty and appropriateness is Brief D and the lowest scoring brief for usability is Brief A (2.62).

The mean scores for novelty, appropriateness, and usability with the standard deviation bars for each brief's responses are shown in Figure 7 on a scale from 1-5. It is also shown the three different responses in contrast to the control brief represented by a red dashed line.

Table 3. The 2-level full factorial experimental design layout

2 ³ full-factorial	Factors			Response Variables					
	Q	V	P	Novelty		Appropriateness		Usability	
				Mean	Std	Mean	Std	Mean	Std
Brief A	-	-	-	3.03	± 0.06	2.58	± 0.23	2.62	± 0.17
Brief B	+	-	-	2.98	± 0.17	2.77	± 0.48	2.64	± 0.18
Brief C	-	-	+	2.72	± 0.12	2.69	± 0.08	2.65	± 0.17
Brief D	+	-	+	2.46	± 0.04	2.52	± 0.28	2.80	± 0.04
Brief E	-	+	+	2.67	± 0.29	3.08	± 0.08	2.84	± 0.13
Brief F	+	+	+	2.52	± 0.12	2.96	± 0.22	2.69	± 0.19
Brief G	-	+	-	2.84	± 0.21	2.72	± 0.35	2.88	± 0.22
Brief H	+	+	-	2.51	± 0.23	3.13	± 0.20	3.02	± 0.04
Overall Mean				2.72	± 0.15	2.81	± 0.24	2.76	± 0.14

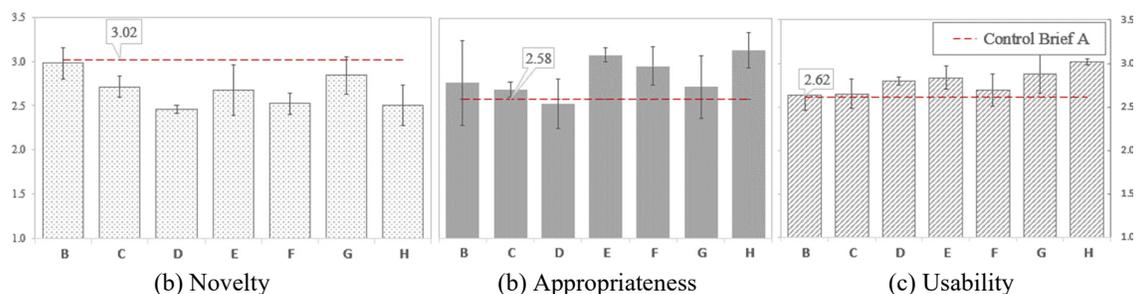


Figure 7. Charts of responses in comparison to control brief

3.1. Statistical analysis and interpretation of results

The next part involves analysis using Minitab software to evaluate the influence of the main effects and interaction effects. Figure 8 illustrates the Pareto plots of effects for all responses at 5% significance level. The grey bars indicate the absolute mean scores, and any effect that extends past the red reference line is potentially important for the model. The level of statistical significance (alpha level) was set at $p < 0.05$ with the corresponding confidence level at 95%. As expected, the 3-way interaction ($Q \times V \times P$) is not statistically significant for the mean response values.

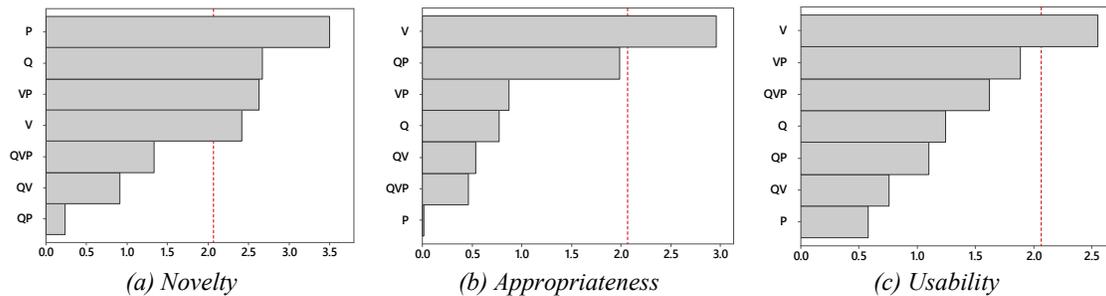


Figure 8. Pareto charts for the responses at $\alpha = 0.05$ confidence level

Novelty: The plot for Novelty shows that the main effects were statistically significant, namely Q ($p=0.021$), V ($p=0.013$), and P ($p=0.001$) at $\alpha=0.05$ confidence level. The interaction of visual and physical stimuli ($V \times P$) was also statistically significant ($p=0.013$).

Appropriateness: The assessment indicates that the main effect V is statistically significant ($p=0.005$). We repeated the regression, removing one insignificant term each time, until only statistically significant terms remained. However, to keep the model hierarchical none of the main effects were omitted. On the reduced model; in which the nonsignificant interactions are removed, $Q \times P$ is borderline significant at ($p=0.046$). It is interesting to note that although Q and P have a no impact on the response on their own, their effect on appropriateness is dependent upon their interaction $Q \times P$.

Usability: Figure 8c indicates that only the main effect V is statistically significant ($p=0.017$). On the reduced model, we eliminated terms from the model as explained above, and the only significant interaction, $V \times P$ is significant at ($p=0.026$) while the two remaining interaction effects had scores far beyond significance ($p > 0.3$).

3.2. Main effect and interaction plots

By plotting the means for each value of the variable for main effects and interaction effects, one can determine the best levels for each of the three factors. The charts in Figure 9 display the means for each group within a categorical variable. The steeper the slope of the line is, the greater the magnitude of the main effect.

For **novelty**, the plot of Figure 9a shows that P is the most dominant factor (significant at 5%), followed by Q and finally by V. Figure 9b proves that the second order interaction is strong among $V \times P$. As the main effect plots show, for novelty, it is best not to provide any additional information at all for all factors and to instead opt for the control condition (Brief A). This is also confirmed by the interaction plot in Figure 9b ($V \times P$), which implies that if a visual stimulus is given then a physical stimulus should be avoided as it will further decrease novelty.

The best factor settings indicated for novelty are the following:

- Quantitative: No (low-level)
- Visual: No (low-level)
- Physical: No (low-level)

In Figure 9c, the magnitude of the graph indicates that the visual stimulus has the most dominant effect on **appropriateness**, followed by quantitative and physical stimuli, with the last factor having little influence on the response. From the above three, only the visual stimulus is statistically significant ($p < 0.05$) among the main effects. Figure 9d indicates that there exists a strong interaction between

quantitative requirements and physical stimulus (Q×P), implying that the effects of quantitative requirements at different levels of physical stimulus vary. The interaction effect plot of Figure 9d also validates this setting; if we hold Q on a high level then providing P contributes to decreased appropriateness overall. If a physical example ought to be provided with the brief, then it will decrease appropriateness. Therefore if it is essential for P to be at a high-level, then Q should be kept at a low-level because if both Q and P are provided together (high levels), then the response will be further decreased. However, such a brief will still score higher than a brief in which no additional quantitative information or physical examples are provided.

The best settings of factors for maximizing appropriateness are as follows:

- Quantitative: Yes (high-level)
- Visual: Yes (high -level)
- Physical: No (low-level)

This factor set corresponds to Brief H, and this finding is in line with the results from Table 3 on which Brief H has the highest mean score for appropriateness.

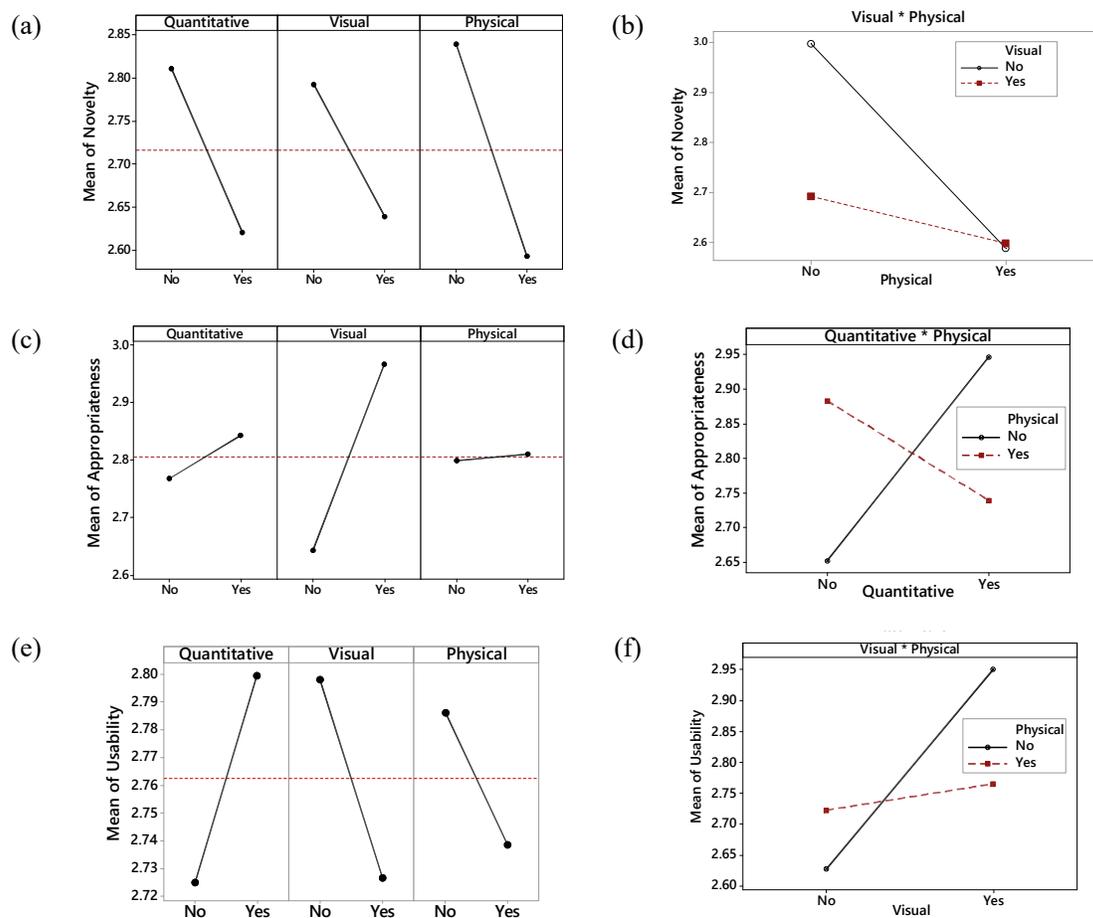


Figure 9. Main effect plots (left column) and interaction plots (right) for all factors

Figure 9e shows the main effect of *usability* with quantitative being the most active factor followed by visual and physical, but again we cannot interpret the main effects without considering their interactions. As mentioned previously in Section 3.1, only the visual effect was statistically significant. The interaction plot of Figure 9f shows an interaction between V×P, indicating that high means are obtained when visual is on high-level and physical is on low-level. The optimal factor settings for maximizing usability are the following:

- Quantitative: Yes (high-level)
- Visual: Yes (high-level)
- Physical: No (low-level)

This factor set corresponds to Brief H and is in line with the results from Table 3 on which Brief H holds the highest mean usability score compared to the other factors.

3.3. Effect of variability on creativity with analysis of variance (ANOVA)

We seek for the terms (main effects and second-order interactions) which are significantly related to differences in the variability of each response Std and produce less variable results. In Minitab® 17 software, this is done by using log transformations to model the standard deviation of the responses analyzed. ANOVA analysis was performed to investigate how the three variables - quantitative data, visual and physical stimuli- affect the variability of the response variables. A coded design matrix is used to determine which terms affect process variability. Novelty, appropriateness, and usability with both Std and Ln(Std) (natural logarithms of standard deviation) are calculated for each of the experimental conditions. First, the least squares estimation (LSE) is used to determine the factors which are significantly related to the responses of interest. In that way, we determine which ones to retain in our model retroactively. Next, we identify a reduced model with only the significant factors on which we used the maximum likelihood estimation (MLE) to determine the most precise effects and their coefficients. Table 4 presents the results of this reduced model and the MLE coefficients and show the changes of the effects which are associated with changes in the response variable. At this point, we note that in most cases all main effects were retained to keep the model hierarchical, but non-significant lower order interactions were omitted.

For **novelty**, the results indicate that the main effects are not statistically significant for variability. The MLE analysis showed that only Q×V is statistically significant at 0.05 α -level (p=0.04). The factorial plots of Figure 10a show that best levels of minimum variability in novelty obtained by keeping Q and P in low levels. As for **appropriateness**, the results indicate that the main effects Q and P are the most critical factors affecting variability. Moreover, minimum variation is attained when we keep Q at low and P at its high-level setting as seen on the plot for AStd on Figure 10b.

It was a surprise to observe from Table 4 that combined factor effect Q×V×P (p=0.003) is the most dominant factor AS which causes variation in **usability**, followed by V×P (Figure 10c) and lastly Q. Furthermore, the main effect plot on Figure 10c points out that minimum variation is obtained when Q is kept at a high-level setting.

Table 4. Maximum likelihood coded coefficients of Ln(Std)

Response	Term	Effect	Ratio	Coeff	P-Value
Novelty- Ln(NStd)	Q	0.242	1.273	0.121	0.402
	V	0.523	1.688	0.262	0.070
	P	-0.099	0.584	-0.049	0.733
	Q×P	-0.537	0.551	-0.269	0.049*
	LSE Summary: R-sq=98.96%; Rsq(adj)=96.36%; R-sq(pred) 82.80%				
Appropriateness - Ln(AStd)	Q	0.688	1.990	0.344	0.012*
	V	-0.137	0.872	-0.068	0.616
	P	-0.767	0.464	-0.384	0.005*
	LSE Summary R-sq=81.15%; Rsq(adj)=67.02%; R-sq(pred) 24.61%				
Usability - Ln(UStd)	Q	-0.634	0.531	-0.307	0.034*
	V	0.005	1.005	0.003	0.986
	P	-0.170	0.844	-0.085	0.556
	V×P	0.677	1.868	0.338	0.019*
	Q×V×P	0.848	2.351	0.427	0.003*
LSE Summary: R-sq=98.70%; R-sq(adj)=95.46%; R-sq(pred) 79.25%					

* Statistically significant values at $\alpha=0.05$

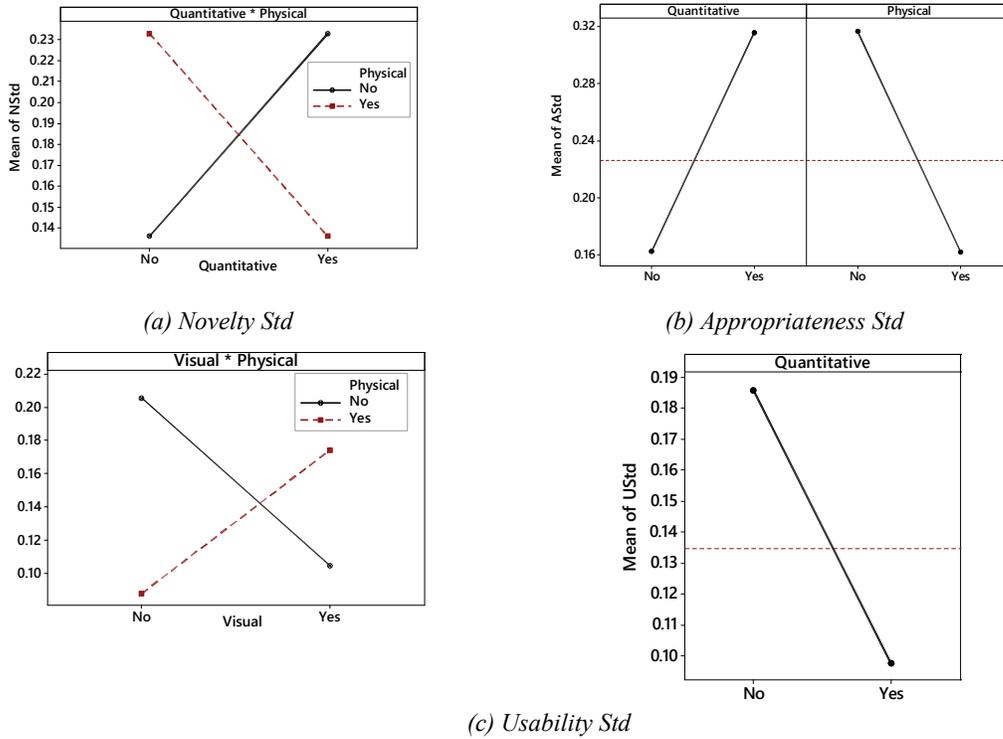


Figure 10. Main effect plots for the statistically significant mean of means (Std)

3.4. Multiple response optimizer

Using the composite desirability (D) and optimizer plots seen in Figure 11, we evaluate which settings optimize our set of responses. Optimizer searches for a combination of input variables that jointly optimize a set of responses by satisfying the requirements for each response in the set. At this point, we set our target to maximize all three responses on Figure 11a and then minimize their standard deviation (Std) in Figure 11b. To accomplish these goals, the optimizer found that the combination of factor levels that produce the best results within an acceptable range for each response using the regression model created from the replicate points we keyed in. The text displayed at the top of each column shows the current factor level settings that maximize novelty, appropriateness and usability scores (in red) which set to fulfill our two goals. The optimization plot shows the effect of each factor (columns) on the responses or composite desirability (rows). The horizontal blue lines and numbers represent the responses for the current factor level.

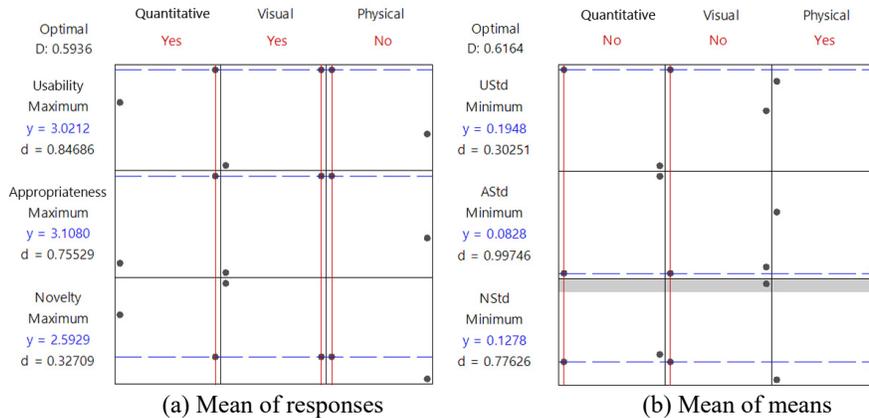


Figure 11. Response optimizer plots with optimal solutions

The results suggest that the brief for maximizing all three responses mean values is the Brief H (D=0.59), which has the following factor levels: Quantitative=Yes, Visual=Yes, Physical=No. For a brief with minimum Std the best brief is Brief C (D=0.61) in the following settings: Quantitative=No, Visual=No, Physical=Yes. In a second run, when choosing for optimum response means and minimum variation, Minitab calculates that design outcome is maximized as a whole when all factors are at their lowest settings (D=0.51), namely Brief A.

4. Conclusions and limitations

This study identifies which types of input data influence different components of the design outcome of student designers during a C-Sketch activity. The factorial ANOVA results show that the inclusion of visual and physical examples in the design brief lowers the novelty scores of student concept drawings. Therefore, it appears to be more effective to give less specific requirements and no examples when opting for novel idea generation. The novelty results are consistent with other studies which show that examples of existing solutions can cause design fixation (Smith et al., 1993) and cognitive fixation which results in replicated features (Goldschmidt and Smolkov, 2006). On the other hand, numerical/quantitative data improves usability and appropriateness, which suggests that it is essential to introduce good examples of successful products and more detailed specifications in the design process to ensure that students' concepts are aligned with the brief guidelines. Appropriateness and usability scores were highest when both quantitative product requirements and the visual example were present in the design brief.

With the composite desirability tool, we evaluated how well the settings optimize our set of responses overall considering collectively all the responses mean values and their standard deviation. We therefore suggest that providing student designers with minimum quantitative requirements and stimuli can foster creativity overall and enable them to come up with more original and workable ideas in our setting. However, if creating an appropriate design is the most important priority, providing visual examples and quantitative requirements may improve alignment with the brief guidelines, and then it shall be essential to introduce this set of stimuli to the students.

Weighing and evaluating metrics and requirements over hundreds of drawings may cause fatigue and there may be inaccuracies to judgment. We allowed judges a flexible and manageable timeline to evaluate so as to reduce the human errors of judgment that are made during this process. Ultimately based on research studies, higher values of the Cronbach's alpha (between 0.6 to 0.9 coefficient) are recommended for increased reliability (Nunnally, 1967; Nunnally and Bernstein, 2010). For this reason, we intend to improve these values in future studies at first by increasing the number of judges and providing them with a further updated rubric including technical aspects, aesthetic appeal, etc.

Acknowledgment

This work is supported by the Singapore University of Technology and Design (SUTD) and the SUTD-MIT International Design Centre (IDC, idc.sutd.edu.sg), project IDG 11600102. We are grateful to the participating batch of first-year SUTD student and teaching assistants. The authors acknowledge the assistance of their IDC colleague/friend Pei Zhi Chia that improved the manuscript.

References

- Amabile, T.M. (1982), "The Social-Psychology of Creativity: A Consensual Assessment Technique", *Journal of Personality and Social Psychology*, Vol. 43 No. 5, pp. 997-1013. <https://doi.org/10.1037/0022-3514.43.5.997>.
- Amabile, T.M. (1996), *Creativity in Context*, Westview Press, Boulder, CO.
- Baer, J. (2008), "Commentary: divergent thinking tests have problems, but this is not the solution", *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 2 No. 2, pp. 89-92.
- Binyamin, G. and Carmeli, A. (2017), "Fostering members' creativity in teams: The role of structuring of human resource management processes", *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 11 No. 1, pp. 18-33. <https://doi.org/10.1037/aca0000088>
- Bonnardel, N. (2000), "Towards understanding and supporting creativity in design: analogies in a constrained cognitive environment", *Knowledge-Based Systems*, Vol. 13 No. 7, pp. 505-513. [https://doi.org/10.1016/S0950-7051\(00\)00067-8](https://doi.org/10.1016/S0950-7051(00)00067-8)

- Bourgeois-Bougrine, S., Buisine, S., Vandendriessche, C., Glaveanu, V. and Lubart, T. (2017), "Engineering students' use of creativity and development tools in conceptual product design: What, when and how?", *Thinking Skills and Creativity*, Vol. 24, pp. 104-117. <https://doi.org/10.1016/j.tsc.2017.02.016>
- Chen, B.-B. (2016), "Conscientiousness and everyday creativity among Chinese undergraduate students", *Personality and Individual Differences*, Vol. 102, pp. 56-59. <https://doi.org/10.1016/j.paid.2016.06.061>
- Cropley, D. and Cropley, A. (2005), "Engineering Creativity: A Systems Concept of Functional Creativity", In: Kaufman, J.C. and Baer, J. (Eds.), *Creativity across domains: Faces of the muse*, Lawrence Erlbaum Associates, Mahwah, NJ, US, pp. 169-185.
- Demirkan, H. and Afacan, Y. (2012), "Assessing creativity in design education: Analysis of creativity factors in the first-year design studio", *Design Studies*, Vol. 33 No. 3, pp. 262-278. <https://doi.org/10.1016/j.destud.2011.11.005>
- 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. <https://doi.org/10.1016/j.destud.2006.01.002>
- Hinton, P.R., McMurray, I. and Brownlow, C. (2004), *SPSS explained*, Routledge, London, UK.
- Howard, T.J., Dekoninck, E.A. and Culley, S.J. (2010), "The use of creative stimuli at early stages of industrial product innovation", *Research in Engineering Design*, Vol. 21 No. 4, pp. 263-274. <https://doi.org/10.1007/s00163-010-0091-4>
- ISO (2013), ISO/TS 20282-2:2013 Usability of consumer products and products for public use - Part 2: Summative test method, ISO, p. 50.
- Kampylis, P.G. and Valtanen, J. (2010), "Redefining Creativity — Analyzing Definitions, Collocations, and Consequences", *The Journal of Creative Behavior*, Vol. 44 No. 3, pp. 191-214. <https://doi.org/10.1002/j.2162-6057.2010.tb01333.x>
- Kim, E. and Kim, K. (2015), "Cognitive styles in design problem solving: Insights from network-based cognitive maps", *Design Studies*, Vol. 40, pp. 1-38. <https://doi.org/10.1016/j.destud.2015.05.002>
- Nielsen, J. (1993), *Usability Engineering*, Academic Press, Cambridge, MA.
- Nunnally, J.C. (1967), *Psychometric theory*, 1st ed., McGraw-Hill, New York.
- Nunnally, J.C. and Bernstein, I.H. (2010), *Psychometric theory*, Tata McGraw-Hill, New Delhi.
- 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", *Research in Engineering Design*, Vol. 24 No. 1, pp. 65-92. <https://doi.org/10.1007/s00163-012-0138-9>
- Onarheim, B. (2012), "Creativity from constraints in engineering design: lessons learned at Coloplast", *Journal of Engineering Design*, Vol. 23 No. 4, pp. 323-336. <https://doi.org/10.1080/09544828.2011.631904>
- Ryd, N. (2004), "The design brief as carrier of client information during the construction process", *Design Studies*, Vol. 25 No. 3, pp. 231-249. <https://doi.org/10.1016/j.destud.2003.10.003>
- Sauer, J. and Sonderegger, A. (2009), "The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion", *Applied Ergonomics*, Vol. 40 No. 4, pp. 670-677. <https://doi.org/10.1016/j.apergo.2008.06.006>
- Shah, J.J., Smith, S.M. and Vargas-Hernandez, N. (2003), "Metrics for measuring ideation effectiveness", *Design Studies*, Vol. 24 No. 2, pp. 111-134. [https://doi.org/10.1016/S0142-694X\(02\)00034-0](https://doi.org/10.1016/S0142-694X(02)00034-0)
- Smith, S.M., Ward, T.B., and Schumacher, J.S. (1993), "Constraining effects of examples in a creative generation task", *Memory and Cognition*, Vol. 21 No. 6, pp. 837-845.
- Storme, M., Myszkowski, N., Çelik, P. and Lubart T. (2014), "Learning to judge creativity: The underlying mechanisms in creativity training for non-expert judges", *Learning and Individual Differences*, Vol. 32, pp. 19-25. <https://doi.org/10.1016/j.lindif.2014.03.002>
- Toh, C.A. and Miller S.R. (2015), "How engineering teams select design concepts: A view through the lens of creativity", *Design Studies*, Vol. 38, pp. 111-138. <https://doi.org/10.1016/j.destud.2015.03.001>

Dr. Georgios Koronis, Postdoctoral Research Fellow
 Singapore University of Technology and Design, International Design Centre
 33B Siglap Hill, 456100 Singapore, Singapore
 Email: georgios_koronis@sutd.edu.sg