**Proceedings of the ASME 2019**
**International Design Engineering Technical Conferences**
**and Computers and Information in Engineering Conference**
**IDETC/CIE2019**
**August 18-21, 2019, Anaheim, CA, USA**

**IDETC2019-97819**

# ENGINEERING DESIGN CONCEPT GENERATION: THE EFFECT OF CONCEPT COMBINATION AND CLASSIFICATION (DRAFT COPY)

**Saurabh Deo[1], Katja Hölttä-Otto**
Aalto University
Espoo, Finland

**Yogesh Bhalerao, Abhijeet Malge**
MIT Academy of Engineering
Pune, India

## ABSTRACT

*World economic forum reported, creativity is one of the most sought after skills by employers globally. Preliminary research lead to multiple initiatives on enhancing creativity and innovation. To contribute in this field, we investigated the effect of two interventions on the creativity of undergraduate engineering students, particularly on engineering design concept generation. The primary focus of investigation was on assessing the effect of two interventions, 1) combining and 2) classifying concepts, on the originality and quantity of the concept produced.*

*In this research, we used the Decision Tree for Originality Assessment in Design (DTOAD) as a measure of concept originality. Statistical analysis showed that both the combine and classify interventions lead to more original concepts. During quantity assessment, we found students produced the higher quantity of radically different concept, i.e. concepts with originality score 7.5 and above, after interventions despite a decrease in overall quantity.*

*These interventions do work and thus can be encouraged as the part of an ideation method or an engineering problem solving task in undergraduate engineering education to help them develop creative skills.*

Keywords: Creativity, Originality, DTOAD, Quantity

## 1. CREATIVE SKILLS AND RESEARCH MOTIVATION

Creativity and innovation skills are essential elements to become a good product designer. Levitt quoted, "Creativity is coming up with new ideas and innovation is putting them to work' [1]. It is not uncommon, at first, if these words invoke the connection of thoughts to artistic fields since people often appreciate artistic creations using these words. However, over the years these words have translated from rather the 'expression of appreciation' to 'desired skills'. For example, a recent article in World economic forum (WEF 2019) shows the top ten demanded skills for 2019. Among the top five skills, the WEF pointed at the creativity of the employees as the highest ranked. [2] European commission launched the University of Future (UoF) project that aims to enrich overall education system in the Europe and one of the objectives is to accelerate the innovation practices. [3] Also, the Centre for Educational Research and Innovation (CERI) [4], the part of Organization for Economic Co-operation and Development (OECD), initiated a project on creativity and critical thinking skill building among students. The current interest on creative and innovative employees is significant. These earlier initiative are a source of motivation for this article.

As the importance of creative skills grew, scientific community started exploring techniques to assess the creativity. If one can measure creativity, it could potentially be improved. Researchers clustered the creativity assessment techniques in three broad fields. The creativity assessment of a person/personality, a product and the group of concepts. Each of these fields have developed its own the measures of assessments. [5] Creativity can be incorporated in many ways to product development and engineering education. We can make curriculum level changes, course improvement as well as teach small creativity improvement techniques. We are interested in creativity related to engineering design. Therefore, the focus of this article is on the product concept generation.

## 2. LITERATURE

Concept generation is the part of all product development. To aid in concept generation, numerous techniques are available in literature. In *'The whole brain business book'* alone, more than seventy ideation methods are listed. [6] To mention few, ideation techniques include Brainstorming [7], TRIZ [8], C Sketch/ 6-3-5 [9] etc. Each of these methods have their unique

[1] Contact author: saurabh.deo@aalto.fi

characteristics and the selection of suitable technique depends on the aim of ideation. Many methods requires participant to be in a group. Since this research focuses on individual level creativity rather than group level, and since we are investigating the interventions independent of any particular method, we selected simple ideation in this investigation instead of adopting any specific technique.

Furthermore, to enhance the concept generation during ideation, others have investigated the use of analogy [10-12] in which, external elements or distant connection with other similar concept aids in improving concept generation. Linsey et al. [13] showed that concept viewing and representation method affect ideation. These methods or interventions along with others as studied by Silk et al. [14] had positive influence on concept generation. In our research we introduce participants to two different interventions, 1) Combine and 2) Classify. In the combine intervention, the participants are expected to combine product concepts that they have themselves already created to form or inspire new concepts; and in the classify intervention, they are expected to categorize their initial concepts to produce new concepts. More details about these intervention can be found in Section 3.

In the last few decades, scholars developed several metrics to assess product concepts. Shah et al. [15] derived a metrics to assess variety, novelty and the quality of product concept. Variety refers to the number of different concepts proposed by any participant, novelty refers to how unique particular concept is within all sample, quality refers to the ability of concept proposed to meets design specification and quantity is total concepts produced. [15] Shah's Novelty metric measures novelty by identifying key features essential for its function. The measure combines the novelty of each feature into a concept level novelty value. Srivathsavai et al. [16] criticized the metric for only focusing on the novelty within the generated concepts and not novelty in reference to products already out in the market. Further, they found low inter rater reliability (r=0.24) for the metric. Sarkar et al. [17] argued against the metric's use of uniqueness to measure novelty. Many other limitations were also highlighted by Brown. [18]

Charyton et al. [19] developed Creative Engineering Design Assessment (CEDA) for measuring the originality, fluency i.e. amount of concepts and the flexibility which represent different types of concepts. Charyton et al. [19] claimed higher inter rater reliability (r=0.84) between two raters. But, during CEDA development the author did not share the number of concepts rated to achieve the inter rater reliability. To calculate originality, the metric uses an 11 point scale with words such as Dull showing lowest to Genius as the highest spectrum of scale. Kershaw et al. [20] also used CEDA to developed modified originality metrics. Furthermore, in another research Charyton et al. [21] reported lower r=0.59 with five raters. CEDA uses scale with words in the assessment which can vary as per the personal interpretation of words used in scale, and accordingly Genco et al. used a modified 5 point scale version of it. [22] Brown [18] agreed to this limitation along with few more.

Kershaw et al. [23] created the Decision Tree for Originality Assessment in Design (DTOAD). DTOAD (Figure 1) allows system level originality assessment rather than feature level by focusing on assessing if the innovation is beyond the industry norm and then how integral the innovation is to the concept. It uses a 5 point ordinal scale to represent lowest to highest originality score ranging from '0' to '10' with 2.5 increment between each level as shown in Figure 1. Kershaw et al. [23] achieved the inter rater reliability of r= 0.70 and validated DTOAD metric. DTOAD compares concepts with existing products in the market rather than comparing within sample alone. Due to provisions available in this metric, it presents itself as compelling alternative to previous metrics. Therefore we chose this metric in our study. We use it to study if the minor interventions of classifying or combining concepts during concept generation have an impact on concept originality.

Therefore, we raise two research questions:

1) Do combine and classify interventions aid in improving the originality of product concepts generated by undergraduate engineering students?

2) Can interventions affect the quantity of original concepts generated?

To answer these questions, we compare interventions against a control group. Detailed approach adopted for this study is explained in subsequent sections.

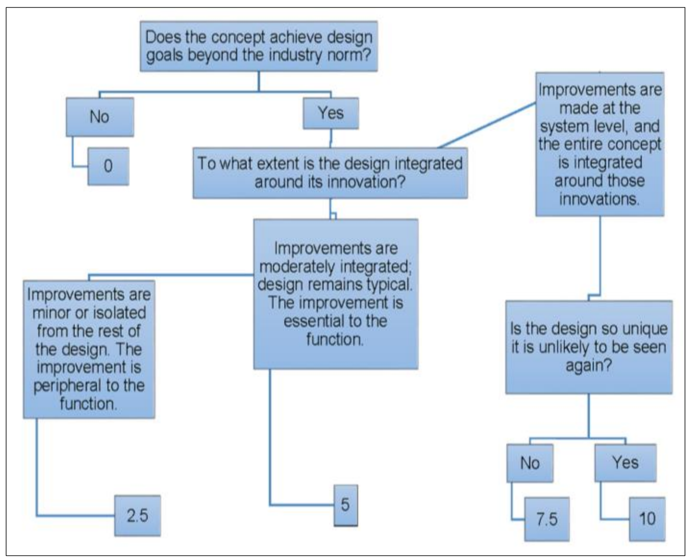


**FIGURE 1:** DECISION TREE FOR ORIGINALITY ASSESSMENT IN DESIGN (DTOAD) [23]

## 3. EXPERIMENTAL APPROACH

Experimental approach adopted for this research is detailed below and summarized in Figure 2. We follow a similar process used also by Kershaw et al. [24].

 

### 3.1 Participants

In this study, 57 freshman engineering students of the academic year 2019-20 from the Maharashtra Academy of Engineering (MIT AoE), Pune, India participated.

### 3.2 Execution Procedure

#### 3.2.1 Pre-experimentation

Participants were attending regular academic sessions and were not specifically instructed to participate in the research. The research theme and general purpose was conveyed, but no information about the experimental set up was revealed. They were verbally informed that participation was completely voluntary. The participants were provided a consent form. After obtaining consent, they were randomly assigned to three groups: control group, combine intervention group and classify intervention group. Participants were given the design challenge on a piece of paper and ideation template that included three pages. In subsequent sections, naming scheme as mentioned above has been used during comparisons. Experiment was executed in following sequence for each group separately.

**I:** We setup the room in preparation for the experiment. Sufficient student sheets, pens and markers were made available. A timer or stopwatch was also handy.

**II:** When participants arrived in the class, we assigned them unique identifying numbers, from henceforth termed only as 'Identifier'. Before the session, we created the master list of identifiers for all students so that data can be traced back to a participant number or a group participant belongs to. (No personal information such as Name or University PRN (Permanent Registration Number) were gathered in order to maintain the anonymity of participants).

**III:** Individual student identifier was written on each piece of paper and all questions were answered before proceeding.

#### 3.2.2 Design Challenge

The participants were asked to "propose concepts for a next generation garbage picker". Same example as in [24] was used. The design challenge did not have any design requirements. As done in previous studies, no reference example physically or in any other form of communication was shown. [24]

#### 3.2.3 Execution

One group (N=20) was assigned to the control condition. They were not subject to any specific intervention. The second group (N=19) was assigned to combine intervention and third group (N=18) to classify intervention. Details about both of these interventions and control condition are explained in subsequent section. Ideation session was done individually and therefore no sub groups within the each three group were formed. Each page on ideation template included written instructions about task to be completed on that page. We made sure to time each stage correctly and collect papers when the time is up.

**Control Condition:**

Stage 1: Each participant drew as many product concepts as he/she could for next generation garbage picker. They were allowed to use phrases or comments to help convey their but those had to be mainly represented through drawing. After 10 minutes Round 1 was complete and they flipped to page 2 for further instructions.

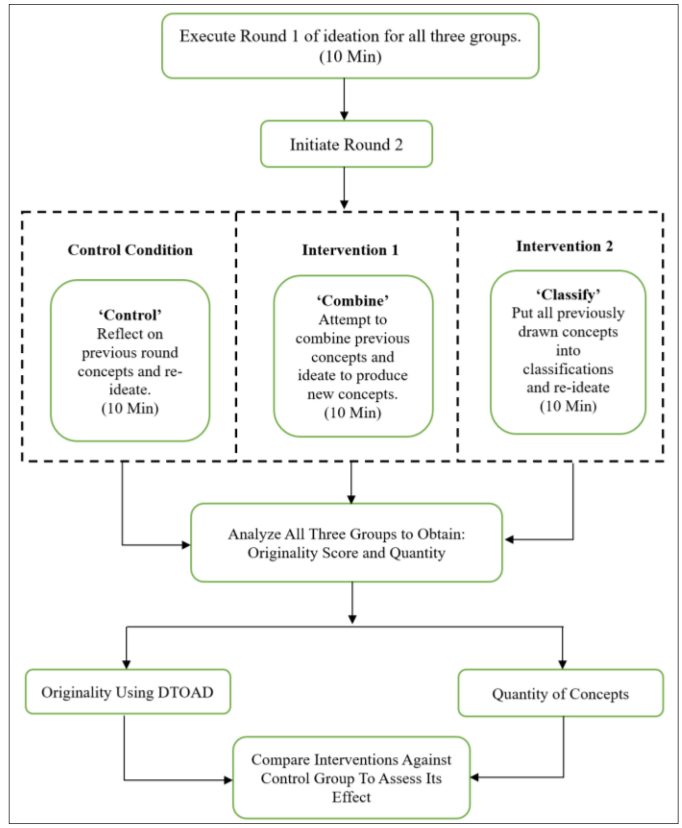


**FIGURE 2:** EXPERIMENTAL APPROACH

Stage 2: Page 2 instructed participants to reflect on their Round 1 product concepts, but not to write on this page. After reflection, everyone moves to page 3.

Stage 3: On page 3, participants repeat ideation to produce any new product concept. Here, we conclude Round 2. For Stage 2 and 3, total 10 minutes were allotted.

**Combine Intervention:**

Stage 1: The process described at Stage 1 of the control condition is repeated at this stage.

Stage 2: The process described at Stage 2 of the control condition is repeated at this stage.

Stage 3: Page 3 instructs participants to combine concepts from Round 1 and draw as many new product concepts as possible for given design challenge.

**Classify Intervention:**

Stage 1: The process described at Stage 1 of the control condition is repeated at this stage.

Stage 2: Page 2 instructed participants to classify their Round 1 product concepts into groups based on e.g. their similarity or the way of functioning etc. After classification, Round 1 concludes.

Stage 3: On page 3 participants again ideated to produce new product concepts for next generation garbage picker. For stage 2 and 3, ten minutes were allotted as in previous groups.

Each group had in total 20 minutes. Experiments were conducted separately for the three groups. Each participant was seated individually to avoid interactions with other participants. The method of individual ideation was a deliberate attempt to prevent unwanted thought stimulation leading to biased results.

### 3.2.3 Post Experimentation

At the end of Round 2, participants were thanked for assisting and were asked to refrain from discussing the details of this experiment or the concepts with their classmates. We also asked them not to discuss this experiment with future students as similar experiments might be carried out in the future.

## 4. ASSESSMENT

The concepts were assessed for originality using the *Decision Tree for Originality Assessment in Design (DTOAD) metric.* [23] Before assessment, we trained raters similar to the procedure used by Kershaw et al. [23]. We used two raters for this experiment. The raters had either a Master or a Doctorate degree in Mechanical Engineering. For rater training in applying DTOAD, we used 20 concepts from a different data set, but one that had the same design brief as this study. Coders completed three rounds of 20 concept each with discussion in between. After 3rd round Cohen's weighted kappa of 0.70 was achieved. The calculated Cohen's weighted kappa can be considered as 'Substantial' based on literature [25,26]. Disagreements between the raters on few concepts were overcome with rules both rater agreed upon. All 381 concept in experiment were then coded at the end of agreement by a single coder.

### 4.1 Originality Assessment

To demonstrate the originality assessment, representative cases from the study were chosen. One was rated 2.5 and the other 7.5. These two originality scores cover the second lowest to the second highest score in our samples. Concepts with scores 0 were omitted from demonstration, because they indicated the replication of commonly used products present in the market or contained non-design elements such as animals, human laborers etc. None of the concepts were rated 10.

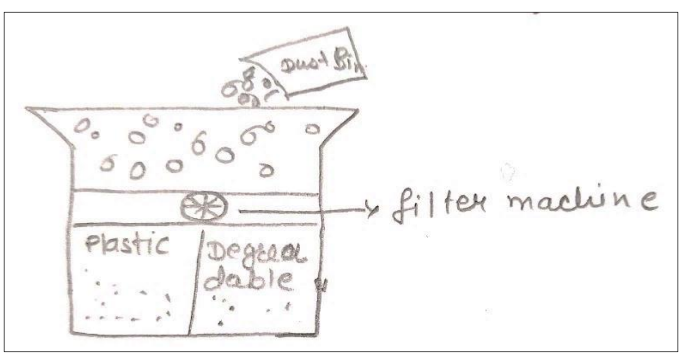


**FIGURE 3:** SAMPLE ILLUSTRATION 2.5 SCORE

With reference to Figure 3, the participant proposed a type of dustbin which has some kind of filter machine at the middle. This dustbin would sort and store the garbage as plastic and degradable separately in different compartment unlike manually done in most houses. Referring back to DTOAD tree (Figure 1), the first level states whether the concept achieves design goals beyond industry norms.

At the time of writing this article, our research did not show any product widely used similar to this concept. The feature of filter machine is additional functionality to dustbin which is beyond the current industrial practice of using two separate dustbins. Therefore it satisfies the condition to reach level 2. At this level we try to understand how well the design is integrated around innovation. This concept does add a feature which is not an industry norm. However, this feature is minor and one addition to existing product. Dustbin with inbuilt partitions for different garbage types is commonly available. Therefore, coders concluded that, this improvement is isolated from the rest of the design. As the product concept is not a norm in the industry and embodies minor improvement, hence, it was rated with a score of 2.5.

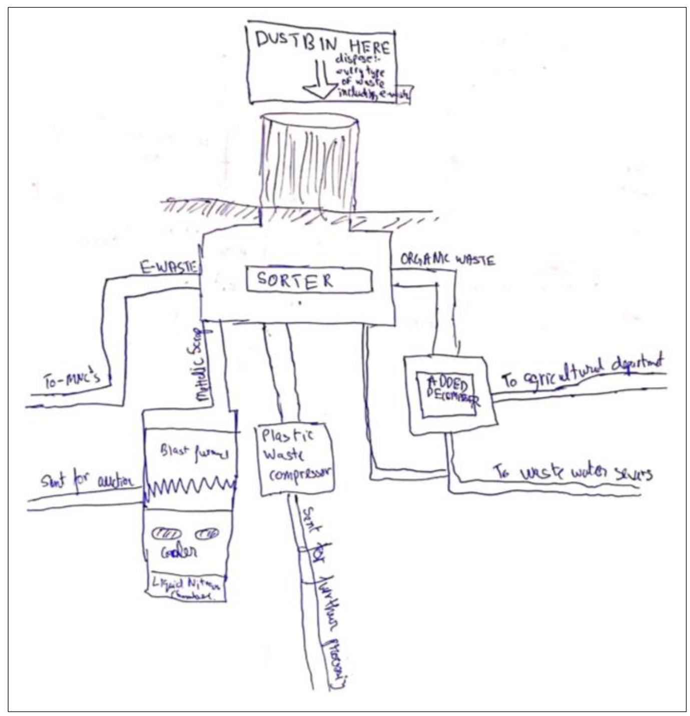


**FIGURE 4:** SAMPLE ILLUSTRATION 7.5 SCORE

Figure 4 shows a design which has multiple features and functionalities incorporated in it such as sorting waste into e-waste, organic, metallic and plastic waste. Garbage is further processed in appropriate manner, which involves burning, compressing, decomposing or simply transferring to waste sewers. Currently, transferring trash to landfills through underground piping system does exist, but such systems do not process it to the extent shown in this concept automatically. To design a system consisting of these features requires considerable system-level infrastructure improvement and the integration of current product and processes for garbage collection and transport. In all the concepts we rated, this concept was not repeated, but some other underground transport systems were found. Nevertheless, it is highly unlikely that this concept will never be seen again. Therefore, considering all the improvements made and integration demonstrated in concept, this concept was rated with a 7.5 score. This is how originality score was measured.

### 4.2 Quantity Assessment

Shah et al. [15] suggested, while calculating the quantity of concept produced, focus needs to be on all concepts documented by individual and not on non-repeated concepts alone since those are measure by variety. Therefore, we considered the overall the number of concepts produced by each group as well as each participant for quantity measurement.

## 5. RESULTS

The intent of this investigation was to compare the effect of two interventions on the engineering students' creativity, in terms of originality and the quantity of concepts produced. The results consist of two sections, first, we compare originality scores from Round 1 to 2. Then we calculate the quantity of concepts produced in each round per group, per student. Kershaw et al. [23] defined product concept radically different if it received a score of 7.5 or 10. We also compare the quantity of radically different concepts from Round 1 to 2.

### 5.1 Originality Analysis

At the beginning of the analysis, the originality scores of Round 1 for both interventions were compared with control group to identify if students performed equally across all groups. As there was no intervention introduced in Round 1, it was expected that students would perform similarly. The Round 1 originality scores violated the conditions of normality hence, a non-parametric test was conducted. Mann Whitney U test showed, originality score distribution across all groups in Round 1 were statistically significantly different ($p>0.05$) and that made them incomparable. This could be due to the factors beyond control of facilitators such as students' versatile background, experiences, other ambiguous factors or simply because of mistakes that went undetected during the execution of experiment. Even though comparison between interventions was not possible, within subject statistical analyses were possible.

Descriptive statistics for groups are in Table 1. It shows the number of students as 'N' per group and the number of concepts in that particular category as 'n'. We can also observe mean originality score for control (M=2.333), combine (M=1.571) and classify (M=1.891) groups. Numerically, difference in mean originality between the two rounds for each group is visible. However, appropriate statistical analysis would confirm whether this difference is statistically significant.

We started the analysis with the control group. Round 1 and Round 2 concepts were checked for normality using SPSS. Concept from Round 1 and 2 had the negative skewness of 0.540 and 0.301 respectively. Test statistics are listed in Table 1. Data was non-parametric and voids the fundamental assumption for t test, therefore, it could not be used to check significance. To find any statistically significant difference in originality between Round 1 and 2 for each group, non-parametric Mann Whitney U test was conducted. It should be noted that, Mann Whitney U test compares the mean rank between two independent variables not mean or median values, in our case originality score. [27]

Null hypothesis for Mann Whitney U test was, there is no statistically significant difference between mean rank between Round 1 and 2. This test, unlike t distribution in t test, follows Z distribution. [27] Table 2 shows Mann Whitney U test statistics. We can numerically compare unit difference in mean ranks between rounds. For control group $p = 0.406$ ($>0.05$) which is considerably above the acceptable norm. This implies, from Round 1 to 2, statistically insignificant difference existed in the mean rank originality and null hypothesis holds true. In other words, Control condition did not assist the students in producing more original concepts in Round 2.

Mean originality score in a bar chart with standard error and 95% confidence interval for control group is show in the Figure 5. Standard error was, 0.331 for Round 1 and 0.475 for Round 2.

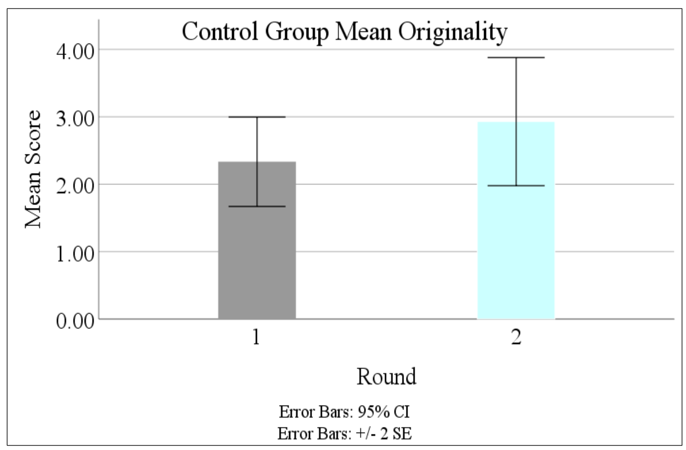


**FIGURE 5:** CONTROL GROUP ORIGINALITY COMPARISON

**Table 2**
**Mann Whitney U Test Statistics**

| | Mean Ranks | | Asymp. Sig. (2-tailed) |
|---|---|---|---|
| | Round 1 | Round 2 | |
| **Control** | 38.69 | 42.83 | 0.406 |
| **Combine** | 54.85 | 68.41 | 0.023 |
| **Classify** | 81.13 | 108.20 | 0.000 |

 

**Table 1**
**Descriptive Statistics for Control Group and Combine, Classify Interventions**
**(N= No of participant) (n = No. of concepts within group)**

| | Round 1 | | | | Round 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean Originality (M) | Standard Deviation (SD) | No. of Participants (N) | No. of Concepts (n) | Mean Originality (M) | Standard Deviation (SD) | No. of Participants (N) | No. of Concepts (n) |
| **Control** | 2.333 | 2.223 | 20 | 45 | 2.928 | 2.810 | 20 | 35 |
| **Combine** | 1.571 | 1.958 | 19 | 70 | 2.650 | 2.595 | 19 | 50 |
| **Classify** | 1.891 | 2.186 | 15 | 115 | 3.219 | 2.471 | 15 | 66 |
| **Total** | | | **57** | **230** | | | **57** | **151** |

For the combine intervention group, normality test showed negatively skewed data with the value of 0.955 and 0.560 in Round 1 and 2 respectively. Therefore data was not suitable for the t test of significance. Non parametric analysis using Mann-Whitney U test showed, difference in means ranks between Round 1 and 2 were statistically significant with $p = 0.023$ ($p < 0.05$). Hence, we reject the null hypothesis of equal mean ranks. This also means that, the combine intervention indeed made a difference in increasing the originality of concepts in Round 2. Bar chart is shown in Figure 6. Standard error in Round 1 and Round 2 was 0.234 and 0.367, respectively.

Finally, we investigated the classify intervention group data for normality. Round 1 followed similar trend with the negative skewness of 0.739 and but Round 2 was positively skewed with 0.075 skewness. Non-parametric Mann-Whitney U test shows statistically significant difference between Round 1 and 2. In fact, highest significance was found in this group with p = 000. Figure 7 shows, mean originality score bar chart with the standard error of 0.203 and 0.304.

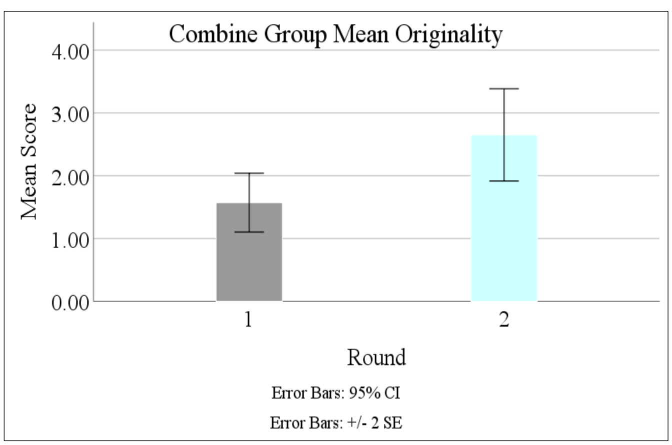


**FIGURE 6:** COMBINE GROUP ORIGINALITY COMPARISON

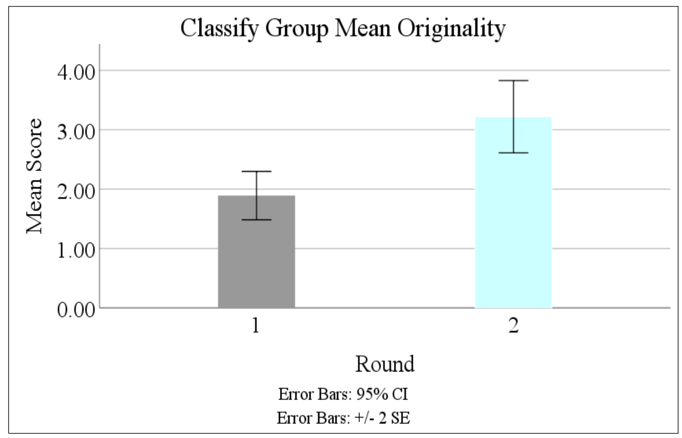


**FIGURE 7:** CLASSIFY GROUP ORIGINALITY COMPARISON

### 5.1.1 Effect size estimator

Albeit, these tests of statistical significance depict the means ranks as being statistically different or not, the real life impact can be measured by using effect size estimators. Effect size can be calculated using z distribution score obtained in Mann-Whitney U test. [27] Leech et al. [28] emphasized that, for non-parametric results, reporting effect size is statistically equally vital as for parametric results and claimed the most of the research often does not report the effect size or does it incorrectly.

We calculated the effect size estimator, known as Eta Squared ($\eta^2$) using the guideline by Cohen [29,30] for all three interventions using the Z statistics obtained during analysis. Z statistics and effect size are shown in the Table 3.

$$\eta^2 = \frac{Z^2}{n-1} \quad (1)$$

Where,
n : No of concept in each intervention
Z: Z distribution score

**Table 3**
**Effect Size Estimations for Mann Whitney U Test**

| | Z Score | n | $\eta^2$ |
|---|---|---|---|
| Control | -0.831 | 80 | 0.008 |
| Combine | -2.269 | 120 | 0.043 |
| Classify | -3.548 | 181 | 0.069 |

Effect size estimation implies, less than the 1% (0.008) of the variance can be explained for the given population due to control condition. But the variance of 4.3% in dependent variable can be explained as due to combine intervention. Classify interventions showed the highest effect size estimation of 6.9% ~ 7% and can explain highest variance. As per the Cohen's rule of thumb [29,30], anything below 0.3 is not large enough effect. From 0.3 to 0.5 is medium effect and anything above 0.5 is considered as significant effect size. Graphically, this is shown in Figure 8.

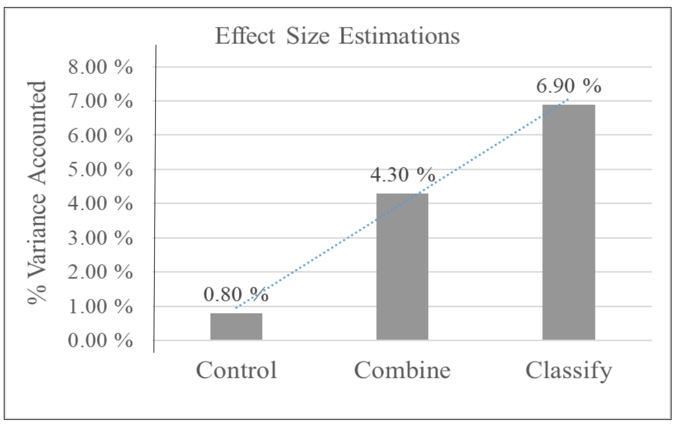


**FIGURE 8:** EFFECT SIZE ESTIMATIONS

## 5.2 Quantity Assessment

In this section**,** the quantity of concepts produced by each group as well as individual student are discussed. The comparison of the number of product concepts proposed from Round 1 to 2 are shown in Sections 5.2.1 and 5.2.2.

### 5.2.1 Group level quantity generation

The comparison of total quantity of concept produced by each group is shown in Figure 9. Control group produced 80 concepts including Round 1 and Round 2, with each round contributing 45 and 35 concepts respectively. Combine group shows slightly better ideation with each round contributing 70 and 50 concepts respectively. In total combine group produced 120 concepts. Classify group, which was asked to group concepts into suitable categories produced the highest count of concepts in both the rounds. In total, this group contributed 181 concepts. Round 1 have 115 and Round 2 have 66 concepts. Table 4 shows, the mean quantity of concepts produced by each group.

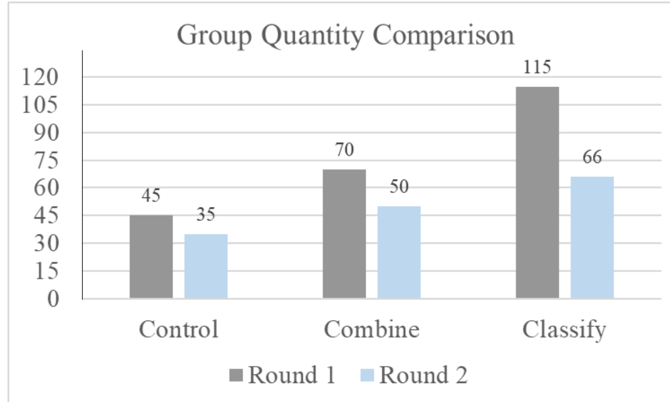


**FIGURE 9:** GROUP QUANTITY GENERATION

Classify group students produced on average six product concepts in Round 1. Other two groups produced on average three and two product concepts per student.

With reference to pie chart in Figure 10, at a glance the distribution of product concepts between Round 1 and 2 across all the groups can be seen. Round 1 contributed approx. 60% (+/-4) of all the concepts. Round 2 nearly 40% (+/-4) for all conditions.

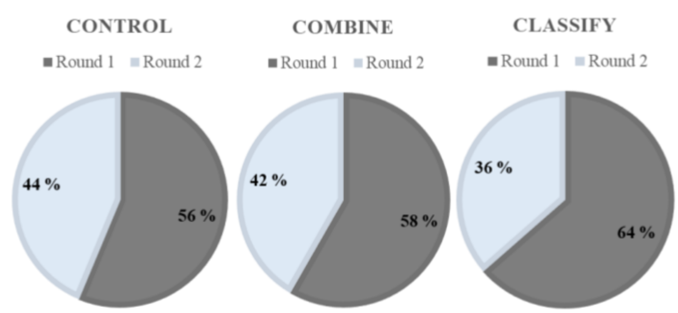


**FIGURE 10:** PIE CHART SHOWING QUANTITY DISTRIBUTION

**Table 4**
**Summary of Group Quantity Generation**

| | Mean | | Std. Deviation | |
|---|---|---|---|---|
| | Round 1 | Round 2 | Round 1 | Round 2 |
| Control | 2.250 | 1.750 | 1.070 | 0.910 |
| Combine | 3.684 | 2.632 | 1.600 | 1.422 |
| Classify | 6.389 | 3.722 | 2.033 | 2.052 |

## 5.2 Participant level quantity assessment

Group quantity assessment helped us briefly glance at the overall performance of each group. This section details individual level contribution. Figure 13 to 15 shows individual level concept generation.

Control group students produced fewer concepts in both Round 1 and 2. The highest number of concepts by any individual student from control group was four in Round 1 and three in Round 2. Combine group performed slightly better than the control group. The highest number of concepts produced by any individual for given design challenge in Round 1 was 6 and 5 in Round 2. Within classify group, highest contribution by any individual was 11 concepts in Round 1 and 7 for Round 2. During statistical analysis it was revealed that, data was mostly negatively skewed towards '0' score. It is understandable that not many students can produce very original concepts at first. We further investigated if, interventions contributed to reducing the quantity of non-original concepts scoring '0' or '2.5' and increase the quantity of radically different [23] product concepts scoring 7.5 and above. Table 5 shows the summary of concepts produced in each round with rating from 0 to 7.5, recall there was no concept with 10 score.

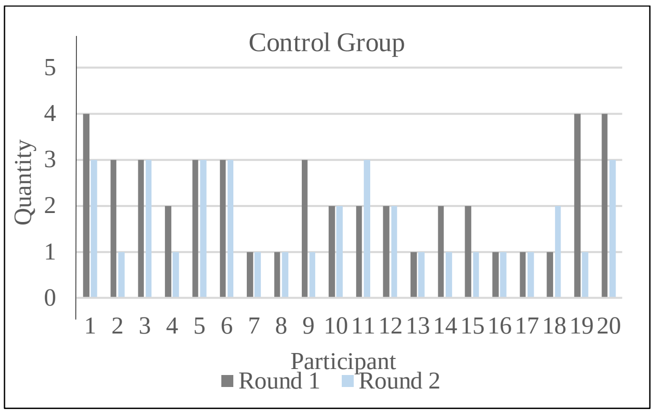


**FIGURE 13:** PARTCIPANT LEVEL QUANTITY BY CONTROL

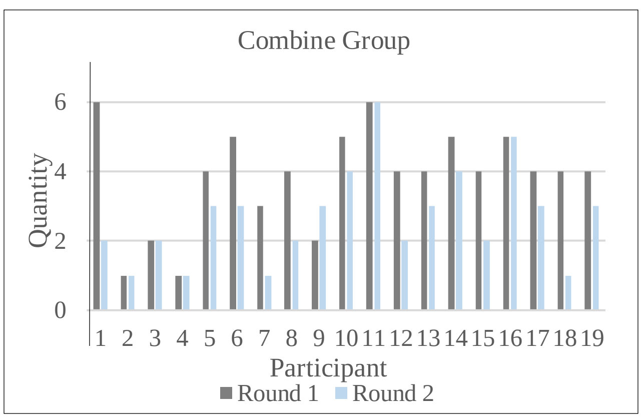


**FIGURE 14:** PARTCIPANT LEVEL QUANTITY BY COMBINE

Table 5 shows an interesting comparison, in all the groups we investigated, non-design or already existing product concepts with the score of 0, reduced considerably in Round 2. Similar was the case with concepts scoring 2.5. Concepts which scored 5 however, were almost identical except minor drop in classify group.

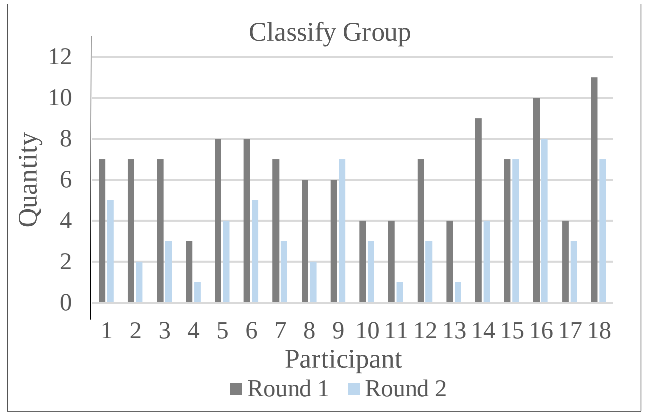


**FIGURE 15:** PARTCIPANT LEVEL QUANTITY BY CLASSIFY

**Table 5**
**Quantity Comparison at Each DTOAD Scale**

| DTOAD Scale | Control | | Combine | | Classify | |
|---|---|---|---|---|---|---|
| | Round | | Round | | Round | |
| | 1 | 2 | 1 | 2 | 1 | 2 |
| **0** | 17 | 14 | 39 | 19 | 59 | 20 |
| **2.5** | 16 | 6 | 20 | 15 | 30 | 18 |
| **5** | 10 | 10 | 10 | 10 | 23 | 22 |
| **7.5** | 2 | 5 | 1 | 6 | 3 | 7 |
| **10** | 0 | 0 | 0 | 0 | 0 | 0 |

It must be realized that, in Round 2, the total quantity of concepts reduced considerably compared to Round 1, yet, original concepts increased perhaps due to interventions. These findings shows, interventions have a positive impact on radically different concept [23] (originality score ≥ 7.5) generation.

## DISCUSSION

In this article, we answered two research question. Firstly, we investigated the effect of two interventions, combine and classify on the originality score using the Decision Tree for Originality in Assessment in Design (DTOAD) [23]. The outcome showed that interventions have a statistically significant effect in the originality of concepts generated. Secondly, we assessed the effect of interventions on quantity. Our investigation found that the quantity did not increase from Round 1 to 2, however, the quantity of radically different concepts with originality score 7.5 and above improved post intervention.

Combine intervention specifically asked students to combine the concepts and classify to distribute concepts into classifications during ideation. In control condition, student simply reflected on their concepts. Each intervention was designed to, sort of, channelize the thinking of student and test if it helps in concept generation. Unfortunately, we could not compare which one works best because Round 1 results before any intervention were statistically different. There could be a number of explanations for this difference. One plausible reason is the motivation as suggested by Bergendahl et al. [31]. We did not have provision to give credits or any direct academic benefits to students which might have affected motivation to contribute despite research being voluntary.

Further, we conducted within group investigation to measure intervention's effect. Investigation showed statistically significant difference in the mean rank of concept originality from Round 1 to 2. For combine intervention significance was high ($p<0.05$) and for classify very high ($p<0.01$). However, control group did show similar results ($p>0.05$) for both rounds. From statistical significance, classify seems to works better in more original concept generation. From results we can also assume, these two interventions allow students to look at their own concepts from unique perspectives, stimulating thought process. Chan et al. [32] suggested distant inspiration sometime gives best ideas. Perhaps, classifying or combining concepts stimulated the mechanism of concept generation similar to Chan's and other studies in literature [11-13]. We also calculated effect sizes and found that classify had the highest impact followed by the combine condition. Control had the smallest effect size. However small, nevertheless, it is an impact and these two interventions could be combined with an existing creativity method to augment engineering design concept generation. This addresses the first research question.

Kudrowitz et al. [33] claimed, higher the quantity of concepts, higher will be the creativity. Since student in this group produced the highest count of concepts, this might be the reason for classify group to show very high statistical significance. Interestingly, quantity measurement showed, despite decrease in the overall quantity of concepts produced in Round 2, radically different concept [23] (originality score $\geq 7.5$) increased. In other words, reduced the quantity of concepts in Round 2 could be mostly non-original concepts. Here, we have answered second research question. It is also possible that unequal quantity may be due to the diverse backgrounds, culture or the exposure of students. It would be interesting to group the students from similar background together and repeat the experiment to better understanding the role of culture or background on original concept generation.

During the assessment of concepts using DTOAD, we experienced similar challenges as in [23]. Concepts at both top and bottom end were quite easy to rate however, most difficult were 2.5 to 5 score because, each coder had their own perception about what is as moderately integrated or isolated. We used coders from two different countries and we found, it was quite challenging to reach r= 0.70 as recommended by Cohen. [25]. In the future, analysis for concept feasibility and variety analysis could add another layer of confidence in interventions as useful tool in improving creativity.

## 6. CONCLUSION

We conducted an exploratory study to assess the effect of combine and classify interventions on originality and the quantity of concepts produced by engineering students. These interventions were compared against a control group.

We found statistically significant impact of the interventions, combine and classify on concept generation within group. However control did not yield improvements in originality. The quantity of concept produced decreased in Round 2 but, the number of radically different concepts either remained same or increased. We can deduce this improvement was due to interventions we studied. Results proves, combine and classify aid in engineering design concept generation.

Outcome from this investigation shows, it is indeed feasible to enhance the creativity in engineering design concepts generated by engineering students with even small interventions. This adds to the body of knowledge on the factors that can have a positive effect of engineering student creativity from a course or curriculum level [20, 24] to creativity method or tool level [34, 35].

## REFERENCES


[1] Levitt, Theodore, "Creativity Is Not Enough," *Harvard Business Review*, (2002).

[2] Charlton, Emma, "These Are the 10 Most In-demand Skills of 2019, According to LinkedIn." World Economic Forum. January 14, 2019. Accessed February 24, 2019. https://www.weforum.org/agenda/2019/01/the-hard-and-soft-skills-to-futureproof-your-career-according-to-linkedin/.

[3] European Commission, "Universities of the future," Accessed February 24, 2019. https://universitiesofthefuture.eu/

[4] Organisation for Economic Co-operation and Development, "Centre for Educational Research and Innovation-CERI," Accessed February 24, 2019. http://www.oecd.org/education/ceri/

[5] Oman, Sarah K., Tumer, Irem Y., Wood, Kris and Seepersad, Carolyn, "A comparison of creativity and innovation metrics and sample validation through in-class design projects," *Research in Engineering Design*, Vol. 24, No. 1 (2013): pp. 65–92.

[6] Herrmann, Ned and Herrmann-Nehdi, Ann *The Whole Brain Business Book*. McGraw Hill, New York (1996).

[7] Osborn, Alex F. *Applied Imagination*. Scribner, New York (1957).

[8] Altshuller, Genrikh and Shulyak, Lev, *40 Principles: TRIZ Keys to Innovation*. Technical Innovation Center, Worcester, MA (1997).

[9] Shah, Jami, Vargas-Hernandez, Noe, Summers, Joshua and Kulkarni, Santosh, "Collaborative Sketching (C-Sketch) — An Idea Generation Technique for Engineering Design," *Creative Education Foundation*, Vol. 35, No. 3 (2001): pp. 168-198.

[10] Casakin, Hernan and Goldschmidt, Gabriela, "Expertise and the use of visual analogy: implications for design education," *Design Studies*, Vol. 20, No. 2(1999): pp. 153-175.

[11] Linsey, Julie S., Wood, Kris L., and Markman, Arthur B. "Increasing Innovation: Presentation and Evaluation of the

Wordtree Design-by-Analogy Method," *Proceedings of the ASME IDETC/CIE*. DETC2008-49317: pp. 21-32, Brooklyn, New York, August 3–6, 2008.

[12] Knoll, Stefan W. and Horton, Graham, "The Impact of Stimuli Characteristics on the Ideation Process: An Evaluation of the Change of Perspective 'Analogy'," *Proceedings of 44th Hawaii International Conference on System Sciences*, pp. 1-10, Kauai, HI, January 4-7, 2011.

[13] Linsey, Julie S., Clauss, Emily F., Kurtoglu, Tolga, et al., "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 (2011): pp. 031008

[14] Silk, Eli M., Daly, Shanna, Jablokow, Kathryn, et al., "Interventions for ideation: Impact of framing, teaming, and tools on high school students' design fixation," *American Educational Research Association Annual Meeting (AERA '14)*, Philadephia, PA, April 3-7, 2014.

[15] Shah, Jami, Smith Steve, and Vargas-Hernandez Noe, "Metrics for measuring ideation effectiveness." *Design studies* Vol. 24 No. 2 (2003): pp. 111-134.

[16] Srivathsavai, Ramesh, Genco, Nicole, Hölttä-Otto, Katja and Seepersad, Carolyn C., "Study of Existing Metrics Used in Measurement of Ideation Effectiveness," *Proceedings of the ASME IDETC/CIE*. DETC2010-28802: pp. 355-366, Montreal, August 15–18, 2010.

[17] Sarkar, Prabir and Chakrabarti, Amaresh, "Assessing design creativity," *Design Studies*, Vol. 32, No. 4 (2011): pp. 348-383.

[18] Brown, David, "Problems with the calculation of novelty metrics," *Sixth International Conference on Design Computing and Cognition (DCC'14)*, pp. 23-25, London, June 2014.

[19] Charyton, Christine, Jagacinski, Richard J., and Merrill, John A., "CEDA: A research instrument for creative engineering design assessment," *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 2, No. 3 (2008): pp. 147-154.

[20] Kershaw, Trina C., Seepersad, Carolyn C., Hölttä-Otto, Katja, et al., "The Effects of the Undergraduate Curriculum and Individual Differences on Student Innovation Capabilities," *Proceedings of the ASME IDETC/CIE*. DETC2014-35540: pp. V003T04A005, Buffalo, NY, August 17–20, 2014.

[21] Charyton, Christine, "An overview of the relevance of creative engineering design: Background," *Creative Engineering Design Assessment*, Springer, London (2014), pp. 1-10.

[22] Genco, Nicole, Johnson, Danny, Hölttä-Otto, Katja, et.al., "A Study of the Effectiveness of Empathic Experience Design as a Creativity Technique," *Proceedings of the ASME IDETC/CIE*. DETC2011-48256: pp. 131-139, Washington, DC, August 28–31, 2011.

[23] Kershaw, Trina C., Bhowmick, Sankha, Seepersad, Carolyn C., et al., "A Decision Tree Based Methodology for Evaluating Creativity in Engineering Design," *Frontiers in Psychology*, Vol. 10, (2019): pp. 1-19.

[24] Kershaw, Trina C., Peterson, Rebecca L., McCarthy, Molly A., et al., "A Cross-Sectional and Longitudinal Examination of the Development of Innovation Capability in Undergraduate Engineering Students," *Proceedings of the ASME IDETC/CIE*. DETC2015-47650: pp. V003T04A008, Boston, MA, August 2–5, 2015.

[25] Cohen, Jacob, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit," *Psychological Bulletin*, Vol. 70, No. 4 (1968): pp. 213-220.

[26] Fleiss, Joseph L., *Statistical Methods for Rates and Proportions*. John Wiley and Sons, New York (1981).

[27] Field, Andy, "Non parametric tests", *Discovering Statistics Using SPSS*. SAGE Publications Ltd, London (2009): pp. 539-550.

[28] Leech, Nancy and Onwuegbuzie, Anthony, "A Call for Greater Use of Nonparametric Statistics," *Annual Meeting of the Mid-South Educational Research Association*, ED471346, pp. 1-25, Chattanooga, TN, November 6-8, 2002.

[29] Jacob Cohen, *Statistical power analysis for the behavioral sciences,* 2. Ed., Erlbaum, Hillsdale, NJ (1988).

[30] Cohen, Jacob, "A power primer," *Psychological Bulletin*, Vol. 112, No. 1(1992): pp. 155–159.

[31] Bergendahl, Magnus, Magnusson, Mats and Björk, Jennie, "Ideation High Performers: A Study of Motivational Factors," *Creativity Research Journal* , Vol. 27, No. 4 (2015): pp. 361-368.

[32] Chan, Joel, Dow, Steven P. and Schunn, Christian D., "Do the best design ideas (really) come from conceptually distant sources of inspiration?," *Design Studies*, Vol. 36, (2015): pp. 31-58.

[33] Kudrowitz, Barry M. and Wallace, David, "Assessing the quality of ideas from prolific, early-stage product ideation," *Journal of Engineering Design*, Vol. 24, No. 2 (2012): pp. 120-139.

[34] White, Christina, Kristin Wood, and Dan Jensen. "From brainstorming to C-sketch to principles of historical innovators: ideation techniques to enhance student creativity." *Journal of STEM Education: Innovations and Research* Vol.13, No. 5 (2012): p. 12

[35] Farel, Romain, and Bernard Yannou. "Bio-inspired ideation: Lessons from teaching design to engineering students." *Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies,* Vol. 7: Human Behaviour in Design, pp. 327-336, Seoul, August 19-22, 2013,

 