

Comparative Evaluation of a Novel Concept Design Method

Damian Rogers

Ryerson University

Filippo A. Salustri

Ryerson University

digital.library.ryerson.ca/object/156

Please Cite:

Rogers, D., & Salustri, F. A. (2013). Comparative evaluation of a novel concept design method. *Proceedings of the ASME 2013 International Mechanical Engineering Congress and Exposition: Vol. 12. Systems and Design* (pp. V012T13A028). San Diego, CA: ASME.

[doi:10.1115/IMECE2013-63769](https://doi.org/10.1115/IMECE2013-63769)

IMECE2013-63769

COMPARATIVE EVALUATION OF A NOVEL CONCEPT DESIGN METHOD

Damian Rogers

PhD Candidate, Ryerson University
Toronto, ON, Canada

Dr. Filippo A. Salustri

Associate Professor, Ryerson University
Toronto, ON, Canada

ABSTRACT

Based on their previous work in creating a new method of design, termed the "Design by DNA" method, the authors are now experimentally validating the method against other, known methods. The goal of the experiment is to determine if Design by DNA promotes creative designs. Specifically, the authors are seeking to measure and compare creativity resulting from the use of Design by DNA and from other, known design methods. However, few have conducted empirical experiments in the past, and further, the literature on comparatively evaluating creativity of different design methods is relatively sparse. Therefore, the authors are developing a framework for defining and executing meaningful experiments that can accommodate various design methods, including Design by DNA, and also provide meaningful data to comparatively evaluate those methods, with the goal of determining whether Design by DNA impacts creativity in design. The experimental framework is described, and results of a pilot experiment are given. In that framework, creativity was characterized by novelty, usefulness, and cohesion. Due to small sample sizes, confidence in the results is not particularly high. Even so, some results do indicate several points of interest. An analysis of the results suggests that Design by DNA can offer advantages in engineering design, ranking higher in both the 'usefulness' and 'cohesion' categories of the creativity assessment. Hypotheses are given to explain why the experimental results show the slightly lower score in the 'novelty' category. Experiment participants were also evaluated on the NASA Task Loading Index (TLX) to evaluate how taxed they were using the different design methods and results are shown. Here, the Design by DNA method accrues better scores in 5 of the 6 NASA TLX categories, suggesting that it was less strenuous on the participants than the other methods. Statistical analysis of both the creativity scores and the TLX document shows confidence levels of between 65% and 96%, which is acceptable for very low populations. As this was a pilot experiment, the authors foresee future work to improve the results presented here. First, larger sample sizes are needed to improve statistical significance of our conclusions. Secondly, the authors wish to set out a series of experiments

whereby each test is run by pitting one specific design method against the Design by DNA method, to better show a 1-on-1 comparison between the methods and highlight the strengths and weaknesses of each.

INTRODUCTION

It is in the concept development stage of any design method that the form of the design starts to take shape and sets the stage for the detailed design and further, the finished product at the end of the day. Concept development not only lets designers try to find the best solution for the design problem, but also is the time when creativity is allowed to flourish and when innovative designs may occur. Creativity has been acknowledged as an integral part of design and engineering [1-3]. Some researchers have even argued that design inherently includes creativity [4]. On a practical level, creativity is a measure of design success and can be considered an "order winner." While customers will voice functional and performance requirements, e.g.: cargo space, fuel efficiency, etc. creativity will often remain an unvoiced requirement [5], even though the more creative and innovative products will be selected. As such, creativity is highly valued in engineering as well as in business and industry [6][7].

In engineering design, creativity can be implicitly taught by presenting methods that were developed specifically to increase concept creativity, such as brainstorming, TRIZ (the Russian acronym for the Theory of Inventive Problem Solving), random stimuli, etc. [8-12]

Creativity is generally regarded as a recognizable entity albeit difficult to define - something that one recognizes when one "sees it" [13][14]. Many researchers view creativity as the synthesis of a whole from different, often unrelated parts [15-17]. This can involve finding relationships [10][15][16], or transforming [16] those different and unrelated parts. There is general agreement that these "parts" include information and knowledge, which are also crucial elements of design.

Still, even with the current emphasis on design creativity, existent methods generally do not give the designer direction on *how* to be creative; only that there are ways and/or stimuli to

promote creative thinking. That is, in all the creativity methods studied so far by the authors, there is invariably a creative “moment” that is not part of the method itself; put another way, no creativity method reviewed so far can *guarantee* a creative solution. [9][11][12][18][19]

Further, given the overwhelming signs that creativity is a desirable trait in design, there is seldom, if any, consensus on how to measure creativity [20]. Therefore, the authors will define what they have used to evaluate creativity in design for this study.

Engineering design literature indicates *usefulness* as a key measure of creativity [2][3][21]. Ullman [3] specifies that creative ideas must be more than just good ideas; they must *solve the problem* (be useful). De Bono [10] stresses that creative ideas must be *novel* and include an underlying logic and value. In contrast, ideas that are merely novel may be “bizarre,” strange, or even incorrect. Most agree that at the very least, an engineering design must always be useful; which is to say that the design is appropriate, functional, correct, and valuable [10][15][22][23]. It is also shown that creative designs should be *novel*; described as new, original, or surprising [1][24][25].

Still more researchers include others measures of creativity, such as: *fluency*, *flexibility*, *detail* and *elegance* [22][24]. In an attempt to facilitate some of the other factors of creativity found in the literature, the authors have chosen to also evaluate designs in the area of *cohesion*. A cohesive design is one that includes: complexity (details and implications of ideas), elaboration (ideas are explained in detail), fluency (ability to generate many ideas), abstraction (demonstrate abstract ideas or concepts), flexibility (ideas showing a variety of possibilities or realms of thought), and robustness (the design is applicable in a dynamic environment).

However, few have conducted empirical experiments [20] to (1) verify the different component measures of creativity (novelty, usefulness, cohesion), or (2) give designers a way to be creative, not just give pointers on what may, or may not, stimulate creativity. This study will give participants a way in which to be creative, while at the same time give them a method which allows for faster concept generation and potentially better overall designs. Results from the study will test whether the Design by DNA method is able to score higher than other, more conventional design methods in several different measures.

EXPERIMENTAL PROCEDURE

The study was conducted in 2 “runs”: an experiment conducted with a class of Master’s-level students in a design methodologies graduate course, all of whom had industry experience in various sectors (experiment run 1); and an experiment conducted with people directly from industry, from varying backgrounds (experiment run 2). Both experiments were conducted using the same design problem, to design an urban (bi- or tri-)cycle for use by “white-collar workers”. The package of material that was used to present, explain, and support each experiment was identical for both runs. The

package included: a description of what was expected as an outcome of the concept design; a set of requirements for the cycle; an hourly checklist, to be filled in after every hour of working on the design (see Annex A); and a walkout checklist to be completed when subjects were finished their concept design (see Annex B). In each run, there was a test group and a control group; test groups used the Design by DNA (DbD) method, and control groups used other design methods. Packages given to test groups included details on a standard set of “genes” (developed by the authors) for the exercise; packages given to the control groups lacked these details.

For both experimental runs, the final analysis of the results from each participant was conducted by two assessors; namely, the authors. The second author kept a certain distance from the execution of the experiments to help minimize potential assessor bias. Assessors were given everything that each participant handed in by the end of the experimental run. The final assessment of each participant’s work was based on the 3 outlined parameters for creative design: novelty, usefulness, and cohesion. A five-point Likert scale was used to determine how well each participant fared in each of the three areas, where a score of 1 indicated poor performance and 5 indicated exceptional performance. For purposes of confidentiality, and also to negate any personal preferences toward certain participants, all work was made anonymous before the assessors received it for evaluation. Participants were also asked to rank their own designs in terms of novelty and usefulness, as can be seen in the walk-out checklist questions in Annex B. Here, cohesion was purposely excluded from the participant self-evaluations since it is a relatively difficult concept and would have likely caused confusion for the participants. Self-evaluations were used as an extra set of data for analysis, as will be discussed later.

Experimental Run 1

In the first run, the design problem was given to the graduate students in the form of a homework assignment wherein students were expected to complete a concept design for the given problem. All students in the class had been given one hour of instruction in each of three different design methods: systems design, TRIZ, and Design by DNA (DbD). The methods were all taught in the same day. All students were given the same design package for their homework, as outlined above. Students were randomly assigned one of the three different concept design methods taught in class to use for their homework. Participants were expected to work individually on this assignment. While there was no hard limit on the amount of time students could spend on the assignment, they were encouraged by the authors to not exceed 8 hours; however, nearly all students exceeded this, with some reaching as high as 17 hours.

All students were expected to submit deliverables including: all ideation and concept generation sketches, notes, etc.; ranking of concepts; refined concepts and new rankings; a

two page report on the final “winning” concept, including explanation and sketches; hourly checklists for each hour spent on the homework; the completed walk-out checklist; and any other notes, diagrams, etc that they created during the time spent.

Experimental Run 2

The second experiment was run with participants from industry in the form of a design “charrette,” conducted at the university campus. Participants were given a short introduction to the charrette, explaining what was expected from them by the end. All participants were given the same informational package as in Run 1; however, since the second experiment was held in a more controlled setting, participants were also given the NASA TLX document (see Annex C), to be filled in at the end of the experiment. Participants were randomly assigned 1 of 2 design methods, a conventional design method, as can be found in common design literature (Pahl & Beitz [2], Suh [18], Pugh [19], Ullrich & Eppinger [26], etc.), or the DbD method; both of which were introduced to the participants by one of the authors. As in Run 1, participants worked individually on their assignments during the charrette. Total allowable time for the charrette was set at a soft 8 hours, with some small tolerance for overtime; however, it is interesting to note that all participants finished within 5 hours.

Since participants from this run were from different fields and none had seen the authors’ design method before, a short, 1 hour presentation on their assigned design method was given to them 2 days before the charrette. The control group received a presentation on designing via product characteristics, while the test group received a presentation on how to use the authors’ DbD method. Both lectures included a participatory example at the end in order to involve them in the presentation and ensure that the learning was reinforced through doing.

All participants were asked to hand in the final deliverables, including: system architecture documents (identification matrix, system diagrams, etc.); all ideation and concept generation ideas, sketches, notes, etc; rating of concepts; refined concepts and new ratings; a final “winning” concept; if time permitted, a short explanation and sketches explaining the final concept; hourly checklists for each hour spent; the walk-out checklist; and the NASA TLX worksheet.

RESULTS

Results from the walkout checklist questions and creativity assessments (both self and assessor) for both experimental runs were tabulated using a spreadsheet and separately compared. Means for each of the three creativity measures were calculated for each design method in two ways: (1) using all data (committee evaluations and self-evaluations), and (2) using only the committee’s evaluations. This was done to see if the self-evaluations had an overall impact on the results, since it is often found that people will either over- or under-evaluate their own performances [27-29]. It is interesting to note here that the self-

evaluations had no impact on the results for experimental run 2 and only affected one measure from one method in experimental run 1, even though the self-evaluations were often different than the assessors’ evaluations (see the tables of results below). The measures of interest in the walkout checklists were the participants’ answers to: 1) total time spent, 2) if participants would buy their own winning concept, 3) if they were comfortable using the design method they were given, and 4) would they use that design method again.

Results From Experimental Run 1

Since there were three methods being compared in this run, comparisons were done in pairs: DbD vs. systems design and DbD vs. TRIZ. The mean values of the DbD method and the conventional design method were compared; the results of which are shown in the tables below.

Table 1: Systems design vs. DbD creativity assessments

		Novelty	Usefulness	Cohesion
Systems Design	Committee only	3.83	3.75	3.75
	Committee + self	3.78	3.78	N/A
Design by DNA	Committee only	3.06	3.44	3.31
	Committee + self	3.13	3.96	N/A

Table 2: TRIZ vs. DbD creativity assessments

		Novelty	Usefulness	Cohesion
TRIZ Method	Committee only	4.08	3.50	2.67
	Committee + self	3.72	3.94	N/A
Design by DNA	Committee only	3.06	3.44	3.31
	Committee + self	3.13	3.96	N/A

As can be seen in the above tables, the DbD method accrues higher scores than TRIZ in the cohesion category. We note that whether self-evaluations were included in the calculations has a small impact on the usefulness category; whereby, the scenario which added the self-evaluations shows DbD as a category winner against both other methods.

Next, results from the participants’ answers to three of the walkout checklist questions were tabulated and can be seen in the table below.

Table 3: Experiment run 1 walkout checklist results

	Buy Design	Method Comfort	Method Re-use
Systems	4.33	2.67	4.33
TRIZ	4.33	2.50	2.33
DbD	4.50	3.25	4.00

Results from the table above indicate that participants using the DbD method were more satisfied with their final concept design, as shown by the higher scores when asked if they would want to purchase their design if it were available on the market today and also felt more comfortable using the DbD method. On the question of whether they would want to re-use the design method in the future, the results show that participants would want to re-use the systems method the most, with TRIZ being the least favourable, and DbD scoring in the middle. This is not surprising given that the DbD method was new to them and quite different to any other current method. One can note that while the DbD method scores lower than systems design in the re-use category, the values show that it was quite closely scored. This may indicate that participants that had slightly more preparation and background knowledge or experience with the DbD method may score it higher in this category; however, there is not enough data to either confirm or deny this. Given that the DbD method scores the highest in 2 out of 3 questions and is in the middle for the 3rd suggests that the DbD method outperformed the other two methods.

Finally, the average time spent by participants for all three design methods was calculated. Here, the DbD method performed the best, scoring the lowest average time of 8.50 hours, with TRIZ taking an average of 11.67 hours and systems design 14.83 hours. These results reinforce the authors' hypothesis that the DbD method allows for faster concept design over other methods. As mentioned earlier, this variability in time spent seems to be a factor in the results of the creativity assessments for study and will be discussed further in a later section.

Results From Experimental Run 2

Given that the 2nd experimental run contained only two different methods, a direct analysis was made between them; thus pitting the DbD method against the conventional design method. As previously mentioned, the means of the assessments were calculated in each of the three creativity categories and for the two scenarios of: assessors only and assessors plus self-evaluations. The mean values of the DbD method and the conventional design method were compared; the results of that comparison are shown in the table below.

Table 4: Conventional method vs. DbD creativity assessments

		Novelty	Usefulness	Cohesion
Conventional Method	Committee only	4.38	3.13	3.13
	Committee + self	3.75	3.42	N/A
Design by DNA	Committee only	3.42	3.83	3.25
	Committee + self	3.50	3.78	N/A

As can be seen in the above results, the DbD method accrues higher scores in both the usefulness and cohesion categories, while the other design method wins out in the

novelty category. Inclusion of self-evaluations in the calculations does not affect the results here.

Next, results from the participants' answers to three of the walkout checklist questions were tabulated and can be seen in the table below.

Table 5: Experiment run 2 walkout checklist results

	Buy Design	Method Comfort	Method Re-use
Conventional	3.50	3.00	3.50
DbD	4.00	2.67	3.00

These results seem to indicate that participants using the DbD method were more satisfied with their final concept design, as shown by the higher scores when asked if they would want to purchase their design if it were available on the market today. On the other two questions of comfort with the design method used, and if they would want to re-use that method in the future, the results show that participants seemed to be more comfortable with the conventional method. This is not surprising given that the DbD method was new to them and quite different to any other current method.

As mentioned previously, this experimental run also asked participants to complete the NASA TLX worksheet, which measures how tasked a person is for the current activity. The worksheet has six questions to measure the task load, which can be seen in Annex C; higher values indicate a higher loading and thus, a more arduous task and a less desirable outcome. Results from the worksheets were tabulated and can be seen in the table below.

Table 6: NASA TLX worksheet results

	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
Conventional	7.5	2.5	4.75	5.25	6	3
DbD	7.17	1.17	4.00	3.33	6.83	2.33

As is seen in the above table, the DbD method accrued better scores in 5 out of 6 of the task loading measures, showing that on average, participants using the DbD method were less tasked than the users for the conventional design method. Even though the DbD method showed a less desirable result in the "effort" category, it is an expected result given that this was the first time that the participants had seen the DbD method and therefore, would have had to put more effort into the process. A point of future consideration would be to see how this measure changes with higher participant familiarity with the DbD method. The authors hypothesize that if participants had more experience with the DbD method, then results for the measure of effort would favour the DbD method, making it a winner in every facet of task loading.

DISCUSSION

The authors noticed that in run 1, some participants took substantively longer than others to perform the design task. While they were recommended to spend 7-8 hours, the actual range of total time spent ranged between 6 and 17 hours. This kind of variation did not happen in run 2 because of its fixed session length. It occurred to the authors that the amount of time spent may have affected the quality of the final designs that were submitted and, therefore, any conclusions we might draw about the relative ranking of the DbD method with respect to the other methods. It stands to reason that, within certain limits, the more time one spends on a design problem, the better one's results may be. There are, however, other possible influences on the quality of design as well, such as: previous design experience, distractions during work periods, availability of information, etc. The effect of time on the outcomes of the experimental runs becomes evident if one compares the results for run 1 (large time variability) vs. those of run 2 (time consistency). Whereby, run 1 shows the other methods being nearly even with (in the case of TRIZ) or scoring higher than (in the case of systems) the DbD method, while run 2 shows DbD outscoring the conventional method.

The authors are unable, with the available data from the experiments, to control for these factors or even to know the proportional effect each such factor may have. However, if it is the case that time spent and other factors all contribute to design quality, then it is also the case that the influence of time spent can only vary between 0% and 100% influence, and most likely between some small but non-zero value and some large but less than 100% value.

If we can show the points (within the reasonable values of degree of influence) at which the time variability influences the final ranking of the three design methods, then we can start to see how the wide distribution in the time spent by the participants in run 1 affects the experiment outcome for that data set.

The authors have attempted to calculate the impact of varying the influence of time spent by normalizing all data such that we are able to see this temporal impact at several key points; namely, 100%, 20%, 12.5%, and 8.3%, with 0% being the original data. It is at these key points that we can see changes in the results occur, as can be seen in the table below.

Tables 7-10: Temporal influence on creativity assessments

		100% Temporal Influence		
		Novelty	Usefulness	Cohesion
Systems Design	Committee only	1.502	1.516	1.516
	Committee + self	1.506	1.543	1.516
TRIZ	Committee only	2.213	1.888	1.499
	Committee + self	2.012	2.137	1.499
Design by DNA	Committee only	2.217	2.883	2.817
	Committee + self	2.300	3.256	2.817

		20% Temporal Influence		
		Novelty	Usefulness	Cohesion
Systems Design	Committee only	2.914	2.883	2.883
	Committee + self	2.889	2.914	2.883
TRIZ	Committee only	3.459	2.960	2.291
	Committee + self	3.150	3.341	2.291
Design by DNA	Committee only	2.820	3.194	3.080
	Committee + self	2.883	3.673	3.080

		12.5% Temporal Influence		
		Novelty	Usefulness	Cohesion
Systems Design	Committee only	3.211	3.164	3.164
	Committee + self	3.176	3.194	3.164
TRIZ	Committee only	3.670	3.143	2.419
	Committee + self	3.344	3.545	2.419
Design by DNA	Committee only	2.905	3.280	3.162
	Committee + self	2.968	3.773	3.162

		8.3% Temporal Influence		
		Novelty	Usefulness	Cohesion
Systems Design	Committee only	3.388	3.331	3.331
	Committee + self	3.348	3.361	3.331
TRIZ	Committee only	3.794	3.249	2.494
	Committee + self	3.457	3.665	2.494
Design by DNA	Committee only	2.954	3.329	3.209
	Committee + self	3.017	3.831	3.209

We found that for the range of influence of time spent between 0% and 8.3%, the resulting ranking of the design methods does not change. In the range of 8.3% to 12.5%, the DbD method starts to overtake the other methods in the usefulness category, under scenario 1 (IE: assessment via committee only). Further, above 12.5% and below the next key threshold of 20%, the DbD method is shown to overtake the other methods in the cohesion category. Finally, the DbD method outperforms both other design methods in all categories with any temporal influence factor above 20%. While we have not yet been able to analyze these results further, we have reason to believe that the results as presented here support the claim that DbD can perform well compared to other design methods, given a reasonable estimation for the temporal influence.

To ensure that the results from the previous section were statistically significant, *t-tests* for checking confidence were conducted. A *t-test* is commonly used to compare two

populations of values and determine if the conclusions for the hypothesis are statistically significant. This confidence level is reflected in the resulting *P value*, where the confidence interval (CI) = (1-P)*100, given as a percent. Confidence was calculated for all three creativity factors, in both scenarios, as well as for the walkout checklist questions.

Statistical Analysis of Experimental Run 1

After the confidence intervals were calculated for the three creativity categories across both scenarios, the results were tabulated, as can be seen in the two tables below.

Table 11: Systems vs. DbD creativity assessment confidence

	Novelty	Usefulness	Cohesion
Committee only	93.19%	84.44%	83.19%
Committee + self	89.55%	67.34%	N/A

Table 12: TRIZ vs. DbD creativity assessment confidence

	Novelty	Usefulness	Cohesion
Committee only	99.49%	58.65%	87.44%
Committee + self	93.05%	51.47%	N/A

As can be seen by the above table, all but one of the calculated confidence levels in the range (51.47% - 99.49%) are below the commonly accepted values of 95% or 99%. However, given that the data set populations were quite low (IE: far below the accepted norms of 100's or 1,000's of data samples), the resulting confidence levels are found to be acceptable. It is interesting to see a confidence as high as 99.49% in these low populations. Another interesting conclusion is that the confidence in the results from the scenario with the addition of the self-evaluations is lower than that of the committee alone. This lower confidence might reflect the personal bias inherent in self-evaluations, but this sociological topic is beyond the scope of this paper. Next, CI was calculated for the results of the three walkout checklist questions, listed in the previous section. The results are shown in the table below.

Table 13: Experiment run 1 walkout confidence

	Buy Design	Method Comfort	Method Re-use
Systems vs. DbD	59.55%	75.20%	71.19%
TRIZ vs. DbD	59.55%	77.41%	98.47%

Values of 59.55% to 98.47% can be seen in the table above. As before, these values are deemed acceptable for low populations.

Statistical Analysis of Experimental Run 2

Tabulated results for CI in each of the three creativity categories and for both scenarios are shown in the table below.

Table 14: Experiment run 2 creativity assessment confidence

	Novelty	Usefulness	Cohesion	Cohesion'
Committee only	88.05%	84.07%	55.44%	96.83%
Committee + self	65.80%	77.28%	N/A	N/A

As can be seen by the above table, the calculated confidence levels of 55.44% - 88.05% are below the commonly accepted values of 95% or 99%. However, given that the data set populations were quite low (IE: far below the accepted norms of 100's or 1,000's of data samples), the resulting confidence levels are found to be acceptable. It is interesting to note the final column of the table; which recalculates the CI for the category of cohesion, termed cohesion', when the single outlier data point is removed. We see here that the confidence in the result increases significantly, from the original 55.44% to 96.83%, whilst the resulting mean with the outlier removed continues to show that the DbD method scores higher. As with experimental run 1, CI values for the scenario with the added self-evaluations are also lower; again, this is most likely due to the bias in self-evaluations.

Table 15: Experiment run 2 walkout confidence

	Buy Design	Method Comfort	Method Re-use
Confidence	70.46%	60.56%	86.39%

The table above reflects the CI results for the questions that were asked on the walkout checklist. Similarly to the creativity assessments, the CI ranges for the questions were between 60.56% and 86.39%. Again, these are acceptable values given the low populations.

For this experimental run, the CI was also calculated for each of the 6 measures of the NASA TLX worksheet. Results are shown in the table below.

Table 16: NASA TLX worksheet confidence

	Mental Demand	Physical Demand	Temporal Demand	Performance	Effort	Frustration
CI	75.25	87.49	64.49	77.69	70.82	75.25

As with the other calculated confidences for this experiment, the table above shows a range of CI from 64.49% to 87.49%, which are again, acceptable for low populations. It is not surprising that the highest confidence was in the measure of physical demand, since there is almost no demand on people in an experiment setting such as this experiment and all participants scored very nearly the same.

CONCLUSIONS

When comparing the results of the creativity assessments of experiment run 1 to run 2 and factoring in the temporal

influence, one can see that for any temporal influence factor above 12.5%, the results of both experiments coincide. Which is to say that, for any value above 12.5%, the DbD method outperforms all other methods in both usefulness and cohesion. Though the DbD method seems to fall short of the other methods in the novelty category for both experiments, the authors hypothesize that this is due to the fact that the participants were not well versed enough in the DbD method to allow them the broadest scope of its uses, which would hinder the ability to generate the most novel solutions. It is also possible that the participants using the DbD method were not given enough supplemental materials to fully explore the realm of concepts, which would also explain the lower scores for novelty.

Although participants were asked to complete the hourly checklists, the data collected from that source was not used in the study outlined in this paper. The authors felt that the amount of data available in those checklists and the implications of that data were beyond the scope of this paper and the size limit imposed. It is felt that more research and analysis is needed for that particular data set before publishing the results.

Other future work stemming from this study include running more sets of experiments to better validate the results and conclusions presented in this paper. Also of interest are experiments to investigate the difference in results if another experiment is run with the same participants, using the same methods. Here, it might be possible to see how results reflect the increased experience and familiarity that participants have when engaging in a second similar design study.

REFERENCES

- [1] V. Hubka, W.E. Eder, *Design Science- Introduction to the Needs, Scope and Organization of Engineering Design Knowledge*, Springer-Verlag, London, UK, 1996.
- [2] G. Pahl, W. Beitz, *Engineering Design, a Systematic Approach*, K. Wallace, L. Blessing, and F. Bauert, Trans., K. Wallace, Ed., 2/e., Springer-Verlag London Ltd., London, UK, 1996.
- [3] D. Ullman, *The Mechanical Design Process*, Third Edition, McGraw-Hill, New York, NY, 2003.
- [4] A. Hatchuel, B. Weil, C-K design theory: an advanced formulation, *Res. Eng. Design*, 2009, 19:181-192.
- [5] N. Cross, *Designerly Ways of Knowing*, Springer-Verlag, London, UK, 2006.
- [6] T. Kelley, J. Littman, *The Ten Faces of Innovation*. Doubleday, NYC, NY, 2005.
- [7] U. Mahle, "The Path to Invention", *Mechanical Engineering*, September 2007, vol. 129, no.9 pg 37-38.
- [8] A.F. Osborn, A.F., *Applied imagination: principles and procedures of creative problem-solving*, 3/e, Scribner, 1963.
- [9] G.S. Altshuller and L. Shulyak, *And Suddenly the Inventor Appeared: TRIZ, the Theory of Inventive Problem Solving*, 2/e, Technical Innovation Center, Worcester, MA, 1996.
- [10] E. De Bono, *Serious Creativity*, HarperCollins, NY, 1992.
- [11] N. Roozenburg, & J. Eekels. *Product design: fundamentals and methods*. Vol. 2. Chichester: Wiley, 1995.
- [12] J.C. Jones. *Design methods*. John Wiley & Sons, 1992.
- [13] J.S. Gero, Computational Models of Creative Design, Chalmers Design Seminar, March 19, 2008, University of Toronto.
- [14] J. Betz, Assessing creativity in architectural design: Evidence for using student peer review in the studio as a learning and assessment tool, *ASEE Annual Conference*, Austin, TX, June 14-17 2009, AC-2009-428.
- [15] T. M. Amabile, 1983, The Social Psychology of Creativity, *Journal of Personality and Social Psychology*, 43: 997-1013.
- [16] M.A. Boden, What is Creativity?, in M.A. Boden's (ed.) *Dimensions of Creativity*, The MIT Press, 75-117, 1994.
- [17] T.B. Ward, S.M. Smith, & R.A. Finke, Creative Cognition, in R.J. Sternberg's (ed.) *Handbook of Creativity*, Cambridge University Press, 1999.
- [18] N.P. Suh, *The principles of design*. Vol. 990. New York: Oxford University Press, 1990.
- [19] S. Pugh, *Total design: integrated methods for successful product engineering*. Workingham: Addison-Wesley Publishing Company, 1991.
- [20] I. Chiu, & F.A. Salustri. Evaluating Design Project Creativity in Engineering Design Courses. *Proceedings of the Canadian Engineering Education Association*, 2010.
- [21] J.J. Shah, S.V. Kulkarni & N. Vargas-Hernandez, Evaluation of Idea Generation Methods for Conceptual Design: Effectiveness Metrics and Design of Experiments, *Journal of Mechanical Design*, 2000, 122:377-384.
- [22] S.P. Besemer, & D.J. Treffinger, Analysis of Creative Products: Review and Synthesis, *Journal of Creative Behavior*, 15:158-178, 1981.
- [23] O. Akin, & C. Akin, On the process of creativity in puzzles, inventions and designs, *Automation in Construction*, 7:123-138, 1998
- [24] E.P. Torrance, *Torrance Tests of Creative Thinking*, Scholastic Testing Service, Inc., 1974.
- [25] D.C. Brown, Guiding Computational Design Creativity Research, in J. Gero's (ed.) *Studying Design Creativity*, Springer, 2008.
- [26] K.T. Ulrich, & S.D. Eppinger, *Product design and development*. McGraw-Hill, 1995.
- [27] P.M. Podsakoff, & D.W. Organ. "Self-reports in organizational research: Problems and prospects." *Journal of management*, 12.4:531-544, 1986.
- [28] M.M. Harris, & J. Schaubroeck. "A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings." *Personnel Psychology*, 41.1:43-62, 1988.
- [29] L.M. Hough, N.K. Eaton, M.D. Dunnette, J.D. Kamp, & R.A. McCloy. "Criterion-related validities of personality constructs and the effect of response distortion on those validities." *Journal of Applied Psychology*, 75, no. 5:581, 1990.

ANNEX A

HOURLY CHECKLIST

1) What best describes the design stage you are in?

1	2	3	4	5
Identifying design elements	Ideating embodiments	Evaluating embodiments	Refining ideas and concepts	Creating a full product concept

2) Report on progress during this time period

- Briefly touch on what was accomplished
- A point may be: "selected the drivetrain" or "finished the system diagram"

3) What, if any, obstacles have you come across during this time period?

- Briefly report on anything that slowed progress

4) At this point, how much time (to nearest half hour) do you estimate you need to finish?

5) Approximately how much time in this period has been spent re-reading the design notes?

1	2	3	4
Little/none	Some	Fair	Lots

ANNEX B

WALKOUT CHECKLIST

1) Estimate your total time spent on this design problem (round to the nearest 30min interval)

2) How do you think your winning concept rates in terms of: (Use a 1-5 scale with 1 being the worst and 5 being the best)

2.1) **Novelty** (how novel was your winning concept compared to other products on the market?)

- A *novel* design is: new, original, or surprising

2.2) **Usefulness** (does your design fulfill some, most, or all of the requirements? How well does it satisfy those requirements?)

- A *useful* design is: appropriate, functional, correct, and valuable

3) You would want to buy your winning concept if it were to be put on the market tomorrow

1	2	3	4	5
Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree

4) Do you have previous design experience? If so, compare the method you used today against what you've used in the past.

5) How comfortable did you feel with using this method for concept design?

1	2	3	4	5
Very uncomfortable	Somewhat uncomfortable	Neutral	Somewhat comfortable	Very comfortable

6) You would want to use this method for concept design again in the future

1	2	3	4	5
Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree

7) Please add any other comments you wish to share about your experiences using this design method.

ANNEX C

NASA TLX WORKSHEET

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date
------	------	------

Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed were you?

Very Low Very High