

# *Understanding design ideation mechanisms through multilevel aligned empirical studies*

Noe Vargas Hernandez, Dept. of Mech. Engineering, University of Texas at El Paso, El Paso, TX 79968-0521, United States

Jami J. Shah, Dept. of Mech. and Aerospace Engineering, Arizona State University, Tempe, AZ 85287-6106, United States

Steven M. Smith, Dept. of Psychology, Texas A & M University, College Station, TX 77843-4235, United States

*The overall objectives of the study are to identify key components of ideation methods and develop effectiveness metrics. This paper presents experimental results conducted on six ideation components (Provocative Stimuli, Suspend Judgment, Flexible Representation, Frame of Reference Shifting, Incubation, and Example Exposure). These experiments were conducted simultaneously at the Design (engineering) and Lab (cognitive psychology) levels following an experimental procedure previously developed to align these two levels of experiments. Results show the effectiveness of the ideation components in terms of effectiveness metrics and also the extent of alignment of the results from these two levels. Understanding of ideation components has been gained (main effects ease of manipulation and interactions).*

© 2010 Elsevier Ltd. All rights reserved.

**Keywords:** design ideation, creativity, design cognition, engineering design

The overall objectives of the study are to gain an insight into cognitive mechanisms in design ideation by identifying key components of ideation methods and developing effectiveness metrics. The first step is to understand the effectiveness of such ideation components through empirical evaluation. Based on past experience (Shah, Kulkarni, & Vargas-Hernandez, 2000; Shah, Smith, Vargas-Hernandez, Gerkens, & Wulan, 2003) conducting experiments on ideation methods as a whole had major limitations: (1) inability to discriminate between necessary and superfluous components, (2) prohibitive number of experiments required (i.e. one for each method or variation), (3) inability to extrapolate experimental results to different environments, design problems, and human variables, and (4) poor understanding of the causes and effects of ideation. Shah, M.S. Smith, et al. (2003) introduced an experimental methodology to overcome these limitations. Design researchers at Arizona State University (ASU) and cognitive researchers at Texas A&M (TAMU) developed an ‘aligned’ experimental approach that

**Corresponding author:**  
Noe Vargas Hernandez  
nvargas@utep.edu



[www.elsevier.com/locate/destud](http://www.elsevier.com/locate/destud)

0142-694X/\$ - see front matter *Design Studies* 31 (2010) 382–410

doi:10.1016/j.destud.2010.04.001

© 2010 Elsevier Ltd. All rights reserved.

combined the complementary strengths of highly controlled (cognitive) Lab experiments and available atomic process models from cognitive science with design (engineering) experiments from design research. Simultaneous experiments were conducted at both Design and Lab levels on ideation components instead of ideation methods as a whole. Ideation components are defined as mechanisms believed to intrinsically promote ideation or to help designers overcome mental blocks.

This paper presents the main effects of the ideation components and some of their interactions. These results are shown in the Experimental Results section. Since the experiments were conducted at the ASU-Design and TAMU-Lab experimental levels, these results (i.e. main effects and interaction effects) were compared to evaluate the level of alignment. Higher-level observations were made not only from the experimental results but also from the experimental procedure.

## *1 Literature review*

A comprehensive classification of ideation methods developed in the last four decades can be found in [Shah \(1998\)](#), [Shah et al. \(2000\)](#), [Kulkarni \(1998\)](#), and [VanGundy \(1988\)](#). These methods can be classified into intuitive and logical methods. Intuitive methods attempt to overcome mental blocks (e.g. Brainstorming and C-Sketch) while logical methods make use of charts, databases, patent searches, physical principles, etc. (e.g. morphological charts and TRIZ). Intuitive and logical methods are important for the generation of ideas, but intuitive methods are typically the least understood and the ones with higher chances of producing novel ideas since logical methods clearly define the solution space while intuitive methods attempt to expand this space. Intuitive methods are the focus of this research. Ideation methods provide a prescription (i.e. a normative procedure) on how to overcome certain blocks to creativity, such as premature judgment. The empirical evidence of their effectiveness is frail; research studies like this one are needed.

### *1.1 Ideation components and effectiveness metrics*

Kulkarni and Shah observed that many intuitive methods had many common elements, which they termed ideation components. Ideation components ([Kulkarni, 2000](#); [Kulkarni & Shah, 1999](#)) are defined as cognitive mechanisms believed to intrinsically promote ideation or to help designers overcome mental blocks. Evaluating specific ideation methods in their entirety (known as the Direct Method) is complicated, the reason for this being that many components are at play simultaneously. The alternative was to identify components of ideation methods and test them individually (Component Based Method). The effectiveness of specific ideation methods could then be predicted by the components present in the method. Another advantage is that ideation components are commonly accepted and understood in engineering design research and cognitive psychology; this allows the possibility of connecting

the engineering design and cognitive psychology theories on ideation. Some of the components that have been identified are: Provocative Stimuli, Suspended Judgment, Flexible Representation, Frame of Reference Shifting, Incubation, Example Exposure, among others (Shah, M.S. Smith, et al., 2003).

In order to carry out empirical studies of design ideation at any level, one must specify how effectiveness of ideation is to be measured. One can focus on the ideation process (as in a protocol study), or the outcome (based on characteristics of ideas generated). Based on previous research done by Shah et al. (2000); Shah, Smith, and Vargas-Hernandez (2003) and creativity literature, four effectiveness metrics were proposed. Procedures for the evaluation of these effectiveness metrics are given in a later section.

## *1.2 Engineering design research*

Several empirical methods have been used for studying the design process and/or its associated cognitive activities. These include case studies (Altshuller, 1984; Marples, 1960; Ward & Sobek, 1996), protocol studies (Christiaans & Dorst, 1991; Christiaans & Venselaar, 1991; Cross, Dorst, & Roozenburg, 1991; Ericsson & Simon, 1984; Ullman, Dietterich, & Stauffer, 1988; Ullman, Stauffer, & Dietterich, 1987; Ullman, Wood, & Craig, 1989; Waldron & Brook, 1994), and controlled tests (Schön, 1991). There is not much reported on experimental studies of specific idea generation methods applied to engineering design, particularly groups engaged in conceptual design (except studies on Brainstorming or “free form” idea generation (Dennehy, Bulow, Wong, Smith, & Aronoff, 1992; Smith, 2000)). Regarding the experimentation on ideation components, there have been some studies on communication such as designers working in teams (Christiaans & Dorst, 1991; Christiaans & Venselaar, 1991; Hale, 1987; Leifer, 1996; Nagy, Ullman, & Dietterich, 1993; Schön, 1991; Wilde, 1999), industrial team review sessions (Hale, 1987), and data representation (Kan & Gero, 2005), among others. While there is no consensus among engineering design researchers on a unique theory or model of design ideation, each resulting theories and models provide valuable insight from their particular perspective.

When compared to cognitive research, Engineering Design research results better simulate real world design (i.e. less controlled environment, more complex tasks, closer to engineering design). One disadvantage, in general, is that experimentation and analysis is too time consuming (e.g. protocol studies). For each extra variable and interaction considered, the work required increases considerably, this because the same designer or team cannot be tested with the same problem but different methods. Another disadvantage is that the results, being empirical, have natural limitations, for example, results cannot be extrapolated to different conditions since there is little understanding of the behavior of the variables involved. A possible reason for the poor intrinsic validity is that many uncontrolled

cognitive processes (variables) are not accounted for. Cognitive research in design studies, among other aspects, studies the cognitive processes occurring during design.

### 1.3 Cognitive psychology

Cognitive Psychology observes and analyzes the cognitive processes occurring during various human reasoning tasks. Models of many cognitive phenomena, such as memory, perception, problem solving, have been developed by psychologists based on controlled experiments that often use very simple tasks or problems in order to focus on specific aspects of the ideation process. It is fair to say that there is no complete model that explains what happens in design ideation, and that their simplicity to a complex issue, such as ideation in the design process, may lack some degree of ecological relevance. For example, The Darwinian Model (Simonton, 2003) focuses on the production (i.e. quantity and variety of ideas); the Wallas model (Wallas, 1926) identifies 4 stages of creation: Preparation, Incubation, Illumination, and Verification. The Geneplore model (Finke, Ward, & Smith, 1992) divides the creative mental processes into generative and exploratory. Nonetheless, these and other cognitive models may offer guidelines for isolating the effects of independent variables, and they can also be used to address the interactive effects of combinations of factors.

### 1.4 Levels of ecological validity

Direct *macro level* experiments at the ideation method level, shown in Figure 1, such as those conducted in past ASU studies, simulate real world design better but are unable to discriminate between necessary and superfluous components, require prohibitive number of experiments, and are unable to explain the performance of methods under different conditions. On the other hand, a large body of knowledge on creative processes exists in cognitive science. However, the models in cognitive theories are derived from highly controlled *micro level* (i.e. cognitive process level) Lab experiments involving simple and isolated tasks that model atomic cognitive processes. There is little similarity

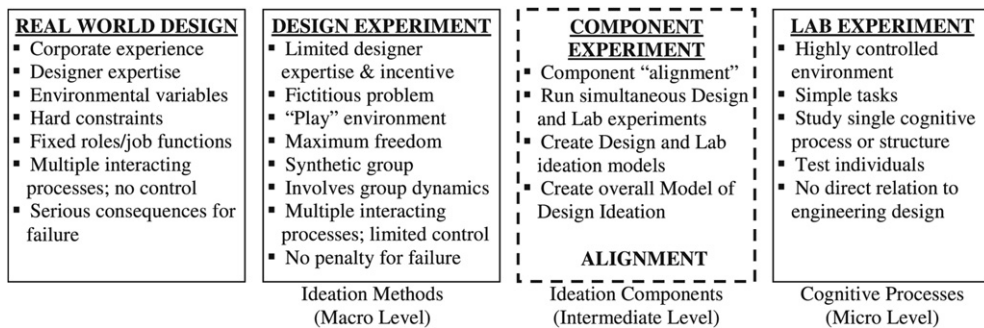


Figure 1 Levels of ecological validity

between the conditions for these experiments and design concept generation in the real world. The objective in this research was to increase both internal validity (cause-and-effect conclusions) at each level, and external ecological validity using an aligned experimental approach across levels. Experiments at an *intermediate level* (i.e. ideation component level) on specific ideation components have various advantages; “aligned” experiments can be run simultaneously at the Design and Lab experiment levels, results from such experiments can be compared across levels, and dual Design and Lab models for ideation components can be constructed based on these results.

The authors did not know a priori whether the ASU and TAMU experiments were comparable; this is precisely the whole idea of what we termed ‘alignment’. That is, our question is whether the two are comparable, and the experiments show how and when they are comparable, and how and when they are not. That is one of the main questions of alignment, whether we can generalize from one to the other.

## 2 *Research approach*

Experiments were conducted at two levels: micro (lab, cognitive) and macro (ideation components). The purpose is to see if experiments on the same combination of components will give the same results at both levels. As a consequence of this connection, more of the simpler Lab experiments could be run, and hence, collect more empirical data on specific ideation methods and ideation components. Another advantage of the alignment approach is the possibility of accessing the rich understanding that cognitive psychology has on mental mechanisms while engineering design provides more realistic data to support those theories. Details of the experiments can be found in [Shah, M.S. Smith, et al. \(2003\)](#) and [Vargas-Hernandez \(2007\)](#).

How to simultaneously run and compare results at the Lab and Design experiment levels? The fundamental issue was how to compare results from two different levels (Lab and Design experiments). The alignment approach was based on *three* key concepts: (1) Agreement on the ideation components to study, (2) similar factorial Design of Experiments (DOE), and (3) use of equivalent effectiveness metrics for assessment.

### 2.1 *Pre-alignment*

In order to compare experiment results across levels, design engineers at ASU and cognitive researchers at TAMU formulated the following three basic pre-alignment agreements:

#### 2.1.1 *Same ideation components to study*

More than a dozen ideation components were identified in previous work ([Kulkarni, 1998](#); [Kulkarni, 2000](#); [Kulkarni & Shah, 1999](#); [Mckoy, Vargas-Hernandez, Summers, & Shah, 2001](#); [Shah, Vargas-Hernandez, Summers, &](#)

**Table 1** Selected ideation components

<i>Ideation component</i>	<i>Description</i>	<i>Example method</i>
Provocative stimuli	Trigger new ideas by exposing the subject to related and unrelated pointers, pictures, sounds,...	C-Sketch, 635
Suspended judgment	Postpone premature decisions or dismissing an idea	PMI, Brainstorming
Flexible representation	Use representation mediums that are easier to manipulate, e.g. graphical representations	C-Sketch more flexible than 635
Frame of reference shifting	Change how objectives and requirements are being viewed, perceived, interpreted	Inversion, synectics
Incubation	Add programmed delay to allow sub-conscious processing to take place	Can be added to any method
Example exposure	Excite ideas by exposing the subject to a solution for the same problem	Gallery

Kulkarni, 2001). Due to time and resource constraints, only the most relevant were selected for experimentation: Provocative Stimuli (P), Suspend Judgment (J), Flexible Representation (R), Frame of Reference Shifting (F), Incubation (I), and Example Exposure (E); these components, shown in Table 1, appeared to be some of the most common among typical ideation methods. These constitute the experiment's dependent variables (i.e. factors).

### 2.1.2 Same DOE at both experimentation levels

Since a full factorial design with 6 factors and two levels requires  $2^6 = 64$  runs, and due to limitation of time and resources it was decided to study only the main effects for Provocative Stimuli (P), Suspend Judgment (J), and Flexible Representation (R), referred as Set 1, were studied individually with simple comparative experiments; to contrast each treatment. Provocative Stimuli (P) had 2 treatments: (1) presence and (0) absence. Suspend Judgment (J) had 2 treatments: (−1) focus on quantity and (+1) focus on quality. Flexible Representation (R) had 2 treatments: (−1) sketch only and (+1) free format. On the other hand, it was decided to study the main and interaction effects only for 3 ideation components (referred as Set 2): Frame of Reference Shifting (F), Incubation (I), and Example Exposure (E). These 3 ideation components were selected for full factorial design with 2 treatments each: (1) presence and (0) absence for a total of 8 runs (i.e.  $2^3$  runs). Table 2 presents a summary of the treatment condition and experimental runs for each ideation component.

The reason to study interactions only for F, I, and E was the interest from TAMU cognitive psychologists on conformity and persistence (two forms of fixation in design) of Frame of Reference Shifting (F) and Example Exposure (E) with respect to Incubation (I) (Jansson & Smith, 1991). Interaction experiments on P, J, and R could be run afterwards if needed. During experimentation, designers are subjected to one or more components at a time; this combination of components recreate a design ideation method that may resemble existing ideation methods. Various experiments were conducted for

**Table 2 Design of experiments**

<i>Ideation component</i>	<i>Treatment conditions</i>	<i>Experiments at ASU and TAMU</i>	
		<i>Main effects</i>	<i>Interaction effects</i>
Provocative stimuli	(1) Presence (0) Absence	Simple comparison 2 runs for (P): (1) vs. (0)	N/A
Suspended judgment	(−1) focus on quantity (+1) focus on quality	Simple comparison 2 runs (J): (−1) vs. (+1)	N/A
Flexible representation	(−1) sketch only (+1) free format	Simple comparison 2 runs (R): (−1) vs. (+1)	N/A
Frame of reference shifting	(1) Presence (0) Absence	2 <sup>3</sup> Full factorial design 8 runs (FIE): (100), (101), (110), (111), (001), (010), (011) and (000—control group)	
Incubation	(1) Presence (0) Absence		
Example exposure	(1) Presence (0) Absence		

this research during the last four years at the Design (ASU) and Lab (TAMU) levels.

### *2.1.3 Equivalent effectiveness metrics to assess the outcome*

These constitute the experiments' independent variables (i.e. responses). The four previously identified metrics Quantity, Quality, Novelty, and Variety (Shah et al., 2000; Shah, S.M. Smith, et al., 2003) are the average scores. These were extended to include the best quality and best novelty since they measure the possibility of attaining a maximum score, which may be of interest in contrast to an average score. The metrics used were: Average Quantity (AQN), Average Variety (AVR), Average Quality (AQL), Best Quality (BQL), Average Novelty (ANV), and Best Novelty (BNV).

## *2.2 Test hypotheses*

The experiment was formally proposed with the following hypotheses.

### *2.2.1 Hypothesis 1*

Ideation component X improves effectiveness metric Y, where X = P, J, R, F, I or E and Y = AQN, AVR, AQL, BQL, ANV or BNV.

### *2.2.2 Hypothesis 2*

Every main effect and interaction effect observed at the design level has the same sense (positive or negative) as at the Lab level.

## *2.3 Design task*

Students in the design experiment were asked to design a device (i.e. to generate sketches) to 'transport' a ping-pong ball the farthest distance powered only by a standard issue compression spring. The device is to be constructed with a limited set of given materials (e.g. balsa wood, wire and Styrofoam). The laboratory tasks used for Sets 1 and 2 at TAMU were primarily generation



tasks; requiring participants to produce multiple responses, rather than one correct response. Students were asked to generate simple tools (i.e. no need for electricity, motors, computers, etc.) for an intelligent species in another planet. They had to draw, label and describe as many tools as possible.

## 2.4 Independent variables (factors): ideation components

The ideation components selected constitute the independent variables of interest (i.e. factors) in the experiment. Each factor has the treatments shown previously in Table 2. This is how each treatment condition was manipulated:

### 2.4.1 Provocative stimuli (P)

In Design experiments, students were exposed to three figures: a frisbee, fire, and water, shown in Figure 2. The selection of these stimuli was difficult since depending on the stimuli chosen it will incite a different response. Since almost anything would qualify as a stimulus, a decision was made for the 3 instances selected to contain simple characteristics (flying, heat, and liquid) that could be easily identified in the generated ideas. Provocative Stimuli was not tested in Lab experiments.

### 2.4.2 Suspended judgment (J)

In Design and Lab experiments, subjects were asked to generate as many ideas as possible for level -1 (focus on Quantity), while subjects in level +1 (focus on Quality) were asked to generate better ideas without restrictions on quantity of ideas generated.

### 2.4.3 Flexible representation (R)

In Design experiments, subjects were asked to generate ideas using sketches only for level -1, while subjects in level +1 were asked to generate ideas using any type of representation (i.e. text, sketch, etc.). Flexible Representation was not tested in Lab experiments.

### 2.4.4 Frame of reference shifting (F)

In Design experiments students were given four hypothetical situations used to shift the subject's frame of reference, shown in Figure 3. The objective was to give the subject a different perspective, even if this perspective was not

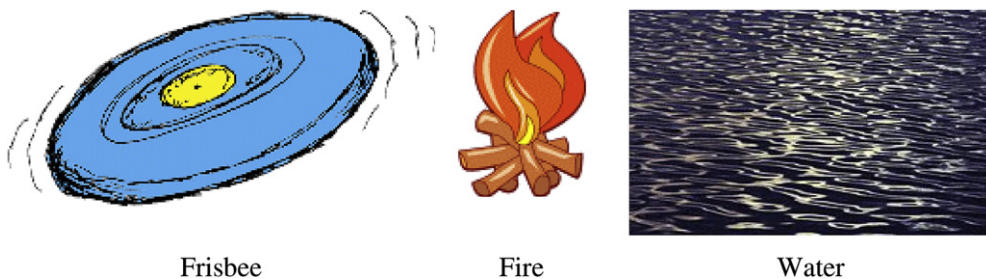


Figure 2 Provocative stimuli concepts used for design experiments



THINK OF THIS!  
 What if:

- The device you are designing is for catching a ball, instead of throwing it.
- The device you are designing is for underwater use, not open air.
- The ball is fixed (doesn't move) and the rest of the device is what travels.
- The device will be used with different types of balls (e.g. baseball, basketball, bowling, etc)

Figure 3 Frame of reference shifting text given for design experiments

complying with the original specification of the design problem. For Lab experiments, two frames of reference were used: tool-salesperson and tool-manufacturer.

### 2.4.5 Incubation (I)

In Design experiments subjects were given 2 days. No specific instructions were given to the subjects other than meeting again to resume the generation of ideas. In Lab experiments, subjects were given 10 min for incubation between ideation sessions.

### 2.4.6 Example exposure (E)

In Design experiments, subjects were exposed to a well-defined solution example, as shown in Figure 4a, an airplane launched by a slingshot, for the design problem. For Lab experiments, the examples used, shown in Figure 4b, were the “Rid-O-Pest”, a device to manually dispense pesticides. Both examples are simple enough to be quickly understood by the subjects and with a good level of novelty to be able to trace its conformity and persistence effect in the ideas generated if desired in the future. This is important since future work analysis may require tracing the effect of persistence and conformity (two forms of fixation in design).

The implementation of each of these components remains an open issue since these ideation components are not necessarily discrete; for example,

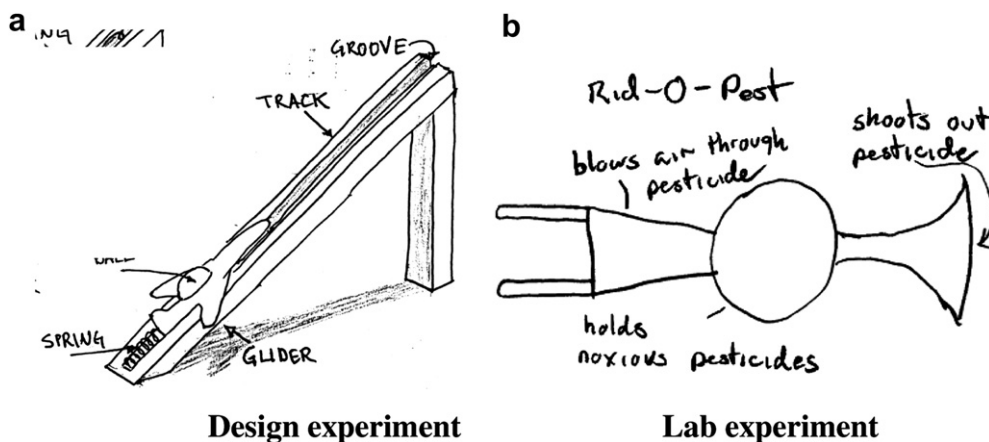


Figure 4 Examples used to induce conformity in design and lab experiments

incubation involves suspending the design activity for a period of time, it can be 1 min, 1 day or more. The effectiveness, level of detail, and accuracy in the definition of each component is a topic for future work.

## 2.5 Dependent variables (responses): effectiveness metrics

The dependent variables or responses are the outcome metrics obtained through the scoring process. A summary of these effectiveness metrics is shown in Table 3. These are: Average Quantity (AQN), Average Variety (AVR), Average Quality (AQL), Best Quality (BQL), Average Novelty (ANV), and Best Novelty (BNV). The six responses were scored for the Design experiment results while for Lab experiments only the average responses were scored.

**Table 3 Effectiveness metrics summary (Shah et al., 2000; Shah, S.M. Smith, et al., 2003)**

Effectiveness metrics	Calculation
<b>Quantity</b> <i>Definition:</i> total number of ideas generated.	$N$
<b>Variety</b> <i>Definition:</i> how different concepts are from each other. <i>Method:</i> A group of generated ideas is characterized based on a genealogy-like tree structure that has branches at various levels of physical abstraction: Physical Principle, Working Principle, Embodiment, and Detail.	$M_3 = \sum_{j=1}^m f_j \sum_{k=1}^n \frac{S_{3jk} b_k}{N}$ <p>where</p> <p><math>M_3</math> overall variety score</p> <p><math>m</math> is the total number of functions</p> <p><math>j</math> is the function being evaluated</p> <p><math>f_j</math> is the weight assigned for function <math>j</math></p> <p><math>n</math> is the number of stages (conceptual, embodiment, etc.)</p> <p><math>k</math> is the stage in the genealogy tree</p> <p><math>S_{3jk}</math> is the score given at function <math>j</math> and stage <math>k</math></p> <p><math>b_k</math> is the number of branches at stage <math>k</math></p> <p><math>N</math> is the total number of ideas</p>
<b>Quality</b> <i>Definition:</i> feasibility and conformance to design specifications. <i>Method:</i> The score for a given function in an idea is calculated based on answers to a qualitative and quantitative questionnaire.	$M_1 = \sum_{j=1}^m f_j \sum_{k=1}^n S_{1jk} p_k$ <p>where:</p> <p><math>m</math> is the total number of functions</p> <p><math>f_j</math> is the weight assigned for function <math>j</math></p> <p><math>n</math> is the number of stages (conceptual, embodiment, etc.)</p> <p><math>S_{1jk}</math> is the quality score given at function <math>j</math> and stage <math>k</math></p> <p><math>p_k</math> is the weight assigned to stage <math>k</math></p>
<b>Novelty</b> <i>Definition:</i> how unusual or unexpected an idea is as compared to other ideas. <i>Method:</i> Two approaches can be used. a priori: Before evaluating the ideas, judges predefine known or expected ideas. A-posteriori: Ideas are evaluated based on their occurrences.	$M_1 = \sum_{j=1}^m f_j \sum_{k=1}^n S_{1jk} p_k$ <p>where</p> <p><math>M_1</math> overall variety score</p> <p><math>m</math> is the total number of functions</p> <p><math>j</math> is the function being evaluated</p> <p><math>f_j</math> is the weight assigned for function <math>j</math></p> <p><math>n</math> is the number of stages (conceptual, embodiment, etc.)</p> <p><math>k</math> is the stage level</p> <p><math>p_k</math> is the weight assigned to stage <math>k</math></p> <p><math>S_{1jk}</math> is the score given at function <math>j</math> and stage <math>k</math></p> <p>(a priori – based on a table or a posteriori – using the formula to count occurrences:</p> $S_{1jk} = \frac{T_{jk} - C_{jk}}{T_j} \times 10$ <p><math>T_j</math> is the total number of ideas being evaluated</p> <p><math>C_j</math> is the count of the occurrence of that solution for that function</p>

## *2.6 Experiment subjects*

In both, lab and design experiment cases, the participants are assumed to have similar level of design expertise for the corresponding design task. This is appropriate for the experiments since the focus is on the components (i.e. incubation), not on the expertise level. Also, participants were randomly assigned to the treatment conditions, that is, the groups did not have prior knowledge of the treatment conditions and runs were randomly assigned to each participant group. Approximately 350 undergraduate students participated in the various Design experiment runs; an average of 25 students from 14 groups corresponding to the 14 runs (not counting preliminary and discarded experiments). No student participated in more than one run. The subjects were mechanical engineering undergraduate students with basic engineering design knowledge. From each run, only the ideas from 10 individuals, randomly selected were evaluated; this means that each run had 10 replicants since each participant in the same run is essentially a replication, that is, a repeated manipulation of the independent variable and of the dependent measures. A total of 256 undergraduate student volunteers from Introductory Psychology courses participated in this Lab experiment. Participants could choose from a variety of experiments to take part in, or they could choose to write a paper in order to fulfill a research requirement for their introductory psychology course.

## *2.7 Other variables*

Variables that may significantly affect the outcome response must be controlled or blocked; non-significant variables may be left uncontrolled to become part of the acceptable environmental noise. All runs were conducted in a classroom setting during regular class hours, avoiding extraordinary or unfamiliar situations for the subjects. The subjects were student taking introductory engineering design courses or Introduction to Psychology courses. The authors believe that the subjects have similar background and skills, and that minor differences are balanced by averaging. Being said that, the authors also recognize that each subject introduces an abundance of important variables not included in the scope of this work (e.g. personality, expertise level, motivation, problem solving style, among others). Simple design tasks were used for the experiments; these did not require specific knowledge or background other than the average understanding from an engineering or psychology student.

## *2.8 Experiment procedure*

The procedures for both the Lab and Design experiment were similar. A different group of subjects participated in each run. Every experiment run started with an instruction on general rules, the Design or Lab problem was introduced and explained, the students were allowed to ask any questions in order

to assimilate the problem. In the first session each student generated ideas that were collected; after this, each group was exposed to one, two or three ideation component manipulations sequentially, except for the control group who worked continuously from ideation session #1 to ideation session #2. The members of the group were instructed not to use any other components such as ideation strategies (although this was difficult to police during the Design experiments). Subjects in each run generated ideas individually for the same design problem (following a between-subjects model). Of particular interest were the ideas generated during ideation session #2, when the effect of the ideation component could be measured. The ideas generated by each group were collected.

## 2.9 Scoring of experimental data

The scoring of the ideas obtained in the experimentation follows the effectiveness assessment procedure described in previous research work (Shah et al., 2000; Shah, S.M. Smith, et al., 2003; Shah, Smith, & Vargas-Hernandez, 2005). In general, the approach for measuring ideation effectiveness is physically based since it maps the design space coverage as defined in traditional systematic design theory (Pahl & Beitz, 1996). The assigned scores are the numerical results to analyze and these range from 0 to 10 where 0 represents the lowest Quantity, Quality, Novelty, Variety, and 10 the highest possible score.

Figure 5 shows some of the sketches produced by the participants of the Design experiment. These sample sketches are grouped into low and high novelty sets. The measurement method followed is described at length in Shah

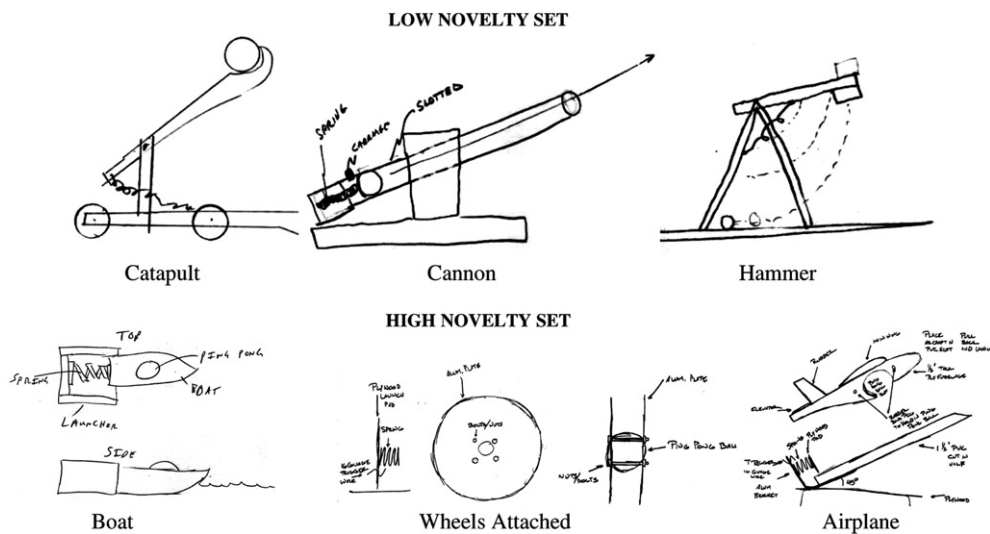
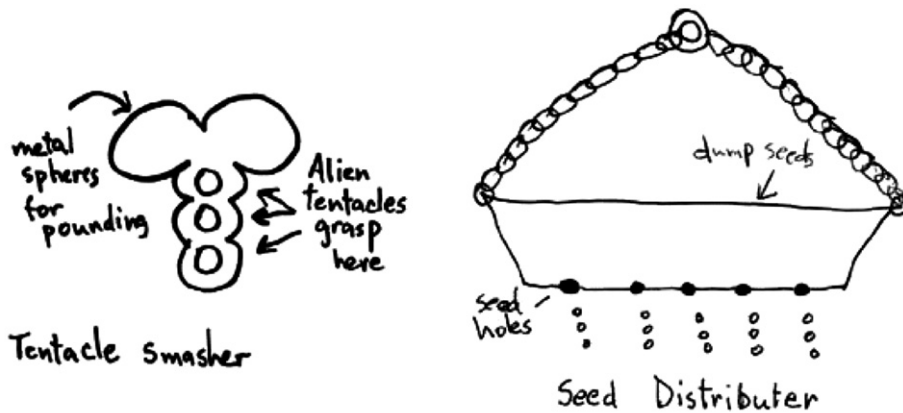


Figure 5 Sample sketches produced in the design experiment

et al. (2000); Shah, S.M. Smith, et al. (2003) and is summarized here. Every idea was first characterized (i.e. solution method for each attribute is described); four attributes were identified from the Design experiment problem statement: Propulsion (i.e. impulse mechanism), Medium (e.g. fly, roll, float), Motion (e.g. sliding, rolling), and Number of parts. For Novelty scoring, the instances of each solution method were counted. The more a particular solution method was used the lower the novelty score assigned. For example, the catapult, cannon and hammer, shown in Figure 5, were more common than the boat, wheel attachment method, and airplane. Each idea's novelty score is computed by multiplying the novelty scores of each attribute by its corresponding weight (e.g. Propulsion = 0.35, Medium = 0.35, Motion = 0.20, and Number of Parts = 0.10). For Variety scoring, the ideas are organized in a genealogy-like tree. Instead of using the four attributes, it was decided to use only the overall function: Ball Throwing. At the highest level ideas are branched according to the physical principle used. Subsequent levels branch the ideas according to working principle, embodiment, and detail differences. The nodes of this tree carry the number of ideas for that category and level. Upper levels have higher variety scores than lower levels. For example, the catapult and hammer shown in Figure 5 use the same working principle (i.e. linear spring potential energy with a lever mechanism) but have different embodiments. The variety score, which applies to the entire group of ideas, is calculated by multiplying each level's score by the number of corresponding branches. For Quality scoring, each idea was assessed with respect to four characteristics: Distance (i.e. estimated achievable distance), Operation (i.e. violation of operation rules defined in the problem statement), Manufacturing (i.e. how difficult it is to construct), and Materials (i.e. comply with the given material list). Because the early state of the concept sketches, judges were employed to score the four characteristics. Each idea's quality score was computed by multiplying the judges' average score for each characteristic by the corresponding weight (e.g. Distance = 0.35, Operation = 0.15, Manufacturing = 0.20, Materials = 0.30). For Quantity scoring, the average number of ideas produced by each individual was calculated.

The sketches shown in Figure 6 are example responses from the tool-generation task used in the Lab experiment. The Low Novelty Set are tools from common categories of generated tools (i.e. 'hand' tools and farm tools), resemble existing tools (i.e. hammer and seed spreader), and use commonly used mechanical principles, so are rated low in novelty. The High Novelty Set, involve less commonly given categories of tools, have no direct existing counterparts, and utilize less commonly used principles. Quality scores were assigned by independent judges who were instructed to use the same standards and the same scale to assess novelty. Inter-judge reliability scores were high, indicating that the independent judges usually gave the same quality score. Novelty was assessed by constructing a master list of all tools generated by all subjects,

## LOW NOVELTY SET



## HIGH NOVELTY SET

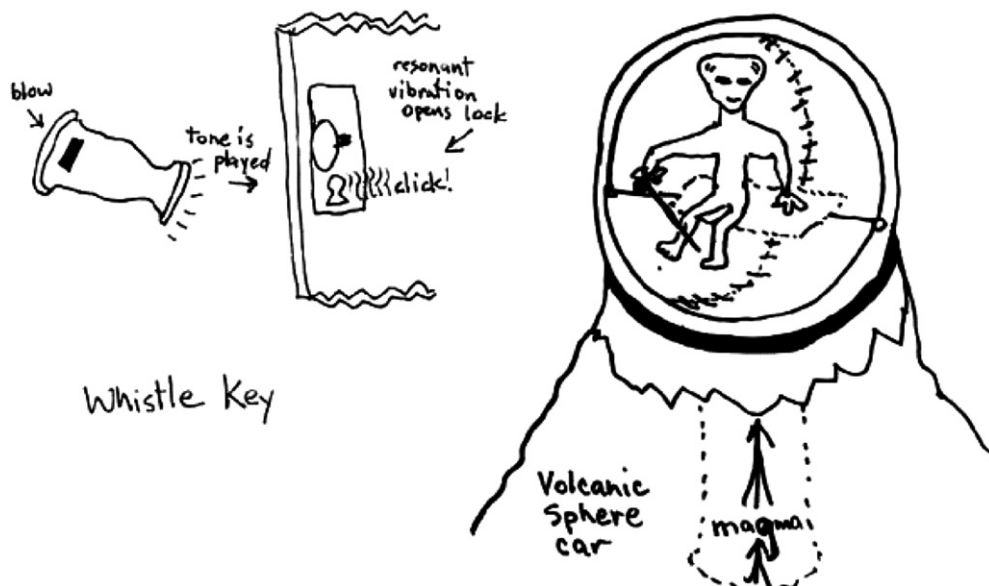


Figure 6 Sample sketches produced in the lab experiment

and then tabulating the frequency of each idea. An idea's novelty score was the frequency divided by the total number of subjects.

### 3 Data analysis

#### 3.1 Hypothesis testing

Data was analyzed using ANOVA with respect to Hypothesis 1, i.e. to determine if an ideation component improves or worsen ideation as measured by each response (i.e. ideation metrics). Hypothesis 2 involves alignment

comparison of results from “equivalent runs” in Design and Lab experiments, that is, runs corresponding to the same ideation component and responses corresponding to the same Effectiveness Metric. For example, Incubation’s Average Quantity (I-AQN) improves *or* worsens simultaneously both Design and Lab experiments will be marked as aligned (i.e. YES), but if it improves *and* worsens in any combination it will be marked as non-aligned (i.e. NO).

### 3.2 Data available

For Design experiments results include the main effects of set 1 (PJR) and set 2 (FIE) and the interaction effects of set 2 (FIE) for the following metrics: Average Quantity (AQN), Average Variety (AVR), Average Quality (AQL), Best Quality (BQL), Average Novelty (ANV), and Best Novelty (BNV). For Lab experiments results include the main effects of set 1 (only for J) and set 2 (FIE) and the interaction effects of set 2 (FIE) for the following metrics: Average Quantity (AQN), Average Variety (AVR), Average Quality (AQL), and Average Novelty (ANV). The alignment comparison between Design and Lab experiments contrasts the main effects of set 1 (J only) for Average Quantity (AQN), Average Variety (AVR), Average Quality (AQL), and Best Quality (BQL), the main and interaction effects of set 2 (FIE) for Average Quantity (AQN), Average Variety (AVR), Average Quality (AQL), and Average Novelty (ANV).

### 3.3 Numerical difference, statistical significance and trends

The numerical difference (or percentage difference) indicates the size of the observed change; this measure alone doesn’t guarantee the reliability of the results. Statistical significance refers to observations that are unlikely to occur by chance and that therefore indicate a systematic cause (Montgomery, 2001). For each contrast a “*t*-test” (also known as ANOVA for one factor) was conducted (2-sample, 1-tail, equal variance). In this case, each run is a sample from two groups. The two sample *t*-test, or *t*-test for two correlated samples, assesses whether the means of two groups are statistically different from each other (Montgomery, 2001). The probability of Type I Error ( $\alpha$ ) is called the significance level and P is the probability of rejecting the hypothesis when the hypothesis is true. The statistical difference significance is calculated using the *t*-test where  $p < 0.05$  is significant and  $p < 0.10$  is borderline significant (marked with \*). Dr. Bernie Bettig’s research group at Michigan Tech independently replicated the statistical analysis conducted at ASU (Kale & Bettig, 2005).

As it will be seen in the results presented, some will be statistically significant and some will not. Even for non-significant results we may be able to observe ‘trends’. Trends are a general indication in which something tends to move (Thompson, 2006). The term ‘trend’ in the behavioral sciences refers to an effect that does not quite reach statistical significance. Because the level of



significance is arbitrarily set at  $p = 0.05$  to avoid excessive Type I errors (rejecting the null hypothesis when no true effect exists) in statistical inferences, there is the danger of Type II errors (failure to reject the null hypothesis when an effect truly exists), particularly in new exploratory research where experimental paradigms have not yet been well-established. Because potentially important and interesting results may be lost in a process that lumps weak effects with non-effects, the term “trend” is consensually understood to represent strong hints at effects that might exist, and that might be worthy of further consideration and research. The behavioral science community has recognized this middle-ground for decades, and understands that trends are not solid, clearly understood phenomena, but rather they indicate hypotheses to be further explored in subsequent research. All comments and observations in this paper based on trends were carefully considered for their value in future research.

### 3.4 Results from ASU design experiments

Figure 7 compares the percentage difference for the means of main effects and interaction effects for Set 1 and Set 2. The effects are represented by rows; single letters (e.g. P) signify main effects of one component and multiple letters (e.g. F:I, FIE) 2 way and 3 way interactions. The horizontal axis represents the size of the effect (i.e. worsen or improve) for each measure (AQN, AVR, etc.) with zero effect at the center.

#### 3.4.1 Main effect of provocative stimuli (*P*)

In general, P benefits Outcome Metrics (Quality and Novelty) while it worsens Process Metrics (Quantity and Variety). A possible explanation is that P produces fixation curtailing the Quantity and Variety of ideas.

#### 3.4.2 Main effect of suspended judgment (*J*)

When comparing level  $-1$  (focus on Quantity) against level  $+1$  (focus on Quality) the Quality metrics (AQL and BQL) improve while the rest of the metrics (ANV, BNV, AQN, and AVR) worsen. The results for Quantity and Quality was as expected since Quantity is requested at level  $-1$  while Quality is explicitly requested at level  $+1$ .

#### 3.4.3 Main effect of flexible representation (*R*)

Flexible Representation improves Novelty (ANV and BNV) and marginally improves Quantity. It may be possible that a flexible representation format allows the designer to explore ideas, producing Novel ones, and not necessarily increasing their Variety and Quality.

#### 3.4.4 Main effect of frame of reference shifting (*F*)

There is some evidence that Frame of Reference Shifting enhances ideation for novelty measures (ANV and BNV) but not for other measures. This result was expected if one considers that F is defined as a way to break implicit and

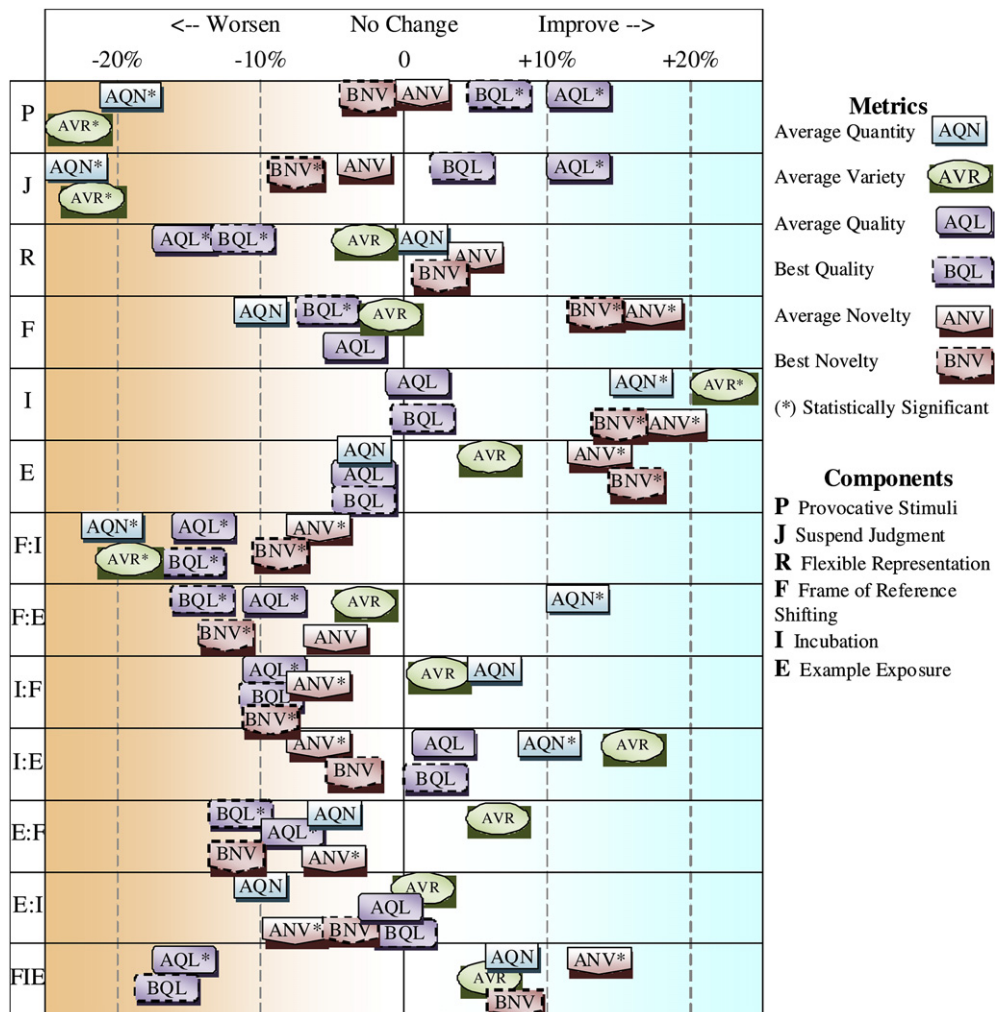


Figure 7 ASU-design experiments results by ideation component

explicit problem constraints by shifting the point of view promoting with this Novel ideas.

### 3.4.5 Main effect of incubation (I)

The effect of Incubation is consistent with hypothesis 1 on all measures in the predicted (positive) direction; average Quality and best Quality are affected to a somewhat lower degree.

### 3.4.6 Main effect of example exposure (E)

Example exposure does not have a universal effect; it improves Novelty and Variety while it marginally worsens Quantity and Quality. Examples can cause conformity, have no effect, or they can make one go into new

directions; it depends on the place one was at in design space prior to seeing the example.

### 3.4.7 Interaction effects for *F* and *I*

Benefits of Incubation not seen if *F* is present and vice versa. The benefits of the two components are not additive, may be redundant. An explanation is that after Incubation, if *F* takes place, the designer gets no benefit from the Incubation. For both Quality measures and both Novelty measures, *F* has a negative effect in presence of Incubation. The effect of Incubation in the presence of *F* was negative on for all measures, except for Variety. Demonstrates the importance of studying interaction effects: both *I* and *F* can be beneficial on their own, but detrimental together.

### 3.4.8 Interaction effects for *F* and *E*

In the presence of Example Exposure, *F* shows a general negative effect except for Quantity. In the presence of Example Exposure, *F* presents a general negative effect except for Variety.

### 3.4.9 Interaction effects for *I* and *E*

In the presence of Incubation, Example Exposure improves all metrics except Novelty. In the presence of Example Exposure, Incubation has an overall negative effect, except for Variety, which is marginally positive.

Figure 8 reorganizes results from ASU-Design experiments by effectiveness metric.

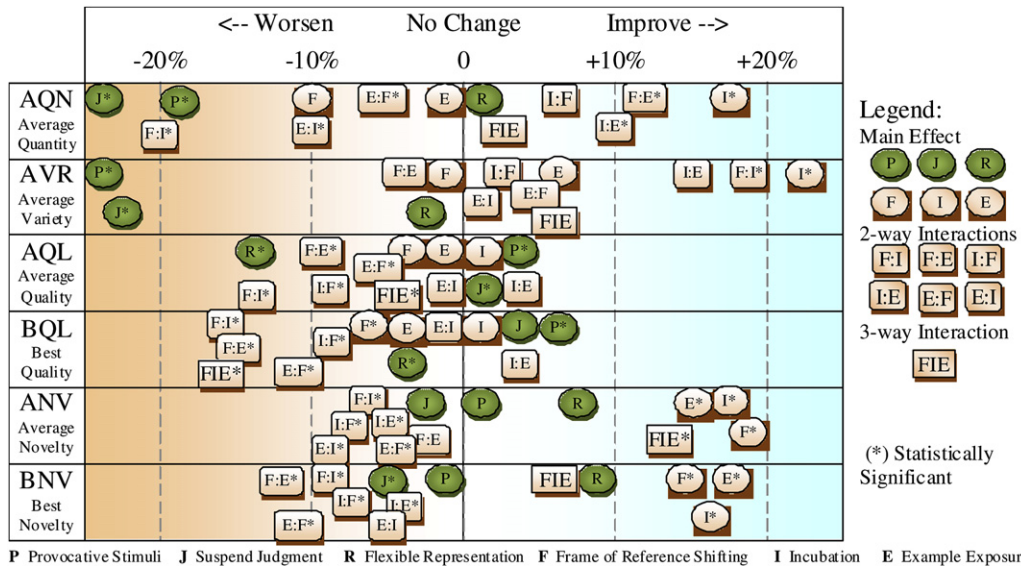


Figure 8 ASU-design experiments results by effectiveness metric

### 3.4.10 Results for quantity (AQN)

In this case Quantity worsens when Suspended Judgment (J) is present as expected since the result compares level -1 (focus on Quantity) against level +1 (focus on Quality). Results are diverse, while P, J and F:I present considerable decline, I and F:E present improvement.

### 3.4.11 Results for variety (AVR)

Variety sharply declines for P and J while it clearly improves with Incubation. Results are diverse, while P, J and F:I show a strong decline, I and I:E show a strong improvement.

### 3.4.12 Results for quality (AQL and BQL)

Quality, in general, tends to worsen for most components. Provocative Stimuli (P) and Judgment (J) improve Quality as expected since the result compares level -1 (focus on Quantity) against level +1 (focus on Quality).

### 3.4.13 Results for novelty (ANV and BNV)

P and J worsen Novelty while R improves it. All set 2 (FIE) main effects improve Novelty, while all two-way interactions worsen it. From [Figure 8](#), it seems difficult to increase novelty; this may reinforce the notion that Outcome metrics (Novelty and Quality) are harder to improve than Process Metrics (Quantity and Variety).

## 3.5 Results from TAMU lab experiments

[Figure 9](#) compares the percentage difference for the means of main effects and interaction effects for Set 1 and Set 2.

### 3.5.1 Main effect of suspend judgment (J)

The effect on Quantity is negative and the effect on Quality is positive; this is as expected since the result compares level -1 (focus on Quantity) against level +1 (focus on Quality). Also, it can be seen how Outcome metrics (Quantity and Variety) worsen while Process metrics (Quality and Novelty) improve.

### 3.5.2 Main effect of frame of reference shifting (F)

The effect is positive for Quantity and Quality while detrimental for Variety and Novelty measures. This result did not produce the expected positive effect on Novelty, but in general, the results for all four metrics only marginally improved or worsened for F.

### 3.5.3 Main effect of results for incubation (I)

The simple main effect of Incubation is marginally positive for all metrics, except for Novelty, which is negative; with this only exception, the effect is as expected.

### 3.5.4 Main effect of results for example exposure (E)

The simple main effect of Example Exposure benefited Outcome Metrics (Quantity, Variety) with a clear negative effect on Process Metrics (Quality

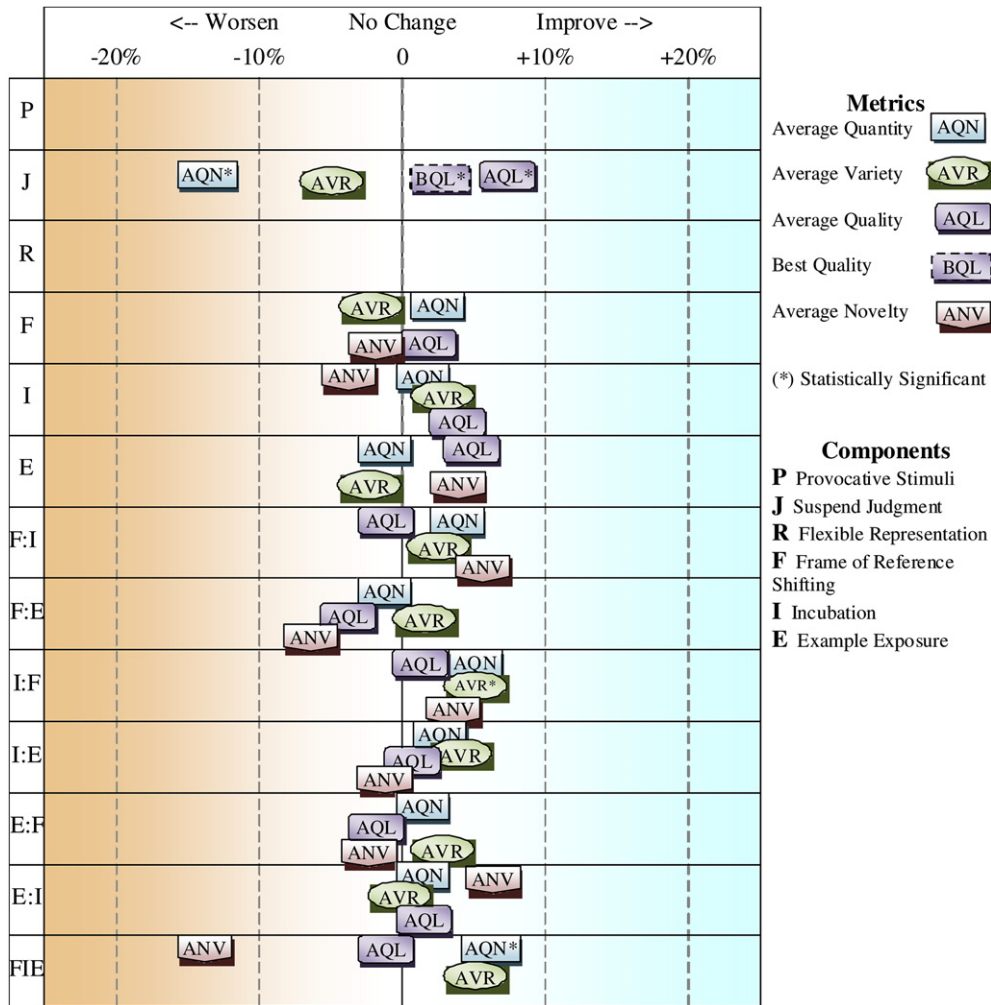


Figure 9 TAMU-lab experiments results by ideation component

and Novelty). In general, it was expected that E would improve at least Quality, but as mentioned earlier, the effect of Examples depends on the place one was at in design space prior to seeing the example

### 3.5.5 Interaction effects for F and I

In the presence of F, Incubation had a beneficial effect on all metrics, except Quality. In the presence of Incubation, F has beneficial effect on Quantity, Variety and Novelty, and a detrimental effect on Quality.

### 3.5.6 Interaction effects for F and E

In the presence of F, Example Exposure's effect is mostly negative for all metrics. In the presence of Example Exposure, the effect of F is

inconclusive in general with marginal improvements and declines for all metrics.

3.5.7 Interaction effects for I and E

In the presence of Incubation, Example Exposure’s effect is positive for all metrics except for Novelty. In the presence of Example Exposure, Incubation’s effect is (at least marginally) positive for all metrics.

Figure 10 reorganizes the results from Figure 6 by Effectiveness Metric.

3.5.8 Results for quantity (AQN)

In this case, Suspend Judgment has a negative effect on Quantity as expected since the result compares level –1 (focus on Quantity) against level +1 (focus on Quality). For set 2, the effects are mostly positive (marginal exceptions for E and F:E).

3.5.9 Results for variety (AVR)

The effect of Judgment is negative as expected. For set 2 main effects are negative, except I, while all interactions are (at least marginally) positive.

3.5.10 Results for quality (AQL)

The Quality improvement for Suspend Judgment (J) is consistent with the hypothesis prediction since the result compares level –1 (focus on Quantity)

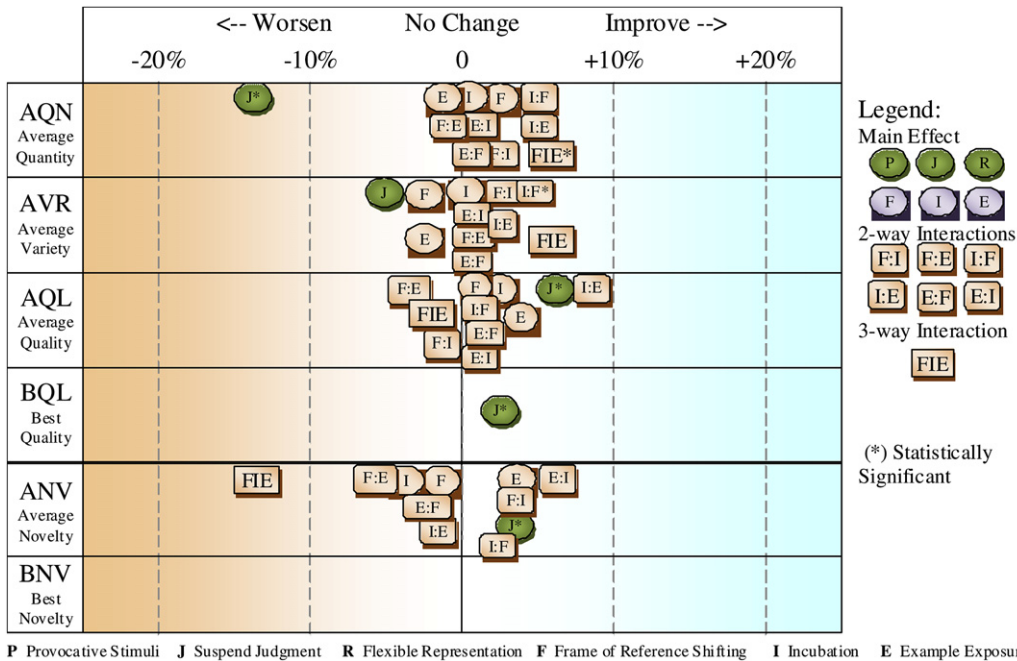


Figure 10 TAMU-lab experiments results by effectiveness metric



against level +1 (focus on Quality). The results for set 2 show marginal improvement or decline, except for Incubation and Exposure, which are consistent with the hypothesis prediction.

### 3.5.11 Results for novelty (ANV)

Results are assorted, improvements and declines make difficult to make a generalization.

## 3.6 Alignment of ASU design and TAMU lab experiment results

Available results from ASU-Design experiments and TAMU-Lab experiments are compared in Table 4. A ‘YES’ indicates matching trends for a particular metric/effect (e.g. I/AQL) between Design and Lab Experiments (i.e. both positive or both negative) while a ‘NO’ indicates a mismatch in the trends (i.e. one positive and one negative). For example, the Incubation (I) effect for the metric Quantity (AQL) has positive matching trends for Design experiments (See Figure 4) and Lab experiments (See Figure 6) and hence is marked “YES” in Table 1 while Novelty is marked with a “NO”. The total alignment percentages are calculated dividing the number of matches (i.e. “YES”) by the number of available comparisons to give an overall indication of the alignment agreement.

Based on Table 4 results, the overall alignment percentage is 59%. It’s worth mentioning the relative high alignment of I, I:E, E:F,FIE, and specially J with a perfect alignment; F, F:I, and E:I present the lowest alignment levels with 25% each. Variety and Quality present the highest alignment percentages

**Table 4 Comparison of trends between ASU-design and TAMU-lab experiments**

<i>Effect compared</i>	<i>Effectiveness metrics</i>						<i>Total</i>
	<i>Process</i>		<i>Outcome</i>				
	<i>AQN</i>	<i>AVR</i>	<i>AQL</i>	<i>BQL</i>	<i>ANV</i>	<i>BNV</i>	
P							
J	YES	YES	YES	YES			4/4(100%)
R							
F	NO	YES	NO		NO		1/4(25%)
I	YES	YES	YES		NO		3/4(75%)
E	YES	NO	NO		YES		2/4(50%)
F:I	NO	NO	YES		NO		1/4(25%)
F:E	NO	NO	YES		YES		2/4(50%)
I:F	YES	YES	NO		NO		2/4(50%)
I:E	YES	YES	YES		YES		4/4(100%)
E:F	NO	YES	YES		YES		3/4(75%)
E:I	NO	YES	NO		NO		1/4(25%)
FIE	YES	YES	YES		NO		3/4(75%)
Total	6/11(55%)	8/11(73%)	7/11(64%)	1/1(100%)	4/10(40%)		26/44(59%)



(73% and 64% respectively) while the lowest is Novelty (40%). The ramifications and causes are presented in the discussion section.

### *3.7 Hypothesis testing summary*

Although Hypothesis 1 states that every ideation component (i.e. P, J, R, F, I, and E) improves each metric (AQN, AVR, AQL, ANV, and BNV), the authors did not predict this hypothesis to be true for all ideation components and effectiveness metrics; it simply was a starting point for the experimental design. Results from ASU-Design and TAMU-Lab experiments clearly show a mixture in the trends of the results, some positive, some negative, and with different magnitudes. Only in two cases the hypothesis seemed true: Incubation (I) from ASU-Design experiments (see [Figure 7](#)) and the two-way interaction F:I from TAMU-Lab experiments (see [Figure 9](#)). In both cases, the effect of incubation (I) is consistent with the hypothesis prediction for all metrics. But even in these cases, only some of the results were statistically significant. As mentioned earlier, the authors recognize the meaning of statistically significance results, nevertheless, important things can be learned from non-statistically significant results; one of them are trends. In general and regardless of the initial hypothesis, these results show the actual effect each component has for each metric. Because of the alignment approach, future work will focus on attempting to explain these effects based on related cognitive processes.

Hypothesis 2 refers to the alignment of results between ASU-Design and TAMU-Lab experiments. The authors predicted that all or most of the results to be aligned, but the results show only partial alignment. From [Table 4](#) it can be seen that the trends between levels of available results matched in 59% of the cases, and only in some cases there was common statistical significance. Since the overall objective of the study is to gain insight into design ideation, these results were used to try to understand the reasons for misalignment; the sources of the misalignment, discussed in the next section, and the derived discoveries represent a valuable guide for our future work.

## *4 Discussion*

This is the first study of its kind; aligned experiments conducted at two levels for a set of ideation components and outcomes measured by objective measures. The study provided insights both into the experiment methodology and certain ideation mechanisms. As an initial point for hypothesis testing, it is expected that every ideation component main and interaction effect improves each effectiveness metrics. But the main objective is not to prove an expected effect, but to discover what the effect is and understand the reasons behind it. As an example, it was expected for Incubation to improve all metrics as predicted by the hypothesis, this occurred for all Design experiments and for most Lab experiments (Novelty was a marginal exception). A clear example of expected results was Suspend Judgment, Quantity declined while

Quality improved as expected. The rest of the results were divided and have been analyzed and commented individually.

#### *4.1 Process vs. outcome metrics*

Process metrics (i.e. quantity and variety) are easier to improve than outcome metrics (i.e. quality and novelty). In past studies the authors treated quantity, quality, novelty, and variety as equal. However, the authors now recognize that just having a lot of ideas or a great variety of ideas is in itself not the end goal of a designer. A designer cares about novelty and quality. It has long been widely believed that generating more ideas (quantity) or exploring many types of alternatives (variety) leads to better ideas (novelty, quality). Thus, quantity and variety are means to the end, while novelty and quality are the end goals. Therefore, it was decided to label quantity and variety as process metrics and novelty and quality metrics as the primary outcomes. The results support this decision: Process metrics were improved in 60% (29 out of 48 cases) of the cases while outcome metrics only in 42% (31 out of 73 cases).

#### *4.2 Average vs. best scores*

For quality and novelty, as expected, the average scores follow the best scores in 96% (26 out of 27 cases) of the cases.

#### *4.3 Quality vs. quantity*

Quality followed quantity in 54% (13 out of 24 cases) of the cases. This partially supports some cognitive theories of ideation, such as the Darwinian (Simonton, 2003) and Remote Association (Mednick, Mednick, & Mednick, 1964) models, which relate quantity to quality of ideas generated.

#### *4.4 Manipulation of ideation components*

It was observed that some ideation components (e.g. incubation and judgment) had more of the expected behavior (i.e. incubation improving all metrics while judgment improving quality and decreasing quantity). A possible reason is that, compared to other components, the experimental treatment for incubation and judgment was easier to manipulate; incubation was simply defined as an interruption in the ideation process, and judgment was defined as a request for students to focus either on quality or quantity.

#### *4.5 Strength of ideation components*

Some ideation components show a stronger effect, this is evident when interacting with other (weaker) components. For example, frame of reference shifting has a strong negative main effect, and this persists even when interacting with incubation or example exposure.

#### *4.6 Complexity of ideation components*

Interactions between components are not necessarily predicted by superposing two main effects. For example, the separate main effects on novelty of

incubation and example exposure is positive, but novelty decreases during interaction. This suggests that the interactions may produce compound effects such as synergistic or canceling effects.

#### *4.7 Ideation principles more fundamental than ideation components*

It is possible that ideation components be grouped into higher-level ideation principles according to their effect in the ideation process. For example, F and E may belong to the same ideation principle of abstraction, since they allow the user to overcome fixation by putting the designer in a different context, by using examples or shifting the reference of the problem.

#### *4.8 Alignment/misalignment*

Bias and distortion may affect the results. The sources for bias and distortion can be grouped into Reliability errors and Validity Errors.

Reliability relates to the consistency of the experiment, by following the same experimental model, can one obtain the same results? Is the experiment repeatable? Typical sources of reliability error are: sampling size, noise from uncontrolled variables, among others. Validity relates to the accuracy of the experiment model; does it measure what is supposed to measure? Reliability and Validity are independent; it is possible that the experimental results are reliable, in other words, these are repeatable and with a high level of statistical significance, and at the same time not valid or correct because the experiment is not correctly modeling the desired phenomenon. It is also possible that the experimental results are valid; correctly modeling the effects of the ideation components, but due to various reasons, the results are not repeatable.

With respect to reliability error, it can be said that these first experiments will serve as a precedent for more reliable future DOE. The statistical significance is low in general, although it has been explained previously the importance of trends in this research. Although ideation experiments were conducted in the past, there was no precedent on the magnitude and complexity of these series of experiments. Sampling size may be an issue since experiments were conducted on specific sections of engineering design or psychology students. It was also observed that the classroom setting presents various challenges. Every section was different from each other, the instructor, the time of the semester, the classroom, the current schedule of exams, the day of the week, etc. Added to this is the human factor; every participant student/subject opens a possibility of complexities: background, experience, IQ level, personality, etc. This experiment considered these variables out of scope and relied on the possibility of averaging all these differences.

Even with a perfectly reliable experiment, the question still exists: Is this experiment really testing what is intended? As explained in earlier sections, the

DOE intended to measure the effects of various ideation components. The authors consider that the results presented here are valid for the particular manipulation of dependent and independent variables and generalizations can be made after careful considerations. On one hand, the results for incubation (as an example) are valid for the type of design and lab problems used, for the amount of time assigned to incubate, for the particular characteristics of the engineering and psychology students generating ideas, etc. Nevertheless, these results (the first of its kind aligning two levels) provide the necessary insight into ideation and also sufficient understanding for future work improvements.

When comparing across levels, it is a measure of our overall “alignment” approach, and there may be some reasons why this can’t be achieved. An alignment reliability error may be compounded by each Design and Lab reliability errors. Also, there may be a validity error while attempting to compare these two levels; maybe there is something in the nature of the two levels (Design and Lab experiments) that makes it difficult or impossible to compare. The detail on these causes of misalignment; are important for the validity of this research and for the improvement of future research.

## *5 Conclusions*

This paper presented experimental results for six ideation components simultaneously tested at the Design and Lab levels. It must be stated that although the conclusions drawn from this work are limited to the predefined scope (novice designers working on simple design tasks), the results provide valuable insight into the effects of ideation components and the possibility of alignment between Design (engineering) experiment and Lab (cognitive psychology) experiments. With respect to the ideation components, Provocative stimuli (P) and suspend judgment (J) present similar effects. Suspend judgment clearly improved quality and decreased quantity. Frame of reference shifting (F) and Example Exposure (E) have similar effects. Incubation shows a constant improvement on all metrics. Frame of reference shifting has a strong effect even when interacting. With respect to the alignment, Suspended judgment (J) had a perfect alignment (100%). For Set 2, the overall alignment was 80%, with Incubation (I) having the best alignment. The overall results for both sets indicate a level of alignment of 59%, but what may be most useful is derived from the other 41%. Evaluating the reasons of no-alignment provided insight into various relevant issues such as: ideation principles as more fundamental than ideation components, better understanding of ideation components (some are stronger, some are easier to manipulate, interactions are complex, etc.), distinction between process and outcome metrics (outcome being harder to improve than process), quantity to quality match (more ideas means higher quality), addition of new ideation metrics (Best Novelty and Best Quality). The experience gained will be useful when improving the experimental procedure for future experiments. Future research will be geared towards

the development of a design ideation model based on experimental results and cognitive models.

## 6 *Future work*

Further research is needed to refine the alignment approach. The experimental results will generate the knowledge needed to not only evaluate ideation methods but also to distinguish between their necessary and superfluous components. The understanding of the interaction of human variables, method variables and design problem attributes, and the relationship of ideation processes to design outcome, will help companies determine which method to use under given conditions and how to constitute design teams. This study could also provide guidance to educators in finding better ways of teaching design synthesis. For example, (1) it will be easier for educators to teach ideation since students will have a better understanding of the effects of ideation components and principles. (2) Understanding the effect of ideation components and principles will reinforce the students' control of their own ideation processes. (3) Different ideation methods will provide different outcomes; knowing this, students will learn which ideation method to use depending on their outcome needs. And (4) if students understand that quantity and variety (i.e. process metrics) are easier to improve, their focus will shift to quality and novelty (i.e. outcome metrics) of ideas. In general, improving the understanding of ideation components, principles, and outcomes, can be seen as an intermediate step towards the development of new theoretically based design idea generation methods to replace ad-hoc methods, while a long-term goal is the creation of a model of design ideation. Present and future results are an important step in the creation of a model of design ideation that relates real world ideation methods, theories and models from cognitive psychology, and effectiveness metrics from engineering design, by having ideation components (and ideation principles) as a common central element. Based on the major findings discussed in this paper, various improvements will be included in future research in order to reduce accuracy errors and modeling errors. Once this model is established, its implementation will be explored in the area of Computer Aided Conceptual Design (CACD) for the intuitive generation of conceptual ideas (Vargas-Hernandez & Shah, 2004).

## *Acknowledgements*

The authors gratefully acknowledge the support of the Ford Motor Co. and NSF grant DMI-0115447. All opinions expressed are those of the authors and not endorsed by NSF.

## *References*

- Altshuller, G. (1984). *Creativity as an exact science*. New York, NY: Gordon and Breach.
- Christiaans, H., & Dorst, K. (1991). An empirical study into design thinking. In *Research in design thinking*. Delft University Press.

- Christiaans, H., & Venselaar, K. (1991). Practical implications of a knowledge based design approach. In *Research in design thinking*. Delft University Press.
- Cross, N., Dorst, K., & Roozenburg, N. (1991). Research in design thinking. In *Proc. of workshop*. Delft: Univ. Press.
- Dennehy, E. B., Bulow, P., Wong, F., Smith, S. M., & Aronoff, J. B. (1992). *A test of cognitive fixation in brainstorming groups*. Boston, MA: Eastern Psychological Association.
- Ericsson, K., & Simon, H. (1984). *Protocol analysis – Verbal reports as data*. Cambridge, MA: MIT Press.
- Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition – Theory, research, and applications*. Cambridge, MA: MIT Press.
- Hale, C. (1987). *Analysis of the engineer design process in an industrial context*. Cambridge, MA: Grant Hill Pubs.
- Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies*, 12, 3–11.
- Kale, V., & Bettig, B. (2005). *DOE analysis of creative design ideation metrics*. Technical Report. Design and Manufacturing Computations Laboratory, Mechanical Engineering—Engineering Mechanics Dept., Michigan Technological University.
- Kan, J. W. T., & Gero, J. S. (2005). Can entropy represent design richness in team designing? In A. Bhatt (Ed.), *CAADRIA '05*, Vol. 1 (pp. 451–457) New Delhi: TVB.
- Kulkarni, S. V. (1998). *Survey of creativity models*. Technical Report ASU/DAL/IG/98-1. Tempe, AZ: Arizona State University.
- Kulkarni, S. V. (2000). A framework for the experimental evaluation of idea generation techniques, *MS Thesis*. Tempe, AZ: Arizona State University.
- Kulkarni, S. V., & Shah, J. J. (1999). *Survey for evidence of components of creativity*. Technical Report ASU/DAL/IG/99-7. Tempe, AZ: Design Automation Lab, Arizona State University.
- Leifer, L. (1996). *Presentation at DARPA-RaDEO Contractors Meeting*, Melbourne, FL.
- Marples, D. (1960). *The decisions of engineering design*. London: Inst. of Eng. Designs.
- Mckoy, F., Vargas-Hernandez, N., Summers, J. D., & Shah, J. (2001). Experimental evaluation of engineering design representation on effectiveness of idea generation. *Proceedings ASME Design Theory and Methodology Conference*, Pittsburgh, PA.
- Mednick, M. T., Mednick, S. A., & Mednick, E. V. (1964). Incubation of creative performance and specific associative priming. *Journal of Abnormal & Social Psychology*, 69(1), 84–88.
- Montgomery, D. C. (2001). *Design and analysis of experiments* (5th ed.). NY: John Wiley and Sons.
- Nagy, R., Ullman, D., & Dietterich, T. (1993). A data representation for collaborative mechanical design. *Research in Eng. Design*, 3.
- Pahl, G., & Beitz, W. (1996). *Engineering design – A systematic approach* (2nd ed.). London: Springer.
- Schön, D. (1991). Teaching and learning as a design transaction. In *Research in design thinking*. Delft University Press.
- Shah, J. (1998). Experimental investigation of collaborative techniques for progressive idea generation, P. *Proceedings ASME Design Theory and Methodology Conference*, Atlanta, GA.
- Shah, J. J., Kulkarni, S. V., & Vargas-Hernandez, N. (2000). Guidelines for experimental evaluation of idea generation methods in conceptual design. *Journal of Mechanical Design*, 122(4), 377–384.

- Shah, J. J., Smith, M. S., Vargas-Hernandez, N., Gerkens, D. R., Wulan, M. (2003). Empirical studies of design ideation: alignment of design experiments with lab experiments. *Proceedings ASME Design Theory and Methodology Conference*, Chicago, IL.
- Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design Studies*, 24(2), 111–134.
- Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2005). *Development and validation of designing ideation models for conceptual engineering design: Final technical report*. Technical Report ASU/DAL/IG/05-1. Tempe, AZ: Arizona State University.
- Shah, J. J., Vargas-Hernandez, N., Summers, J. D., & Kulkarni, S. (2001). Collaborative sketching (C-sketch) – an idea generation technique for engineering design. *Journal of Creative Behavior*, 35(3), 1–31.
- Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives. *Psychological Bulletin*, 129(4), 475–494.
- Smith, S. M. (2000). November, did that really happen, or was it just a dream? *Paper presented at the annual meeting of the Psychonomic Society*, New Orleans, LA.
- Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: Guilford Press.
- Ullman, D., Dietterich, T., & Stauffer, L. (1988). A model of the mechanical design process based on empirical data. *AI EDAM*, 2(1).
- Ullman, D., Stauffer, L., & Dietterich, T. (1987). Preliminary results on an experimental study of mechanical design. *NSF Design Theory and Methods Conference*.
- Ullman, D., Wood, S., & Craig, D. (1989). The importance of drawing in mechanical design process. *NSF DTM Conference*.
- VanGundy, A. B. (1988). *Techniques of structured problem solving* (2nd ed.). NY: Van Nostrand Reinhold Company.
- Vargas-Hernandez, N. (2007). A basis for the development and validation of design ideation models for conceptual engineering design, *Ph.D. Dissertation*, Arizona State University, Tempe, AZ.
- Vargas-Hernandez, N., & Shah, J. J. (2004). 2nd-CAD: a tool for conceptual systems design in electromechanical domain. *JCISE Special Issue in Conceptual Design*, 4(1), 28–36.
- Waldron, M., & Brook, R. (1994). Analyzing inter and intra group information exchanges in conceptual collaborative design. *Proceedings ASME Design Theory and Methodology Conference*.
- Wallas, G. (1926). *The art of thought*. New York: Harcourt.
- Ward, A., & Sobek, D. (1996) Principles of Toyota's set based concurrent engineering process. *Proceedings ASME Design Theory and Methodology Conference*.
- Wilde, D. J. (1999). Design team roles. *Proceedings ASME Design Theory and Methodology Conference*.