

Applied Tests of Design Skills— Part 1: Divergent Thinking

Jami J. Shah

Mechanical & Aerospace Engineering,
Arizona State University,
Tempe, AZ 85287

Roger E. Millsap

Department of Psychology,
Arizona State University,
Tempe, AZ 85287

Jay Woodward

Department of Educational Psychology,
Texas A&M University,
College Station, TX 77843

S. M. Smith

Department of Psychology,
Texas A&M University,
College Station, TX 77843

A number of cognitive skills relevant to conceptual design were identified previously. They include divergent thinking (DT), visual thinking (VT), spatial reasoning (SR), qualitative reasoning (QR), and problem formulation (PF). A battery of standardized tests is being developed for these design skills. This paper focuses only on the divergent thinking test. This particular test has been given to over 500 engineering students and a smaller number of practicing engineers. It is designed to evaluate four direct measures (fluency, flexibility, originality, and quality) and four indirect measures (abstractability, affixability, detailability, and decomplexability). The eight questions on the test overlap in some measures and the responses can be used to evaluate several measures independently (e.g., fluency and originality can be evaluated separately from the same idea set). The data on the twenty-three measured variables were factor analyzed using both exploratory and confirmatory procedures. A four-factor solution with correlated (oblique) factors was deemed the best available solution after examining solutions with more factors. The indirect measures did not appear to correlate strongly either among themselves or with the other direct measures. The four-factor structure was then taken into a confirmatory factor analytic procedure that adjusted for the missing data. It was found to provide a reasonable fit. Estimated correlations among the four factors (F) ranged from a high of 0.32 for $F1$ and $F2$ to a low of 0.06 for $F3$ and $F4$. All factor loadings were statistically significant. [DOI: 10.1115/1.4005594]

1 Introduction

What sets good designers apart from mediocre ones? Is it just experience and domain knowledge, or is there a skill set? Academics and practitioners seem to have an awareness that good designers possess more than just vast domain knowledge; they have certain abilities that make them more effective in using that knowledge to structure ill-defined problems, construct fluid design spaces to facilitate fluency and flexibility of generating solutions and visualizing the detailed working of artifacts in their imagination. Although design skills are alluded to in design textbooks and curricula, there has not been a concerted effort to explicitly identify and measure them.

For the past several years, the principal authors group has been engaged in identifying and characterizing in formal terms, a set of skills found in good engineering designers [1]. We also devised objective measures of these skills. We define a *skill* as the cognitive ability to perform a task. Design skills were derived from observations of design tasks as well as from past cognitive studies [1]. A good designer or design team must possess a wide range of skills to tackle different phases of product development.

From our past work and that of the others, we identified the following design skills: divergent thinking, convergent thinking, deductive, inductive, and abductive reasoning, spatial reasoning, visual thinking, analogical reasoning, sketching, qualitative reasoning, decision-framing and decision making, and designing and conducting simulated or real experiments. Not all of these are independent or unique skills; for example, there is an inexplicable relation between deductive reasoning and convergent thinking, and also between visual thinking and spatial reasoning. Pattern recognition and analogical reasoning may be interpreted in terms of physical, behavioral, or linguistic context, thus being part of visual thinking, spatial reasoning, or qualitative reasoning.

We are now developing standardized tests for a subset of these skills, those related particularly to conceptual design. Our team

consists of an engineer, a cognitive psychologist, an educational psychologist, and a psychometric consultant. We have so far developed tests for DT and VT. Future plans include tests for PF and QR. Possible applications of these tests include evaluating students in design classes, forming of balanced design teams which possess skills necessary for a given project, and evaluating the effectiveness of design courses and curricula.

We previously reported on the construction of the DT and VT tests [2] and preliminary data. This paper focuses on detailed studies of the DT test, test results, data analysis, and reliability studies. We discuss the continuous improvement of the test based on the results of the collected data.

2 DT Test Development

The basis and motivation for the DT test have been reported in our ICED09 paper [2]. Here, we give a detailed account of its contents and rationale.

2.1 DT Measures. In the context of design, DT is commonly defined as the ability to generate many alternative solutions, i.e., the ability to explore the design space. Good designers understand that design space is not fixed; as they generate and explore ideas and get insight into the structure of the space, they continually find ways to expand the space by redefining and restructuring the problem [3,4] illustrated conceptually in Fig. 1(a). Thus, the number of ideas generated (*fluency*) can be one measure of DT.

The number of ideas generated in the course of ideation (i.e., quantity) has always been a key measure of creative productivity [5,6] and makes sense in terms of the Darwinian theory of creativity [7–9], which sees blind (or chance) variation and selective retention as the way that creative ideas emerge and survive. Using only the number of ideas generated as a measure of DT, however, is inadequate because there could be many superficial variations of the same basic design. Therefore, a measure of *variety* (often termed *flexibility*, e.g., Refs. [5], [6], and [10]) is needed to determine how broadly the design space has been explored. From a cognitive science point of view, variety in idea generation is a measure of the number of categories of ideas that one explores

Contributed by the Design Education Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received September 1, 2010; final manuscript received November 21, 2011; published online February 3, 2012. Assoc. Editor: Janis Terpenney.

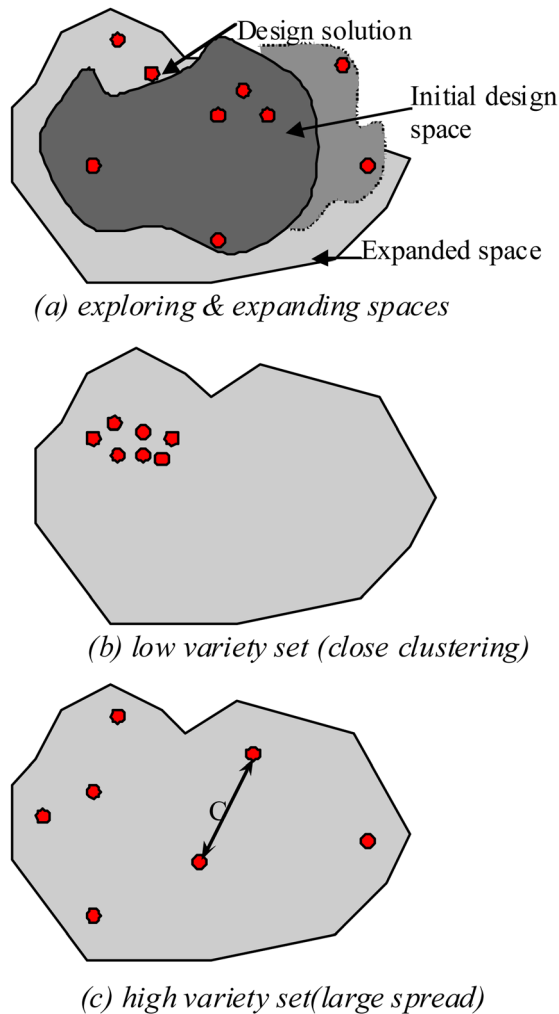


Fig. 1 Abstract representation of design solution points in design spaces

[10]. In Figs. 1(b) and 1(c), there are the same number of ideas (represented as points in design space), but in Fig. 1(b) they are clustered closely, indicating small conceptual variations, leaving the vast design space unexplored. In Fig. 1(c), ideas span a broader spectrum. Researchers often speak of “conceptual distance, C ” as a measure of the extent of differences between ideas. For example, near or local analogies draw upon similar features or relations from the same conceptual domain in which one is working, whereas remote analogies draw from conceptually distinct domains of knowledge [11–15].

While quantity and variety of concepts measure the skill to explore design space, there is another element that needs to be considered: the ability to expand the design space (thinking outside the box). This ability can be measured by the *originality* or novelty of the solutions. In terms of design space, novel designs occupy points that are initially not perceived to be within the design space. In fact, creativity and originality are often considered as interchangeable terms. Creative ability can be measured by the originality of ideas that an individual generates. Expanding the design space offers the opportunity to find better designs that have so far not known to exist. Many idea generation methods provide deliberate mechanisms to view the problem in a different way, to use analogies and metaphors, to play around by loosening the tight grip on goals that engineers generally have. The degree of novelty is a relative measure that requires either a comparative assessment of a set of designs or an enumeration of what ideas are expected with what frequency.

Often one finds that routine approaches to problems can lead to uncreative ideas. In such cases, the original cognitive knowledge structures applied to a problem are inappropriate and insight can be achieved only through what cognitive psychologists have called *cognitive restructuring* [16–19]. The ability to generate a wide variety of ideas is directly related to the ability to restructure problems and is therefore an important measure of creativity in design. Reformulation of problems is facilitated by the ability to abstract or generalize [20–25]. Researchers studying the use of analogical reasoning in design point to the ability to abstract as the key to make connections between the entities across domains [26]. Thus, the ability to abstract (“abstractability”) is an indirect measure of divergent thinking.

The biggest difference between technological and artistic creativity is that in the former there are particular goals or specifications that must be met within certain constraints. Goodness of fit with design specifications is a measure of quality of an idea. Therefore, in engineering we need to “qualify” ideas and not just use fluency, i.e., the ability to generate good ideas that are technically feasible and practical needs to be considered. We term this skill “practicality/quality.”

Design fixation has been identified as a common block to creativity; it is the tendency of a designer to favor a design from previous experience, a design seen or developed by the designer [27]. A symptom of fixation is that new designs share more common features with previous designs. Many design researchers have shown the existence of design fixation [27–31]. Our own studies have demonstrated designers’ susceptibility to design fixation and show how taking breaks from problems can alleviate fixation [32–34]. It is, therefore, important to measure the ability to avoid fixation (“afixability”) on a DT test.

Two other subskills may be of interest. The ability to decompose, decouple complex problems, to identify key issues and conflicts is a mark of good designers. Protocol studies by many different groups on identifying differences between experts and novices have shown this [35,36]. We term this ability as “decomposability” in this paper. Last, being able to think about the workings of a device in a particular environment would certainly have advantages in producing good quality ideas. Gardner terms this “vivid thinking” and may be an indirect measure of DT [37]. We term this skill as “detailability” in this study and measure it by the extent of elaboration in design description.

Based on the above, DT subskills, their definitions, and measures are summarized in Table 1. We have split these into two groups: *direct* and *indirect measures*. Direct measures (fluency, variety, originality, and quality) are ones that can be assessed from a set of ideas generated by an individual. Indirect measures are those that are assumed to aid ideation (afixability, abstractability, and decomposability). They are related more to cognitive processes than outcomes, so questions need to be designed to specifically measure them instead of looking at design ideas generated. Generalized methods for objective evaluation of the direct measures can be found in Ref. [38]. The adaptation of these measures and the assessment of indirect measures in the context of the DT test questions will be presented in Sec. 3.2.

2.2 Survey of Standard Creativity Tests. We examined eight standardized tests of creativity to see the extent to which the above skill indicators are represented [39]. They include Abbreviated Torrance 2002 [40], Meeker test [41], Meeker SOI checklist [42], Torrance [43], Guilford alternative uses [44], Wallach and Kogan [45], and Guilford ARP [46].

The Abbreviated Torrance test uses three divergent thinking activities that represent a merging of Torrance’s previously established verbal and figural batteries. Fluency is defined as a simple count of the number of pertinent responses. Examiners must read each response and make a judgment as to whether it is relevant to the “just suppose” situation. For every relevant response, 1 point is awarded. Originality is defined as the ability to produce ideas

Table 1 DT subskills and measures

Subskill	Definition	Metric
Direct		
Fluency (flu)	Ability to generate many solutions consistently	Quantity of ideas generated
Flexibility (flx)	Ability to explore design space in many directions	Variety of ideas generated
Originality (org)	Ability to “think outside the box,” generate unexpected solutions	Originality of ideas generated
Quality (qlty)	Ability to consider technical, manufacturing, and economic feasibility	Closeness of fit with design goals; tech and economic feasibility
Indirect		
Afixability (afx)	Ability to get out of ruts, not get fixated to past or current solutions	Conceptual distance from exposed example
Abstractability (abst)	Ability to make connections, find relationships, analogies	Number and remoteness of discovered relations
Decomplexability (Demp)	Ability to handle complex problems	Level of decomp, decoupling
Detailability (dtl)	Ability to think at detailed level	Elaboration, embellishments, clarity

that generally are not produced, or ideas that are totally new or unique. The examiner must compare each response to the provided list of “common responses.” For every response that the subject gives that is not on this list, 1 point is awarded. Credit is also given for emotions and humor. Two of the activities are visually oriented, thus mixing up divergent and visual thinking. The Meeker 2000 test also contains activities involving figures and some activities rely on English vocabulary. These characteristics are deemed inappropriate for our use. Another dimension in the Meeker test is making, manipulating, and interpretation of symbols with no particular goal or constraint. One question that we found of relevance was finding relationships between entities listed as words. The ability to make connections between seemingly unrelated entities is a mainstay of many design ideation techniques [45]. William’s creativity packet gives 12 partial sketches (“doodles,” really) and asks subjects to use them in creating something new that no one will think of Ref. [47]. They encourage the use of colors and shading. Fluency points are awarded for the number of categorical transformations; originality points for whether sketches are inside, outside, or on both sides of the given boxes. Points are also awarded for creative titles to sketches. Elaboration points are awarded for symmetry, shading, and colors.

Perhaps the best known creativity test is the Guilford’s alternate uses task which requires subjects to generate as many possible uses as they can think of for a common household item (brick, paperclip, newspaper, etc.). Originality scores are based on normative frequency data. Responses that were given by only 5% of a group are regarded as unusual and awarded 1 point. Responses that are given by only 1% of the group are regarded as unique and given 2 points. Fluency is scored by the total number of responses for each individual. Flexibility is scored by the total number of different categories that are represented across all of a subject’s individual answers. The author notes that as fluency scores go up, so do originality scores. This is an identified contamination problem and can be corrected by dividing the originality by the fluency. Wallach and Kogan also include an alternative use task [45] as well as verbal associations and figural pattern making.

From our survey, we concluded that existing creativity tests require no technical or particular domain expertise. The measures used are fluency, flexibility, and originality. Transformational and analogical skills are not explicitly evaluated by these tests, while we have included items on our DT tests specifically for that purpose. All creativity tests listed above are nongoal oriented, i.e., there is no stated problem for which ideas are being sought. In contrast, design problems have explicit and implicit goals. There are also constraints in any real design problem, while none of the creativity tests try to limit the search space in any way. Another undesirable characteristic of creativity tests is that some use pictures and figures and even grade imagery and visualization. We need to remove the overlap between DT and VT so that each can

be independently evaluated. That does not imply that we should rid the DT test of all questions involving figures or pictures. Instead, we have achieved this by what we grade.

2.3 Derivation of Test Specifications. The current objective is to make the DT test suitable for undergraduate engineering majors at or above the sophomore level. No technical knowledge should be required beyond that level. Also, for practical reasons, it is best to aim for the test to take under 50 min to administer in order to allow it to be taken in one class period. Although this goal has not been achieved yet, it is expected that some test items can be dropped when strong positive correlations are discovered. The primary aim is to measure the four direct metrics and the secondary aim is to assess the indirect metrics. The latter requires exercises that are not ideation exercises but explicate these secondary effects, such as fixation. To go beyond generic creativity tests, an engineering orientation is to be achieved by incorporating goals and constraints in the exercises.

Since we have separate DT and VT tests, we want to minimize reliance on visual representations. This is not entirely possible, but we have attempted to do so—another major difference between our test and many creativity tests. On the other hand, we do not want to have questions that rely on one’s language skills and vocabulary. Finally, gender and ethnic bias needs to be avoided.

We must also consider how the test will be validated later. There are two distinct properties: reliability and validity. *Reliability* is the extent to which a test measures the true difference in individuals versus measurement errors. *Validity* is how well a test measures what it claims to measure [48]. Reliability is evaluated with two criteria: (1) stability—does the test give the same result for the same person each time (test–retest criteria) and (2) internal consistency—are the items on the test homogeneous (related to the same skill). Since test–retest is not practical in our case, we need to include multiple items on the same test to measure the same thing. Internal consistency, determined by Cronbach’s alpha measure [49], finds the correlation between the items on a test and also between the item and the total score. This measure is used to determine whether an item (question) should be included in the test or not.

2.4 DT Test Composition and Rationale. Based on the requirements from Sec. 2, we began developing the DT test as follows. We started with two general (nonengineering) problem types. The first one had no particular constraints or goals, just blue sky imagination, to assess fluency, variety, and originality. The second one was constrained to the use of given components; so in addition to the above measures, *quality* could also be assessed. Several candidate questions were considered for each type and made a part of an alpha test to determine their suitability.

To test design fixation, a simple design exercise was created but we also included one solution to that problem. The purpose

Category Name	Other Category Members
Colors	BLUE, RED, PURPLE, GREEN, YELLOW, PINK
Emotions	BLUE, HAPPY, ANGRY, BORED
Types of Whales	BLUE, SPERM, KILLER, HUMPBACK
Types of Berries	BLUEBERRIES, RASPBERRIES, STRAWBERRIES

Fig. 2 Part of an original exercise to test abstractability (since replaced)

was to determine the extent of fixation by looking at the similarities between the example given and the ideas generated by the individuals. The problem chosen was one for which previous fixation research had considerable amounts of data [27].

To test abstractability, we initially devised an exercise to measure the way that one perceives categories. For each test item, there was a keyword (e.g., blue in Fig. 2 example) which was to be used in as many different categories as possible. Early results from alpha trials revealed strong bias toward linguistic skills in this exercise, and it was eventually replaced. In fact, after the beta trials it became evident that the *ability to abstract* and the *tendency to abstract* were actually different dimensions. The latest version has separate questions to measure each. Another exercise for abstraction requires “discovery” of relations between groups of objects and to put those relations in a particular order.

To test the ability to handle complexity, an exercise was designed for synthesis of specified device categories (tools, toys, weapons, etc.) from a large number of given components. The objective is to use as many of the given components as possible in the synthesis of desired devices. This exercise was modeled after experiments conducted by Finke et al. [6].

To test originality, quality, and flexibility in a technical context, two design exercises were constructed. One involved the resolution of a conflict between two objectives. This particular exercise was also used to assess fixation by provided one example solution, so that we could have at least two different questions measuring fixation, one technical the other general. The second exercise was taken from a design contest conducted in a junior design class many years ago (current students have no knowledge of it).

Table 2 summarizes the questions and what they are designed to measure. From alpha trials, we determined appropriate time allocation for each exercise. As can be seen from Table 2, every subskill is measured at least by two or more questions in order to perform correlations necessary for validating stability and internal consistency of the DT test. Where there are more than two questions capable of measuring the same subskill we have the choice of not using all. What is actually measured on each question is discussed in Sec. 3.2.

2.5 Test Versions. Three major versions of the DT test have been created and used in the past 18 months. The alpha version was used for gauging the range of responses achievable, compare them to expectations, and solicit feedback from test participants. The primary goal was to determine the suitability of the questions,

clarity of the instructions, and time allocation. None of the data collected was scored and consequently not included in norming or reliability studies. Upward of 100 tests were given under the supervision of our own team.

The beta version was designed to collect large amounts of data for use in frequency analysis and categorization necessary for scoring originality and flexibility. Frequencies were also needed for use in normalizing the scores on a uniform scale (1–10). We invited the design academic community to participate in data collection by registering at our test portal [50]. A set of instructions were prepared for those administering the tests and they were asked to run them exactly as we would, so that all data sets would be consistent. Although upward of 500 beta tests were given, just over 300 were used in statistical analysis due to missing items or incompatibility of versions or other types of corruption.

Based on the experience with beta tests and data analysis the final version (gamma) has been prepared.

3 Data Collection and Scoring

The beta test responses were used to look at the following:

- (1) Number of responses, mean, and standard deviation: for use in fluency scoring
- (2) Categorization of responses and frequencies of categories: for use in variety (flexibility) scoring
- (3) Count of features and principles and their respective frequencies: in determining originality scores

It was also necessary to normalize the scores in order to aggregate each metric that was measured on multiple questions. We chose a scale of 1–10, with ten being the best.

In order to score the tests in a uniform and objective way, we have drafted a set of instructions for graders that include category labels, frequencies, and associated scores for every question. Three different people have been involved in scoring the tests. We have cross-checked their scoring against each other to remove biases, inconsistencies, and ensure uniform interpretation of scoring guidelines. In Subsections 3.1 and 3.2, we discuss our scoring methods and rationale.

Cross-scoring between institutions and evaluators can also determine consistency. We have currently eight institutions that have conducted beta tests and half a dozen more have signed up to participate. Beta trials data have been collected primarily from undergraduate engineering students taking design courses at Texas A&M University, Georgia Tech, BYU, Arizona State University, and Monash and Melbourne Universities in Australia. Industry participants have been drawn from Advatech Pacific (a design firm for hire), two different design groups at Intel (equipment design, assembly test development) and HP San Diego.

3.1 Norming Studies. Norming of test scores involves compiling the distribution of test scores in a target population. Norms must be based on large samples (Rose recommends a

Table 2 DT test composition and respective metrics capabilities

Q	Content	Flu	Flx	Org	Qlty	Abst	Afx	Dcmp	Dtl
1	Imagination exercise involving alternative universe; nontechnical	x	x	x					
2	Alternative uses of a common artifact; constrained; nontechnical	x	x	x	x				
3	Example exposure to test design fixation						x		
4	Synthesizing devices from given elements			x	x			x	x
5	Finding unusual semantic relations subject to specified criteria	x	x	x		x			
6	Exercises designed to determine the ability and tendency to abstract					x			
7	Technical conflict resolution problem	x	x	x	x		x		
8	Engineering design problem typical of undergraduate design contests; requires generation of concepts only			x	x			x	x

Table 3 Norming data for fluency scores for Q1

No. of ideas generated	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19+
Normalized score	1	2	3	4	5	6	7	8	9	10

Table 4 Norming Q1 originality scores based on frequencies

Category	No. of responses	Frequencies (%)	Score
1	113	9.21	6.0
2	36	2.93	9.0
3	250	20.37	1.0
4	179	14.6	3.5
5	127	10.35	5.5
6	153	12.47	4.5
7	63	5.13	8.0
8	170	13.85	4.0
9	127	10.35	5.5
10	9	0.73	10

minimum of 400 [51]). We have used 300 samples from beta tests, associating response categories with scores, for use in norming.

For Q1, we found that the average number of ideas generated was between nine and ten, while the max was upward of 20. Based on these numbers, the fluency score was scaled (Table 3).

The same question is also used for originality and variety scoring. In the first pass, we recorded all unique responses found. Then, the responses were categorized into ten groups for convenience, and finally the frequencies of each group were extracted from the data set. Detailed description of each category is given in the scoring instructions so graders do not misinterpret. (Only the category numbers are listed in Table 4 without the any descriptions.) The frequency data were used to determine the score based on the presumption that the more rare the occurrence of an idea category, the greater its originality. The formula used for originality score S for category i is

$$S_i = 9 * \left(\frac{\%H - \%C_i}{\%H - \%L} \right) + 1$$

where $\%H$ is the highest frequency, $\%L$ is the lowest, and $\%C_i$ is the frequency for category i [38]. From Table 4, we see that $\%H$ is 20.37 and $\%L$ is 0.73. So, for example, category 1 which has a frequency of 9.21% is found as

$$S_i = 9 * \left(\frac{20.37 - 9.21}{20.37 - 0.73} \right) + 1 \approx 6.0$$

From the same categorization, we can also determine the flexibility score by counting the total number of categories that all ideas fall into, by constituting a measure of total conceptual distance between ideas, or how well design space is explored.

3.2 Scoring Methods. Figure 3 shows the portion of the DT scoring sheet used to score all measures applicable to Q1 (flexibility, fluency, and originality). Note also that two different originality scores are computed: the average of all ideas (“average Originality”) and that of the best idea (called “max originality”). The reasoning is that it is not just the average idea one is interested in, it is that one great idea that one seeks for. Fluency, originality, and flexibility scores are evaluated the same way for other questions where they are scored, so we will not provide the specifics of every question here.

Afixability on Q3 where one example solution is provided is measured by similarities between the provided solution and

responses. As shown in Fig. 4, physical attribute similarities are less serious than functional (design principle) and thus weighted more heavily. The similarity points are subtracted from 10 in keeping with our 10 scale.

As stated above, quality is measured by goodness of fit with design goals, technical feasibility, and manufacturability. This definition implies that quality is context dependent. The DT test is the measuring quality on Q7 and Q8, both technically oriented problems. However, we have taken different approaches for evaluating quality in the two cases. Q7 responses are categorized in a manner similar to the procedure for Q1 explained in Sec. 3.1. The quality of each has already been predetermined by our team based on design specification and feasibility of that problem. Therefore, to score quality on this test one just needs to categorize the answers and refer to a table of predetermined values. The same table also contains originality scores (from frequency data collected) and afixability (from conceptual distance between the responses and the given solution). An excerpt of this is shown in Table 5, which shows how each metric is independently scored. This method works as long as responses can be categorized into one of the enumerated ones. This is the case for over 99% of the responses we are seeing. When categorization fails, a new category needs to be defined and its quality, afixability scores need to be established. Being so unusual, the originality score will be 10 for such new found categories. For Q8, a different approach is needed as there are several required design specifications and constraints of varying importance. A method similar to weighted objective trees [52] is used.

Detailability is measured by the extent of elaboration and clarity of expression. Different context based checklists have been created for Q4 and Q8, one of which is shown in Fig. 5.

The original problem for testing abstraction ability involved drawing trees of superclass and subclass of a given object. Test subjects found the exercise confusing and it was hard to grade. We experimented by increasing the structure (giving a template to fill in) and also by giving no structure. Neither approach improved the results, and the exercise was abandoned and replaced by a pair of exercises, one to test the ability to abstract and the other the tendency to abstract. The first correlates with the old exercise, but the second one has no equivalence to previous versions of the test. Therefore, the latter had to be treated as a separate variable in the analysis, as will be discussed in Sec. 4. In the tendency test we do not specify which way one can go; they can go up (generalize) or go down (specialize). Only the number of generalizations is counted to compute this score. In the ability exercise, one is asked to go up only and all responses meeting this criterion are counted.

Decomplexability has proven to be the hardest metric to design questions for and to evaluate. The barriers are time constraints on the length of the test and the desire to avoid specific domain knowledge. The best we have been able to do is to come up with a synthesis exercise from a given set of mechanical and structural elements. (We have already heard complaints that this question is biased toward mechanical engineers, although the elements are fairly common types of elements that everyone sees in everyday use.) This question is being evaluated in somewhat of a superficial manner. We count the number of elements used together on a device and their coupling.

4 Test Analysis

4.1 Factor Analysis. The goal of the factor analyses was to determine the number of distinguishable dimensions that underlie the set of measures and to determine which measures are related to which dimensions. Multiple dimensions were hypothesized initially, as these multiple dimensions had inspired the creation of test items as noted above. These hypothesized dimensions may not be accurate representations, however. For this reason, we used exploratory factor analysis (EFA) to explore the factor structure and to arrive at a factor solution that was plausible. As noted

Need to score three sub-skills: fluency, flexibility, & originality

FLUENCY: # of total valid responses given scored on an adjusted scale (based on normative data)

1. Place a check mark into the appropriate scaled score box according to total number of responses.

SCALED SCORE

	1	2	3	4	5	6	7	8	9	10
Fluency Score		X	X			X			X	

of responses listed

	1-2	3-4	5-6	7-8	9-10	11-12	13-14	15-16	17-18	19+
		X								

Fluency
2

FLEXIBILITY: # of total different identified categories responses fall into

2. Place a tally mark into the appropriate category each generated response falls into. Do this for each response that is listed. (Refer to attached scoring sheet for assistance with category and response identification)

Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Cat. 8	Cat. 9	Cat. 10
	X	X			X			X	

3. To determine the flexibility score, make a count of the total number of categories that have tally marks in them. # of tally marks does not matter for this scoring item, only categorical representation.

Flexibility
4

ORIGINALITY: Statistical Infrequency of responses

4. Using the information from above as a guide, re-write (numerically) how many total responses were included in each category below. From there multiply these numbers by the listed category weights and write the resulting product in the oval below each category:

# of items per cat.	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7	Cat. 8	Cat. 9	Cat. 10
	x	x	x	x	x	x	x	x	x	x
Originality Weights	6.0	9.0	1.0	3.5	5.5	4.5	8.0	4.0	5.5	10.0
	=	=	=	=	=	=	=	=	=	=

Ovals: 9, 1, 4.5, 5.5

TOTAL ORIGINALITY RAW SCORE
20

4a. **Mean Originality:** Divide the Total Raw score above by total #response
 $20/4 = 5.0$

4b. **Max originality:** Write the highest originality weight for the categories in which there is a response
9

Mean Originality
5.0

Max. Orig
9

Fig. 3 Scan of scoring sheet for Q1 measures

Count similar features to given example

attribute 1	attribute 2	attribute 3	attribute 4	Ttl similar attributes	functional principle	Ttl similar principles
				x 1.5		x 4.0
				=	+	=

Determine Afixability from $[10 - 1.5*(\text{no. attributes}) - 4*(\text{no. funct principles})]$

Afixability

Fig. 4 Comparison of similarities between fixation exemplar and responses

Table 5 Predetermined category scores for Q7

Category	Subcategory	n	Afixability		Quality		Novelty	
			i	n × i	j	n × j	k	n × k
A	A.1	3	7	21	5	15	3.9	11.7
	A.2	0	2	0	3	0	4.3	0
B	B.1	2	4	8	1	2	8.6	17.2

Elaboration criterion	Yes	No
1. understandable operation/function?		
- thrust		
- motion		
- medium		
- guidance/control of travel direction		
- launch method		
2. Construction understandable?		
3. Dimensions shown?		
4. multiple views, cross-sections, detailed/aux views?		
5. rationale given (goals, constraints, strategy)?		
6. materials elaborated?		
Total		

for every YES answer, 1 point is awarded (max= 10 pts)

Fig. 5 Evaluating detailability from responses

above, some study participants were not given the full set of items, leading to missing data. The EFA software [53] did not offer many options for handling missing data. To appropriately handle the missing data, we used the Mplus [54] program and conducted confirmatory factor analysis (CFA). The CFA was used to evaluate the factor structure reached in the EFA, and to do so while handling the missing data in an appropriate way. Mplus uses a full-information maximum likelihood (FIML) procedure for parameter estimation. The FIML approach uses all of the available information from each individual case in the data. No cases are dropped from the analyses for partially missing data. Missing data are assumed to be missing at random, which means that “missingness” is unrelated to the value, the measure would have had, after conditioning on the nonmissing information in the data [55].

EFA was performed first using a pairwise deletion strategy for handling missing data. Two different estimation methods were used: principle axis factoring and maximum likelihood. The two methods did not give substantially different results, and so maximum likelihood was used for the final analyses. Factor solutions from one to five factors were obtained with oblique rotations for all multiple-factor solutions. The oblique rotations permit nonzero correlations among the factors. Of the original 23 measures, two measures were immediately dropped. The originality score for Q2 was correlated at 0.99 with the fluency score for Q2. We dropped the originality score. Also, a new abstraction score for Q6 was obtained from only 111 participants, leading to problems of low sample size under pairwise deletion. We, therefore, dropped this variable from the EFA. The full variety of factor solutions was used for the remaining 21 variables. It became clear that 5 of the 21 measures did not load meaningfully on any factors in any of the factor solutions. These measures were Q3 affix, Q4 detail, Q4 demp, Q5 abst, and Q6 abst. All of these are indirect measures and of secondary importance as discussed in Sec. 2.1. The remaining 16 measures were confined to Q1, Q2, Q7, and Q8. A four-factor solution was found to be most interpretable solution for these measures, after oblique rotation. The scree plot for the eigenvalues of the correlation matrix for the set of 21 variables is given in Fig. 6. The plot indicates that the four-factor solution is plausible.

In the next step, the 16 measures were analyzed using Mplus in a CFA. The four-factor solution reached using EFA was directly specified in the CFA, with restrictions on the pattern of factor

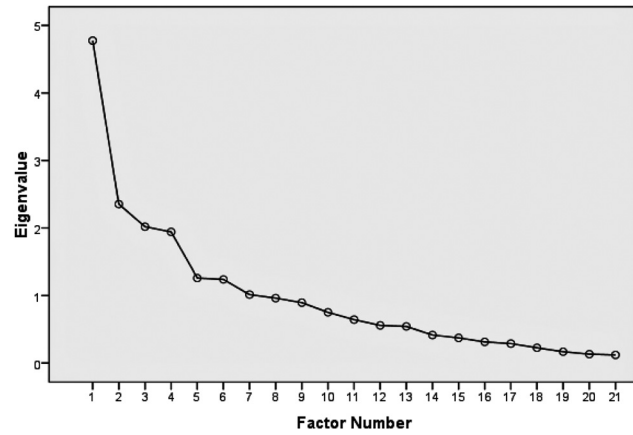


Fig. 6 Scree plot

Table 6 Factor loading for four-factor model

	F1	F2	F3	F4
Q1 fluency	0.89			
Q1 flexibility	0.77			
Q1 average originality	0.35			
Q1 maximum originality	0.38			
Q2 fluency	0.5	0.52		
Q2 flexibility		0.87		
Q2 average originality		0.58		
Q2 maximum originality		0.65		
Q2 flexibility	0.26	0.71		
Q2 average originality		0.64		
Q7 afixability			0.94	
Q7 quality			0.7	
Q7 originality			0.92	
Q8 originality				0.47
Q8 detailability				0.58
Q8 quality				0.84

loadings to force each measure to load one factor only. These restrictions set the CFA apart from the EFA solution, as EFA solutions permit unrestricted loadings. In addition, the CFA effectively used a larger sample due to the FIML adjustments for missing data. The four factors were permitted to correlate without restriction. The resulting CFA solution did not fit well when evaluated using stringent criteria for model fit. The model was rejected using the chi-square test of exact fit (chi-square = 563.61, df = 98, $p < 0.001$). The approximate fit indices were not adequate either (standardized root mean square residual = 0.094). Local fit indices indicated that some modifications to the specified model would improve the fit. Several modifications were adopted. First, two pairs of measures were permitted to have correlated unique factors (Q1 average with Q1 max; Q2 average with Q2 max). Second, two measures were permitted to load on more than one factor (Q2 pflux on both factors 1 and 2; Q2A fluency on both factors 1 and 2). These modifications yielded a four-factor model that showed improved fit. Although the test of exact fit again would reject the model (chi-square = 304.93, df = 94, $p < 0.001$), the approximate fit indices were improved (root mean square residual = 0.063). Estimates of the standardized loadings from the modified four-factor solution are given in Table 6. Note that two measures were permitted to load on more than one factor.

4.2 Correlations. The scale was set up to go from 1 to a max of 10 for each measure with the mean around 5. We have confirmed this from 300 statistically analyzed samples, as shown in Table 7. Note that in Q2 there were two different bases for categorization: one based on device action and the other based on

Table 7 Overall distribution of scores for each question and metric

Q#	Q1				Q2 action				Q2 app		Q3		Q4		Q5		Q6		Q7		Q8	
	Flu	Flx	AvgOrig	MaxOrig	Flu	Flex	AvgOrig	MaxOrig	Flx 2	AvgOrig	MaxOrig	Afix	Dcmpx	Dtl	Abst	Abst	Afix	Qlty	Orig	Orig	Dtl	Qlty
Maximum	10.0	9.0	7.8	10.0	10.0	9.0	8.9	10.0	10.0	10.0	10.0	10.0	9.0	8.0	10.0	10.0	10.0	9.6	9.9	9.2	10.0	10.0
Minimum	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.1	1.0	1.0	1.0
Mean	4.7	5.1	4.3	7.1	4.8	4.8	4.4	8.5	5.6	5.3	4.8	6.2	4.3	4.3	4.5	5.8	4.4	4.6	5.7	4.4	6.3	5.4
Standard deviation	1.7	1.4	1.0	1.8	1.9	1.6	1.7	2.1	2.1	1.4	1.9	2.5	2.1	1.4	2.0	3.3	2.6	2.0	1.9	1.9	2.2	2.2

application. From this analysis, the scaling and normalization appear to be reasonable, except for max originality scores, which are not expected to conform, since it is not independently scaled.

Cronbach alpha values for the four factors were

F1 (q1flu, q1flx, q1ave, q1max, q2aftu) = 0.767

F2 (q2afix, q2aave, q2amax, q2pfix, q2pave) = 0.845

F3 (q7afx, q7qual, q7orig) = 0.879

F4 (q8orig, q8dtl, q8qual) = 0.655

F4 is a bit low but the rest are regarded as a good.

The averages for each metric from all questions that measure it were also computed, using equal weights. We then looked at the correlations between the pairs of the eight metrics (Table 8). Fluency and flexibility show a strong correlation (0.75).

Flexibility and originality also correlate well (0.57). However, fluency and originality have a weak correlation implying that emphasis on quantity of ideas does not necessarily yield original ideas. Quality had no correlation to fluency, flexibility, and originality. Wild and crazy ideas may be very original but not practical. Even this examination of the four direct measures (fluency, flexibility, originality, and quality) indicates that there are possibly at least four independent factors present, a confirmation of the findings of the factor analysis from Sec. 4.1.

The indirect measures show neither strong correlations with each other nor with any of the direct measures. As pointed out before, early results for abstractability motivated us to change the questions. One of fixation exercises did not have enough richness to be a good discriminate between fixated and nonfixated responses. As a result, a new question with greater number of features and more obvious solution principle has replaced the past question. The result of these changes is that abstractability and afixability from the new data sets cannot be compared directly to the old sets.

For the other two indirect measures, decomplexability and detailability, we think test time of less than 1 h is a problem. Decomplexability requires the test to have more complicated questions that would require more time to solve. Detailability can only be demonstrated if time is given to produce more elaborate responses. So, measuring these two subskills is still a challenge.

We did not include maximum originality in our correlation matrix, since it is extracted as the max originality score from an item for each respondent. We looked at the correlation between average originality and max originality and found it to be 0.586. This is fairly strong but may not be strong enough to warrant dropping one measure in favor of the other.

Finally, we looked at the correlation between the two categorizations for Q2 scoring. It turns out to be 0.68 for flexibility and 0.57 for originality. Again, this says to us that both categorizations should continue to be used in scoring.

5 DT Skill Profile

One can arrive at an overall score by aggregating all of the scores for an individual. However, that would assume equal weights for the measures. At this time, we do not fully understand proper weighting of each subskill to come up with an overall score. So, we present the results for each metric and compare the individual to either their peer group or to the entire population that has been tested. To help interpret these results, we also show best and worst scores in a diagram. Figure 7 shows the DT skill profile of two test takers A and B with respect to the best and worst for the test group.

6 Discussion

Several standardized tests of engineering design skills are being constructed. The divergent thinking test is in the most advanced stage. Data have been collected from large numbers of undergraduate engineering students, smaller numbers of graduate students, and practicing designers. From the results of several beta trials,

Table 8 Correlation matrix

	Flu	Flex	Orig	Qty	Dcmp	Dtl	Abst	Afix
Fluency	1.000	0.7544	0.2890	0.0125	0.0714	0.0117	0.2724	0.0668
Flexibility		1.0000	0.5749	-0.0129	0.1016	-0.0023	0.1831	0.0457
Originality			1.0000	-0.0636	0.1285	0.0382	0.0133	0.0222
Quality				1.0000	-0.0153	0.3312	-0.0073	-0.3798
Decomposability					1.0000	0.3594	0.1513	-0.0137
Detailability						1.0000	0.1246	-0.0341
Abstractability							1.0000	-0.0912
Afixability								1.0000

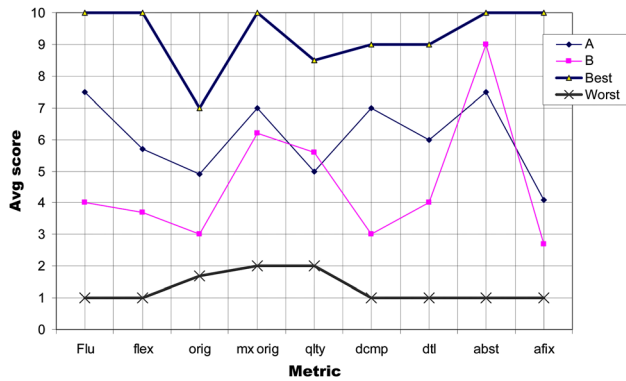


Fig. 7 DT skill profiles

the test has been continuously improved. Statistical analysis shows that the test is a reasonable instrument for testing engineering design oriented divergent thinking (ideation) skills. Only intrinsic validation has been done. Full validation of the tests will require collection of enormous amounts of data from a large number of participants using factorial variants of the tests. This will take several years and would require a community effort and buy-in. We have set up a web site with an open invitation to the design academic community to participate in the beta trials. Any design educator can request the tests, administer them to his/her own students, and return the tests to us for scoring. Figure 8 shows the range of results for all engineering juniors tested so far.

Criterion and construct validity studies are not within the current scope of work. In the future, we propose to determine criterion validity by predicting how one will do on a design task exercise which requires those particular skills. Again, the purpose is to establish preliminary association of the tests with particular skills. Construct validity can only be determined by comparing

the results of a test with other tests that claim to measure the same thing. In some areas, such as lateral thinking, spatial reasoning, there are tests available that we can directly compare to, or at least relate subsets of our items to those sets.

To encourage the development of design skills we must reward out-of-the-box thinking, risk taking, unconventional, and unusual ideas. Factors that influence student attitude include course format, content, problem types used in homework, laboratories, projects, exams, and the evaluation/grading system. The conventional system is “assignments centric;” grades are computed from the weighted sum of homework, exams, and other assignments. The only score that is typically recorded is the aggregate score for each assignment. This single score hides the strengths and weaknesses of an individual. Even if the exercises given were designed specifically to teach/evaluate certain design skills, recording a single score is not adequate. Based on the methods from this study, a new skill based learning and grading system could be implemented with three main elements involved: explication of design skills; association of skills subsets for each design exercise; record keeping and aggregation of scores organized by skills. Each class exercise or assignment could be designed with the objective of teaching, practicing, assessing a particular subset of skills, and students told in advance of the particular skill(s) that are to be graded on each exercise. This skill evaluation may have potential uses in (1) determination of design strengths/weaknesses of individuals for the purpose of corrective action; (2) matching individuals with complementary strengths on design teams; (3) continuous improvement and evaluation of the course content.

This research will connect design research to established cognitive theories of human problem solving and learning, visual and spatial reasoning, pattern recognition, and scientific discovery. It seeks to gain insights into how design knowledge is used and what differentiates good designers from the ones less skilled. This research is a necessary prerequisite for creating a framework for future experiments related to design skills, collection of extensive data to enable establishment of norms for skills, new grading

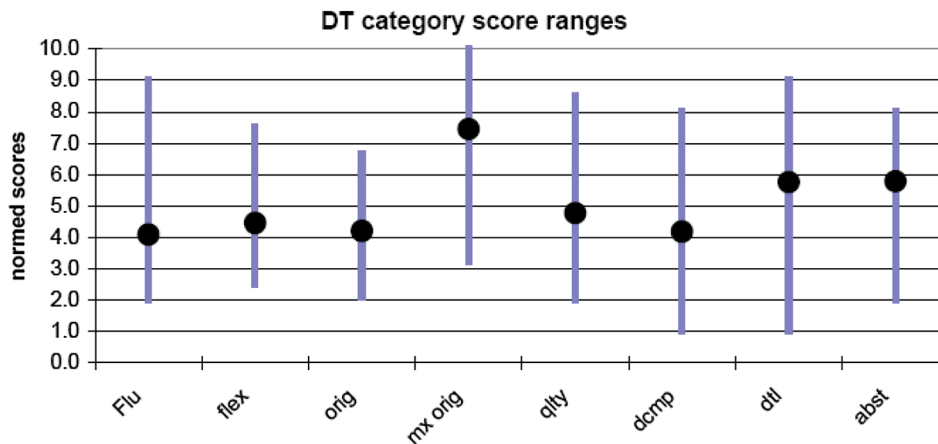


Fig. 8 Comparison of one group of students to its peer group

methods for design classes and curriculum evaluation, and more sophisticated bases for design project team formation.

Acknowledgment

We wish to thank the following individuals for administering DT tests at their institutions/organizations: Professor Bruce Fields, Monash University, Australia; Professor Dirk Schaefer, Georgia Tech; Professor Robert Todd, BYU; Professor Chris Mattson, BYU; Professor Rodney Hill, Texas A&M; Mr. Andy Contes, Intel ATD; Mr. Frank Heydrich, ADVATECH. This research is supported by the US National Science Foundation Grant No. CMMI-0728192. Opinions expressed in this paper are those of the authors and not endorsed by NSF.

This research was originally presented at the ASME Design Theory & Methodology Conference, Montreal, August 2010.

References

- [1] Shah, J., 2005, "Identification, Measurement and Development of Design Skills in Engineering Education," International Conference on Engineering Design (ICED05), Melbourne, Australia.
- [2] Shah, J., Smith, S. M., and Woodward, J., 2009, "Development of Standardized Tests for Design Skills," International Conference on Engineering Design (ICED09), Stanford, CA.
- [3] Dorst, K., and Cross, N., 2001, "Creativity in Design Process: Co-Evolution of Problem-Solution," *Des. Stud.*, **22**, pp. 425–437.
- [4] Maher, M., 1996, "Modeling Design Exploration As Co-Evolution," *Microcomput. Civ. Eng.*, **11**(3), pp. 195–209.
- [5] Torrance, E., 1964, *Role of Evaluation in Creative Thinking*, Bureau of Educational Research, University of Minnesota, Minneapolis.
- [6] Finke, R. A., Ward, T. B., and Smith, S. M., 1992, *Creative Cognition: Theory, Research, and Applications*, MIT Press, Cambridge, MA.
- [7] Campbell, D. T., 1960, "Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes," *Psychol. Rev.*, **67**, pp. 380–400.
- [8] Simonton, D. K., 1999, *Origins of Genius: Darwinian Perspectives on Creativity*, Oxford University Press, New York.
- [9] Simonton, D. K., 2007, "Picasso's Guernica Creativity as a Darwinian Process: Definitions, Clarifications, Misconceptions, and Applications," *Creativity Res. J.*, **19**(4), pp. 381–394.
- [10] Guilford, J., 1967, *The Nature of Human Intelligence*, McGraw-Hill, New York.
- [11] Dunbar, K., 1995, "How Scientists Really Reason: Scientific Reasoning in Real-World Laboratories," *The Nature of Insight*, R. J. Sternberg and J. E. Davidson, eds., MIT Press, Cambridge, MA, pp. 365–395.
- [12] Dunbar, K., 1997, "How Scientists Think: On-Line Creativity and Conceptual Change in Science," *Creative Thought: An Investigation of Conceptual Structures and Processes*, T. B. Ward, S. M. Smith, and J. Vaid, eds., American Psychological Association, Washington, DC, pp. 461–493.
- [13] Mednick, S. A., 1962, "The Associative Basis of the Creative Process," *Psychol. Rev.*, **69**, pp. 220–232.
- [14] Ward, T. B., 1998, "Analogical Distance and Purpose in Creative Thought: Mental Leaps Versus Mental Hops," *Advances in Analogy Research: Integration of Theory and Data From the Cognitive, Computational, and Neural Sciences*, K. Holyoak, D. Gentner, and B. Kokinov, eds., New Bulgarian University, Sofia.
- [15] Wharton, C. M., Holyoak, K. J., and Lange, T. E., 1996, "Remote Analogical Reminding," *Mem. Cognit.*, **24**, pp. 629–643.
- [16] Metcalfe, J., 1986, "Premonitions of Insight Predict Impending Error," *J. Exp. Psychol. Learn., Mem. Cogn.*, **12**, pp. 623–634.
- [17] Metcalfe, J., and Weibe, D., 1987, "Intuition in Insight and Non-Insight Problem Solving," *Mem. Cognit.*, **15**, pp. 238–246.
- [18] Smith, S. M., 1994, "Getting Into and Out of Mental Ruts: A Theory of Fixation, Incubation, and Insight," *The Nature of Insight*, R. Sternberg, and J. Davidson, eds., MIT Press, Cambridge, MA, pp. 121–149.
- [19] Smith, S. M., 1995, "Fixation, Incubation, and Insight in Memory, Problem Solving, and Creativity," *The Creative Cognition Approach*, S. M. Smith, T. B. Ward, and R. A. Finke, eds., MIT Press, Cambridge, pp. 135–155.
- [20] Baughman, W. A., and Mumford, M. D., 1995, "Process-Analytic Models of Creative Capacities: Operations Influencing the Combination-and-Reorganization Process," *Creativity Res. J.*, **8**, pp. 37–62.
- [21] Mumford, M. D., Reiter-Palmon, R., and Redmond, M. R., 1994, "Problem Construction and Cognition: Applying Problem Representations in Ill-Defined Problems," *Problem Finding, Problem Solving, and Creativity*, M. A. Runco, ed., Albex, Norwood, NJ, pp. 3–39.
- [22] Ward, T. B., 1994, "Structured Imagination: The Role of Category Structure in Exemplar Generation," *Cognit. Psychol.*, **27**, pp. 1–40.
- [23] Ward, T. B., Dodds, R. A., Saunders, K. N., and Sifonis, C. M., 2000, "Attribute Centrality and Imaginative Thought," *Mem. Cognit.*, **28**, pp. 1387–1397.
- [24] Ward, T. B., Patterson, M. J., Sifonis, C. M., Dodds, R. A., and Saunders, K. N., 2002, "The Role of Graded Category Structure in Imaginative Thought," *Mem. Cognit.*, **30**, pp. 199–216.
- [25] Ward, T. B., Patterson, M. J., and Sifonis, C., 2004, "The Role of Specificity and Abstraction in Creative Idea Generation," *Creativity Res. J.*, **16**, pp. 1–9.
- [26] Christensen, B. T., and Schunn, C. D., 2007, "The Relationship of Analogical Distance to Analogical Function and Preinventive Structure: The Case of Engineering Design," *Mem. Cognit.*, **35**, pp. 29–38.
- [27] Jansson, D. G., and Smith, S. M., 1991, "Design Fixation," *Des. Stud.*, **12**(1), pp. 3–11.
- [28] Chrysikou, E. G., and Weisberg, R. W., 2005, "Following the Wrong Footsteps: Fixation Effects of Pictorial Examples in a Design Problem-Solving Task," *J. Exp. Psychol. Learn. Mem. Cogn.*, **31**(5), pp. 1134–1148.
- [29] Dahl, D. W., and Moreau, P., 2002, "The Influence and Value of Analogical Thinking During New Product Ideation," *J. Mark. Res.*, **39**, pp. 47–60.
- [30] Linsey, J., Tseng, I., Fu, K., Cagan, J., and Wood, K., 2009, "Reducing and Perceiving Design Fixation," International Conference on Engineering Design, Stanford, CA.
- [31] Purcell, A. T., and Gero, J. S., 1996, "Design and Other Types of Fixation," *Des. Stud.*, **17**(4), pp. 363–383.
- [32] Smith, S. M., and Blankenship, S. E., 1989, "Incubation Effects," *Bull. Psychol. Soc.*, **27**, pp. 311–314.
- [33] Shah, J. J., Smith, S. M., Vargas-Hernandez, N., Gerkens, R., and Wulan, M., 2003, "Empirical Studies of Design Ideation: Alignment of Design Experiments With Lab Experiments," Proceedings of the American Society for Mechanical Engineering (ASME) DTM Conference, Chicago.
- [34] Vargas-Hernandez, N., Shah, J., and Smith, S. M., 2007, "Cognitive Models of Design Ideation," Proceedings of the International Design Engineering Technical Conference/Computers and Information in Engineering.
- [35] Jansson, D. G., Condoor, S. S., and Brock, H. R., 1993, "Cognition in Design: Viewing the Hidden Side of the Design Process," *Environ. Plan. B: Plan. Des.*, **19**, pp. 257–271.
- [36] Condoor, S. S., and Burger, C. P., 1998, "Coupling and Its Impact on the Product Creation Process," *Management of Technology, Sustainable Development and Eco-Efficiency*, L. A. Lefebvre, R. M. Mason, and T. Khalil, eds., Elsevier, Amsterdam, pp. 197–206.
- [37] Gardner, H., 2006, *Multiple Intelligences*, Basic Books, New York.
- [38] Shah, J. J., Smith, S. M., and Vargas-Hernandez, N., 2003, "Metrics for Measuring Ideation Effectiveness," *Des. Stud.*, **24**(2), pp. 111–134.
- [39] Woodward, J., and Shah, J., 2008, "Analysis of Divergent Thinking Tests," Technical Report DAL2008-05, Design Automation Lab, Arizona State University.
- [40] Goff, K., and Torrance, E. P., 2002, *Abbreviated Torrance Test for Adults*, Scholastic Testing Services, Inc., Bensenville, IL.
- [41] Meeker, M., and Meeker, R., 1982, *Structure of Intellect Learning Abilities Test: Evaluation, Leadership, and Creative Thinking*, SOI Institute, El Segundo, CA (revised in 2000).
- [42] Meeker SOI checklist, <http://www.soisystems.com/>
- [43] Torrance, E. P., 1974, *Torrance Tests of Creative Thinking*, Personnel Press, Lexington, MA.
- [44] Guilford, J. P., 1950, "Creativity," *Am. Psychol.*, **5**, pp. 444–454.
- [45] Wallach, M. A., and Kogan, N., 1965, *Modes of Thinking in Young Children: A Study of the Creativity-Intelligence Distinction*, Holt Rinehart & Winston, New York.
- [46] Guilford, J. P., Wilson, R. C., and Christensen, E. C., 1952, "A Factor-Analytic Study of Creative Thinking II. Administration of Tests and Analysis of Results," Report No. 8, Psychological Laboratory, University of Southern California, LA.
- [47] Williams, F., 1980, *Creativity Assessment Packet: Examiner's Manual*, Pro-Ed Publishing, Austin, TX.
- [48] Kline, P., 1993, *The Handbook of Psychological Testing*, Routledge, London.
- [49] Cronbach, L. J., 1990, *Essentials of Psychological Testing*, 5th ed., Harper & Row, New York.
- [50] <http://asudesign.eas.asu.edu/testportal/index.php>
- [51] Rose, R. G., 1993, *Practical Issues in Employment Testing*, Psych Assmntt, Inc., Odessa, FL.
- [52] Pahl, G., and Beitz, W., 1995, *Engineering Design*, 2nd ed., Springer, London.
- [53] SPSS, Inc., 2008, *SPSS Statistics 17.0*, SPSS, Inc., Chicago, IL.
- [54] Muthén, L., and Muthén, B., 1998–2006, *Mplus User's Guide*, 4th ed., Muthén & Muthén, Los Angeles, CA.
- [55] Little, R. J. A., and Rubin, D. B., 1987, *Statistical Analysis With Missing Data*, John Wiley & Sons, New York.