INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN, ICED'09
24 - 27 AUGUST 2009, STANFORD UNIVERSITY, STANFORD, CA, USA

DEVELOPMENT OF STANDARDIZED TESTS FOR DESIGN SKILLS

Jami J. Shah, S. M. Smith[2] and Jay Woodward[3]
(1) Mechanical & Aerospace Eng, Arizona State, (2) Dept. of Psychology, Texas A&M Univ.
(3) Dept. of Ed Psych, Texas A&M University

ABSTRACT
A number of cognitive skills relevant to design have been identified. Formal methods for evaluating these skills in engineering students are needed. Towards that goal we are developing a battery of standardized tests for the most critical of these skills. These tests are based on research in design thinking, human problem solving, cognitive psychology and psychometrics. Tests for Divergent Thinking, Visual Thinking, Spatial Reasoning, Qualitative Reasoning and Problem Formulation are in various stages of development. This paper describes the procedure and rationale for test development. It involves characterization of design skills in terms of measurable indicators; devising skill measurement instruments (problems, metrics, scales); conducting trials to collect data and evaluation of tests for criterion and construct validity. Beta trials are in progress involving undergraduate and graduate students at multiple universities and engineers in industry. Possible applications of these tests include evaluating students in design classes, forming of balanced design teams that posses skills necessary for a given project and evaluating the effectiveness of design courses and curricula.

Keywords: Design thinking; skill evaluation; design education; teamology; Psychometrics

1 INTRODUCTION
Product design is a complex activity that involves the use of human intellect, knowledge and skills. Design tools and methods impact the design process, but the choice of these and how they are used may also depend on the designer(s) involved. Economic factors and organizational environment also have influence, but different designers may respond to such factors in different ways. The point that is being made is that the human designer is the most important element in product development. It is, therefore, important that we attempt to formally characterize designers in terms of key skills relevant to design. A good designer or design team must possess a wide range of skills to tackle various phases of product development, from problem formulation to detailed design. Different problem types (novelty *vs.* routine) may require different skill sets. There are a wide variety of design problems and so one may expect different combination of skills (skill sets) to be required for different problems.

One possible benefit of standardized tests is for designing design curriculums, design exercises and projects, and evaluating their effectiveness by testing students before and after such experiences. Another possible use is in composing design teams which collectively posses all the skills needed for a particular project. The US National Academy report on engineering design specifically mentions the need for developing metrics for evaluating designs and metrics for the effectiveness of the design component of the curriculum [1].

Although design skills are indirectly alluded to in design textbooks and curricula, there has not been a concerted effort to explicitly identify and measure them. In our ICED 2005 paper we proposed a set of design skills and objective characterization of the skills. We discussed how such characterization forms the basis of how we currently grade our students in design exercises and projects in our department. Under an NSF grant we are now developing standardized tests for a sub-set of these skills. The objective of this research is to develop these skill definitions more formally, create metrics and design problems, refine the problems and scoring criteria through a validation process, and to correlate them to other domain independent standard reasoning tests. Such an effort must have input from the larger design education community, the eventual end users of standardized skill tests. We are inviting the ICED community to participate in beta testing of the measurement instruments we have developed.

To skeptics of standardized tests of design skills, we point out that Psychologists have been testing reasoning skills for almost a century: William Stern's IQ test and its derivatives [2,3] which include questions on verbal, numerical and abstract reasoning, Torrance and other creativity tests [4], Watson-Glaser critical thinking tests[5,6], Purdue spatial reasoning tests [7], ACER mechanical reasoning tests [8], visual [9] and verbal understanding tests. Many aspects of these tests are directly or indirectly related to design tasks. For example, creativity tests evaluate divergent thinking and imagination, which are strongly connected to idea generation in conceptual design; critical thinking is related to design evaluation and analysis.

2. SURVEY OF RELATED STUDIES

A wide range of literature is relevant to this study. As design is a type of problem solving, we need to look at cognitive studies of scientific discovery, learning, Human Problem Solving (HPS), spatial reasoning and creativity. We also need to survey measurement techniques related to reasoning skills and creativity. We examined past studies in design research that have identified specific thinking skills and evidence of their relevance to design synthesis. A brief synopsis is provided here organized along two streams: what do we know about reasoning or thinking skills and how are they being measured.

2.1 Thinking Skills

The expertise or skills used in scientific discovery have been under intense study by Psychologists for the past several decades [10-13]. Although design and scientific discovery are different, there are many generic skills that are common. Skills identified in Scientific Discovery literature include: comprehension, inferencing, planning, predicting, experimenting and evidence evaluation skill. These theories have been influential in designing science curricula [14]. There is controversy over whether skills transfer between domains. Theories of situated learning predict they do not, but there is now empirical evidence that there are both domain specific and domain independent skills [15]. Box points out some similarities between scientific discovery and engineering [16].

Human Problem Solving (HPS) is the area of cognitive science concerned with the nature of thinking at a macro level [13]. Problems have been classified in various ways based on structuredness, complexity, and domain specificity. There is a consensus that HPS is very different for well structured problems than for ill-defined problems. For the former, various types of schema based theories have been proposed [17,18]. Routine problem solving requires deductive reasoning. For ill structured problems, scripts do not exist. A plan needs to be devised. Initial conception of the problem is important because it determines the initial mental model. Analogical reasoning is needed to retrieve appropriate scripts.

"Design Thinking" is defined as cognitive processes manifested in design action [19]. Dym *et al* identified desired capabilities of designers in terms of: divergent and convergent thinking, systems thinking, knowing how to handle uncertainty and complexity, decision-making, social skills needed to operate in team, and ability to think in terms of several languages [20]. In an effort to elaborate these high level capabilities, several additional skills are mentioned: qualitative reasoning, pattern recognition, performing experiments, sketching, framing of design decisions and ability to make estimates. These skills are not characterized at a detailed enough level to allow measurement. Dym *et al* make the observation "*…new assessment techniques are needed because new skills are being taught*". This is exactly what we are addressing in this paper.

The most common skills mentioned in design literature are lateral or divergent thinking [21], visual thinking [22,23], Spatial reasoning [24], and Sketching [25,26]. Somewhat to a lesser extent there are studies on the use of analogical reasoning in design, though this is a subject that has been studied a great deal in other contexts. Many have speculated that there is a strong connection between visual thinking and creativity [22,23,27]. The relationship of visual thinking/imagery, provocative stimuli, and flexible representation with creative design has been established in various studies in creative cognition, visual imagery, and design thinking [29-32]. Akin studied the progression of spatial reasoning skills in architecture students and found contradictions between test results and observations by human judges [33]. Sorby and others have focused on developing exercises to enhance spatial abilities of engineering students [9]. Sketches and drawings are languages of engineering that are used not only to document and communicate ideas, but essential to understanding the problem [27], off-loading of information from short term memory [25], and development of ideas [26].

2.2 Cognitive Testing

Design skills are sub-sets of human problem solving skills, which forms part of human intellect. Several tests of intellectual development have been proposed, the most popular being based on Perry's scale [34] and King & Kitchner's RJ model [35]. Intellectual development measures the sophistication with which one handles complex problems or conversely, the maximum complexity of a problem that can be handled. Wilde quantified and normalized MBTI measures of Jung's model [36] for the purpose of forming well balanced design teams [37]. He identified eight roles based on this quantification on each of the two criteria (Perception and Judgment). He has used this classification in composing successful design teams for student competitions.

Divergent thinking and pattern recognition are tested on various creativity tests. The three most dominant tests are Guilford test [38], the Torrance Test (TTCT) [39] and the Remote Associates Test (RAT) [40]. The Guilford test asks subjects to think of different uses of some common object, such as a brick. Answers are graded on originality, fluency (quantity), flexibility (variety) and elaboration (detail). In RAT, images, objects, words are listed in order of weaker and weaker association. The subject needs to discover how they are related; the further down one can go, the higher the score. TTCT consists of a verbal test and a figural test. These tests are discussed further in Sec. 3.2.

Visual thinking and spatial reasoning are often not differentiated in many studies, although it is clear that spatial involves 3D objects while visual thinking may be broader, including 2 or 3D, real objects or iconic figures. Many kinds of visual/spatial tests have been developed, including paper folding, arranging blocks, paper-pencil exercises, film/computer animations, finding an efficient path, and verbal tests [7, 41-42]. Psychological factors that are tested include maintaining and transforming images in working memory [7,33], the importance given to mental models in reasoning [43], ability to fill-in missing detail [23], and spatial reasoning using technical drawings [24]. We discuss these further in Sec 3.3.

3. TEST DEVELOPMENT

The current scope of our tests is limited to skills related to early stages of engineering product design. Also, the primary emphasis of this research is on the evaluation of individual designers; social aspects of design, including group dynamics, are not considered at the present time.

The following procedure for test development is used:

Stage 1: Enumeration of design skills and their association with specific design tasks
Stage 2: Characterization of skills in terms of measurable and verifiable indicators
Stage 3: Development of skill measurement instruments (standard design problems, metrics, scales)
Stage 4: Collection of data from alpha and beta level trials
Stage 5: Test validation using standard Psychometrics methods

We define a skill as the cognitive ability to perform a task. Design skills have been derived from observations of design tasks, as well as, from past cognitive studies. A good designer or design team must possess a wide range of skills to tackle different phases of product development (problem definition, conceptual design, embodiment design, detailed design). The importance and duration of these phases varies with the design problem's novelty and complexity; consequently the skill set needed also varies. Routine design proceeds directly to embodiment and detailed design phases; novel design must start with problem formulation, conceptual design phases and experimentation.

From our past work and that of others, as presented in the previous section, 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, 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, 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. For details about skill characterization, categorization and rationale, we refer the reader to our previous ICED paper [44]. This paper presents test development only.

A team consisting of an engineer, a Cognitive Psychologist, an Educational Psychologist and a Psychometric consultant is developing tests, administering trials and analyzing the data for a subset of design skills: Divergent thinking (DT), Visual Thinking (VT) and Spatial Reasoning (ST). Future plans include tests for Problem Formulation (PF) and Qualitative Reasoning (QT).

3.1 General Principles & Objectives

When developing standardized tests that may be given to large numbers of people, it is important to keep in mind the effort needed for scoring the tests and the consistency of results scored by different people. It is also important to keep in mind the expertise and education level of the test takers. We are targeting engineering students at or above the Junior (2^{nd} year) level. Questions have to be chosen carefully to create a clear discrimination between individuals for any given skill indicator. This has to be done through a series of trials. From the point of view of Psychometric validation the Stability component of Reliability Measure it is important to introduce redundancy by putting multiple items that test the same indicator; this allows reliability determination by correlating the results of the redundant items. Once a positive correlation has been established between redundant items, the redundancy can be removed from future tests. Orhun observes that the concurrent application of different forms of knowledge and skills makes assessing difficult [45]. However, if we use two different types of problems, ones requiring general knowledge and ones requiring technical knowledge at least at the level of engineering students, we will be able to study skill with or without domain knowledge. We take domain independence to imply problems that do not require expert knowledge but may be solvable by a certain group, e.g., ME sophomores. For this reason the Tests contain both engineering oriented and non-engineering questions. Another consideration is that it should be possible to complete each Test in a reasonable amount of time before boredom or fatigue sets in. Ideally, the Tests should be no more than 1 hour so as to complete in a typical class period. If that's not possible, we should split the Test into multiple parts. Another factor we took into account is that any faculty should be able to administer the tests by simply following a set of instructions.

The success of the tests hinges strongly on clear and complete characterization of the skills in terms of indicators of that test. We approached this from three directions concurrently. One is to examine past empirical studies of design activities for any attributes or indicators associated with cognitive processes used in such activities. The second is to look at related tests. The third is to extract the characteristics from theories and models related to that cognitive task.

The following sections give a summary of our efforts so far in developing tests for Divergent Thinking, Visual Thinking and Spatial reasoning.

3.2 Divergent Thinking (DT)

The opportunity for creative design varies with the type of design problem, but *divergent thinking (DT)* is the relevant skill. 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. In creativity literature, fluency or number of ideas generated is identified as an indicator of creative ability. However, using only the number of ideas (quantity) generated as a measure of DT is inadequate because there could be many superficial variations of the same basic design. Therefore, a measure of variety is needed to determine how well 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 can imagine [46]. 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. 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* [46]. 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. Re-formulation of problems is facilitated by the ability to abstract or generalize. In some design problems the ability to synthesize different combinations of elements may lead to many new alternative designs. In fact, many formal design methods, such as Morphological Charts are based on making random connections.

Research shows that rather than provoking novel and creative ideas, outside stimuli can sometimes *obstruct* problem solving by suggesting wrong solutions, or inhibit creative idea generation by promoting constrained lines of thinking [47]. Experimental induction of impasses with stimuli related to experimental tasks has been demonstrated in memory, problem solving, creative idea generation, and engineering design [46,47]. Recently encountered ideas can block or constrain creativity, proof that implicit cognition can block creative idea generation. Memory experiments indicate that effects of examples are implicit, and cannot be voluntarily rejected to allow appropriate responses. The ability to avoid fixation is therefore another indicator of divergent thinking.

In Table 1 we have collected and classified DT indicators into three main types: General fundamental DT, transformational, and technical specific to engineering. The fundamental indicators are ideation fluency, versatility and originality. The ability to get out of a rut involves transformational skills: abstraction and fixation avoidance. For engineering design, the technical feasibility and fit with desired specification, ability to handle complexity with detailed elaboration are valuable.

Table 1: DT sub-skills

Fundamental skills	Transformational skills	Engineering specific
Ideation Fluency Ideation Versatility Ideation Originality	Abstraction-ability Fixation avoidance	Complexity handling Elaboration ability (thinking through) Goal directed (Functional quality/tech feasibility)

We examined eight standardized tests of Creativity to see the extent to which the above skill indicators are represented. They include: Abbreviated Torrance [48], Meeker test [49], Meeker SOI [50], classic Torrance [4], Guilford alternative uses [38], Wallach & Kogan test [51] and Guilford ARP [52]. No technical expertise is needed and no variety measures are used in these tests; fluency and originality are the measures used. Transformational 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 non-goal 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 overlap between DT and VT to the extent possible 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.

An alpha version of the DT test was developed and given to approximately 150 undergraduates. The students were also asked to provide feedback about the questions and clarity of the instructions. Based on the test answers and student feedback we dropped some questions, replaced others and improved the wording to get the beta version. The beta version, currently undergoing trials at a number of schools, consists of eight items, as listed in Table 2. It takes approximately 70 minutes to administer the test, including giving instructions to the test takers. We expect to reduce the number of items on the test by removing redundancies after correlation studies are completed. The actual questions on the test are not included in this paper but one can obtain copies by participating in beta testing. Instructions for that are posted on the Test website [53]. Table 2 gives a generalized summary of items on the DT Test and the intended purpose of each item.

Table 3 shows what sub-skills are being evaluated by each DT question and the metric used. Fluency is measured simply by the count of admissible ideas. Versatility is measured by the variety of ideas generated. In our past studies we have developed a variety metric based on the “genealogy” of ideas, or what may be termed the average “conceptual distance” [54]. The conceptual origins of ideas are analyzed through a genealogical categorization based on how ideas fulfill each design function. In this method, the range of ideas is differentiated by a progressive scale based on the number of different physical principles used, different working principles and embodiments. This method is appropriate for use with Q4,7,8 but for others a simpler categories and members database has been created from responses obtained in alpha and beta trials. To measure originality, we have collected data based on about 500 tests from beta trials so far on relative frequencies of each type of response and category.

The inverse of the frequency is used to measure novelty. The lower the count (i.e., the less a characteristic is found) the higher the novelty. Metrics similar to this have also been used by psychologists to measure creativity [3,47]. If a question requires several ideas to be generated, we can look at the average novelty or the best (strength of novelty); both are included in the score sheet. Use of the databases makes scoring easier and consistent across graders. Abstraction ability is measured by particular questions that require finding relations between apparently unrelated items and by constructing superclass hierarchies. Fixation is tested by looking at how close the responses are to the example given (conceptual distance). Elaboration and functional quality are measured subjectively.

Table 2: DT Test composition

DT Question	**Content**	**Time (min)**
Q1. *"What if"*	Imagination exercise involving alternative universe; non-technical	5
Q2. "*What for*"	Alternative uses of a common artifact; constrained; non-technical	5
Q3. "*What else*"	Example exposure to test design fixation	5
Q4. "*What can it be*"	Synthesizing devices from given elements	6
Q5. "*What next*"	Finding unusual semantic relations subject to specified criteria	4
Q6. "*What's what*"	Exercise designed to determine ability to generalize	5
Q7. "*What to do*"	Technical Conflict resolution problem	5
Q8. "*Whatever*"	Engineering design problem typical of undergraduate design contests; Requires generation of concepts only	15

Table 3: DT Measures and Scoring

		DT QUESTIONS								**Sub-skill Scores**	
DT SUB-SKILLS	**Metric**	1	2	3	4	5	6	7	8	Raw Total	Normalized Total
Fluency	number of ideas	x	x			x		x			
Versatility	variety of ideas	x	x			x		x			
Originality											
Strength	highest novelty	x	x		x	x		x	x		
Mean	average novelty	x	x		x	x		x	x		
Abstraction-ability	no. of levels and relations					x	x				
Fixation avoidance	conceptual distance from exposed example			x			x	x			
Complexity handling	no. of parts, relations				x				x		
Elaboration	functionally relevant details				x			x	x		
Functional quality	how well specs are met				x			x	x		

As can be seen from Table 3, every sub-skill 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. The raw score for each sub-skill is converted to a normalized score using normed data.

3.3 Visual Thinking & Spatial Reasoning

Visual Thinking is the ability to use internal (mental) or external (sketches) graphical representations to understand a problem, work towards a solution or represent ideas. Visual thinking involves the interaction between mental images (imagining), graphical images (drawing), and perceptual images (seeing) [22]. A two pronged approach is being taken to construct tests for visual thinking skills. One is to use existing theories of spatial reasoning, visual perception, memory and pattern recognition. The other is to look at design research literature. The role of visual thinking in ideation has been studied extensively. In architectural design, Goldschmidt [55] studied how "serial sketching" progresses, and how unexpected relationships and new shapes emerged outside the scope intended. Through the cycle of sketching, inspection, and revising, the designer is in a sense having a conversation [56], but this conversation is greatly affected by one's ability to use imagery. According to Verstijnen, [32], creative discovery is the result of a set of mental operations on a visual image. Sketches help in capturing fleeting images and may provide additional connections and visual insight

[22,56]. Sketching is the medium for improving the evaluation and restatement of design problems or even problem comprehension [57]. Since pattern seeking occurs naturally in visual thought, connections are more spontaneously made in the designers mind. A gestalt phenomenon occurs when the designer reviews a sketch and is able to extract information beyond what was originally intended. Ambiguity in a sketch may spark 'unexpected connections', which is a promoter of creativity in design. Freehand sketches, characterized by ambiguity and informational denseness, contribute positively toward creative and explorative aspects of problem solving. Design is, therefore, purported to be a reflective, responsive and opportunistic process whereby designers construct their own reality through a unique design situation [56]. Several empirical studies have tested and confirmed the hypothesis that graphical (pictorial) representation leads to higher variety and novelty than textual (sentential) [26,27].

VT is probably the most complex of all skills to evaluate because there are several different but related visual abilities. We need to measure the frequency of "voluntary use" (without being directed to do so) of graphical representations as an indication of the subjects preferred medium, both internal and external; comprehension of pictorially presented stimuli, including graphical pattern recognition; ability to generate, manipulate, retrieve and transform graphical representations, which may be spatial or non-spatial; level of detail present and accuracy of representation. To sort through the multitude of factors we classify them along two assessment axes: internal/external and spatial/non-spatial representations. A comprehensive test must include these four combinations for both general and technical level tests.

Based on the foregoing analysis we identified the VT skills listed in Table 4. Some of these skills are quite straightforward to assess, *e.g,* the quality, accuracy, detail of external graphical representations (sketches, figures). However, the skills listed in Table 4 are too many to be on a single test; we need to split them into multiple tests.

Table 4: VT skill categories

Visual Comprehension	**Visual reasoning /inferencing**	**Visual memory**	**Freehand sketching**	**Technical sense**
✓Feature recognition • Object recognition • Image completion • Semantic relation recognition	✓Mental transforms (Rotate, reflect,..) ✓views transforms ✓Visual synthesis • Motion simulation • inconsistency discovery in image • representation Re-formulation	✓Memory recall ✓Pattern recognition	• Preference for graphical rep. • Sense of Proportion • Appropriate shapes • embellishments	• Spatial reasoning • 3D mech. motion • Object manipulation thru constrained space • assembly arrngmnt • Assembly sequence • Understanding device function • orthog projection • sectioning

Before developing our VT Tests we studied nine such tests to determine the coverage for the VT skills identified in Table 4: Mental Cutting Test [58], Rey Complex Figure Test [59], Kit of Factor-Referenced Cognitive Tests [60], Mental Rotations Test [61], Purdue Test [7], Cube Comparisons Test [62], Arabic Figures Test [63-64], Developmental Test of Visual Perception [65], Mental Cutting Test "Schnitte" [66]. Common elements in several of these tests include pattern recognition by embedding figures in a background [60,65]; mentally rotating objects and matching to result [60-65]; slicing 3D shapes with planes [58,66]; memorizing line drawings and recalling parts of them [59,60]. Less common elements of these tests are folding 2D patterns into 3D shapes [7], combining simple shapes into specified patterns [60], and viewing objects from different directions [7]. The latter can be considered the inverse of mental rotations, i.e. the observer rotates rather than the object. Image completion is covered in a superficial way by one test [65] by using dashed lines instead of continuous in object recognition. The check marks in Table 3 indicate factors for which standard tests already exist, although not all these items are on any single test. A test developed by combining non-redundant elements of Kit, Purdue and RCFT would cover all the checked marked items. One might note that the items not covered on any of the existing tests are engineering design oriented. For example, arranging a set of connected objects in a limited space (2D or 3D version); detecting parts that cannot be removed from an assembly (given a drawing), constructing 3D object views from 2D obscured or

partial drawings. We are in Stage 3 of creating a test to address the missing elements. Again, as correlation between items is established the list of items and questions will become more manageable.
Sketching and use of graphical representations are external evidence of visual thinking. These are relatively easy to evaluate, based on indicators such as frequency of use of such representation, quality, accuracy, level of detail and effectiveness in representing, evaluating a given issue. Internal representations are not as straightforward to identify and evaluate. However, sub-factors of image manipulation, visual pattern recognition and spatial reasoning are independently testable. Unlike DT Tests which can be completely covered by textual/verbal medium, VT requires use of multiple mediums: text, sketches/pictures, computer models/graphics, and physical models.

3.4 Tests of other skills

Our project includes the development of two additional tests: problem formulation and qualitative reasoning. While we are not as far along in these areas, a quick look about the current status is given.

3.4.1 Problem Formulation (framing)

No matter how simple or how complex the designed system or component, no matter if it is novel or routine design, any designer must first understand what the design requirements are, including the functions, specifications, explicit or implicit objectives and explicit or hidden constraints. Problem formulation is not a one shot process- it co-evolves with the solution. The intent of our tests is to use the *initial encoding* of the problem as the basis of test. The indicators of problem formulation skills that have been identified are:

- ability to discover both explicit and hidden objectives, requirements, constraints;
- detecting missing information, synergistic and conflicting requirements
- identification of key issues (the real problem);
- ability to distinguish between real and fictitious constraints;
- ability to relate needs into technical requirements;
- mapping a problem to higher dimension (asking why is this a requirement)
- relating a problem to similar problems (analogical reasoning)

The above factors can be directly evaluated by posing design problems with hidden requirements, conflicts, fictitious constraints, missing critical information and so on. We have not found any prior standardized tests specifically designed to test problem formulation ability. However, HPS studies consider structuring of problem space as a foundational activity. We would be interested in knowing what mental representation of the problem is being used by the subject. Various graphical structures have been proposed to externalize problem understanding in terms of concepts and relations between the concepts, such as semantic nets and concept maps.

3.4.2 Qualitative Reasoning (QT)

Although not studied as extensively as Visual Thinking, QT is a key skill in engineering design and one that apparently not picked up in engineering science classes. In most engineering classes, students work with precise definitions and complete information; deductive reasoning is the only skill that is developed in this manner. Because design problems are not in this form, students have considerable difficulty applying analytical methods at the conceptual stage (too many unknowns to be able to use the formulas). What is needed is the ability to abstract mathematical formulas into qualitative relations. For example, making an observation like, “if I increase the surface to volume ratio, I can reduce the internal temperature”, or “there is conflict between the objective of capturing the most solar energy and retaining it: one requires surface area increase, the other surface area decrease”.
Many studies have found various types of instructions aimed at increased abstraction can increase the originality, and in some cases, the quality of creative ideas (e.g., [67,68]). One way to make creative ideation more abstract is through the use of analogies; greater abstraction is involved in more distant between-domain analogies as compared to local analogies. In creative conceptual design, both within-domain and between-domain analogies have been observed to be used by real teams of designers [69]. In their study, they found that when designers worked from prototypes (i.e., exemplars of design ideas) they generated few between-domain analogies, indicating that exemplars can constrain ideation by limiting abstraction that can occur through analogizing.
Several Abstract Reasoning indicators need to be measured: the ability to make assumptions when presented with incomplete information; simplifying formulas and knowledge of physical behavior to

reveal key qualitative relationships, ability to evaluate concepts/ideas without detailed analysis and evidence of analogical reasoning. The last factor is important because studies have claimed that analogical reasoning is the primary mechanism underpinning the development of domain expertise [70]. Generalized knowledge schemas embody abstract understanding of problem types and also embody procedural understanding of how to solve problems of a particular type [70]. Protocol studies have been used to determine if such abstracted representations are being used or more literal ones [70]. We propose to create general tests that require minimal technical knowledge to test each of the above factors. One example problem that may be suitable is the design of a device that would raise water temperature in a container (made from limited materials) by the largest amount. This involves some abstract thinking between conflicting objectives: collecting the most energy (high radiation, conduction) and not losing any collected energy (low conduction, radiation). This exercise can be a good for differentiating between two groups: those that can abstract equations and those that cannot.

4. TESTS VALIDATION

There are two different properties that need to be considered, Reliability and Validity. *Reliability* is the extent to which a test measures true difference in individuals versus measurement errors; *Validity* is how well a test measures what it claims to measure [71]. Reliability is evaluated with two criteria:

- Stability – does the test give the same result for the same person each time (Test-Retest criteria)
- Internal consistency – are the items on the test homogeneous (related to the same skill)

Stability can be determined by computing Pearson Correlation Co-efficient for consecutive tests done on the same person over time. The difficulty of evaluating this lies in determining the time interval; if it is too long, some learning might take place that alters the skill level, e.g., a student might take a drawing class in the intervening time. We may only be able to assess stability from individuals who are presumed to have reached a constant proficiency level, which rules out students. Internal consistency, determined by Cronbach's Alpha measure [71,72], finds the correlation between items on a test and also between item and total score. This measure is used to determine whether an item (question) should be included in the test or not.

Reliability only tells us that something is being measured consistently but does not tell us anything about how well a test measures what it claims to measure. This aspect can be evaluated by 2 criteria:

- Criterion validity – can it predict how one will perform on something else
- Construct validity – is the characteristic measured consistent with the its definition

We propose to determine criterion validity by predicting how one will do on a design task 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 sub-sets of our items to those sets. In other areas, such as problem formulation, we might not have anything equivalent to compare to at this time.

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 minimum of 400 [73]). We have compiled a substantial database from beta tests, associating response categories with scores, for use in norming.

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.

Another indirect type of validation is to see how students do on design projects and exams. In fact, we correlated DT scores to how well students did in design contests [44]. The correlation factor was 0.65.

Full validation of the tests will require collection of enormous amounts of data from large number of participants using factorial variants of the tests. This will take several years and would require a community effort and buy-in. Before starting validation studies, we are conducting two sets of trials: alpha and beta level, as shown below. Alpha trials have been completed and beta are in progress.

We have set up a website with an open invitation to the design academic community to participate in the beta trials. Any University can request the tests, administer them to their own students and return the tests to us for scoring. After beta testing is complete we will provide instructions and databases to other educators to score their own tests. Cross-scoring between institutions and evaluators will also determine consistency.

TRIALS	OBJECTIVES	LOCATION	Population
Alpha trials	improve test language, format; add/delete items	internal	100+
Beta trials	generate normative database of responses, categories and members	external	500

5. DISCUSSION

The potential benefits of this work are in three main areas: engineering design education; creation of more effective, innovative design teams in industry; creating a bridge between design theory and methodology and the extant knowledge in cognitive psychology and psychometrics. To encourage the development of design skills in our students we must not only have project based instruction but should also reward out-of-the-box thinking, risk-taking, unconventional and unusual ideas. Even if design exercises given were designed specifically to teach/evaluate certain design skills, recording a single score is not adequate. The conventional grading system hides the specific strengths and weaknesses of an individual. Domain knowledge is necessary but not sufficient for engineering design. We must emphasize the development of design skills and evaluate them explicitly in standardized ways. A designer's profile can be represented in a standard way by scores on relevant design skills. Designer Profiles can be useful 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 course content.

ACKNOWLEDGEMENT

This research is supported by the US National Science Foundation Grant CMMI-0728192. Opinions expressed in this paper are those of the authors and not endorsed by NSF.

REFERENCES

[1] National Research Council (U. S.), Improving Engineering Design: Designing for Competitive Advantage, National Academy Press, Washington, D. C, 1991.
[2] Stern W (1912) Psychologische Methoden der Intelligenz-Prüfung. Barth, Leipzig, Germany:
[3] Terman, L.M. et. al., (1917). The Stanford Revision of the Binet-Simonscale for measuring intelligence Baltimore: Warwick and York.
[4] Torrance E, 1966, Torrance tests of creativity, Princeton, Personnel Press.
[5] Watson-Glaser, 1980, Critical Thinking Appraisal, Psych Corp, Cleveland, OH.
[6] Hicks R, Southey G, 1990, The Watson-Glaser critical thinking appraisal and performance of business management students, Psych Test Bull 3, 74-81.
[7] Guay, R., *Purdue Spatial Visualization Test: Rotations*, Purdue Research Foundation, West Lafayette, IN, 1977.
[8] Kendal I, Izard J, 1988, ACER mechanical reasoning test (revised), Psych Bull 1.
[9] Sorby, S, 2000, The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students, *Journal of Engineering Education,* July.
[10] Zimmerman C, Development of scientific reasoning skill, Dev Review 20, 2000.
[11] Klahr & Dunbar, 1988, Dual search space during scientific reasoning, Cognitive Science, 12.
[12] Newell & Simon H, 1972, Human problem solving, Prentice-Hall.
[13] Kulkarni D, Simon H, 1988, The process of scientific discovery: The strategy of experimentation, Cog Sci, 12
[14] Newstetter W, Khan S, A developmental approach to assessing design skills and knowledge, 27th FIE conf, V2, Nov 1997.
[15] Schunn C, Anderson J, General and specific expertise in scientific reasoning, Proc. 19th annual meeting, Cognitive Science Soc.
[16] Box G, Liu P, 1999, Statistics as a catalyst to learning the scientific method, J. Quality Tech, 31(1).
[17] Newell A, 1990, Unified theories of cognition, Harvard Univ. Press.
[18] Marshall S, 1985, Schemas in problem solving, Cambridge Univ Press.
[19] Cross N., Dorst K., Roozeburg N., Research in Design Thinking, Delft Press, 1991.
[20] Dym C, Agogino A, Eris O, Frey D, Leifer L, Engineering design thinking, teaching and

Learning, J. Eng Educ, Jan 2005.
[21] DeBono, E., 1970, Lateral Thinking: Creativity Step by Step, Harper and Row, New York.
[22] McKim, R., Experiences in Visual Thinking, Wadsworth, Inc., Belmont, CA, 1980.
[23] Park, J. A. and Kim, Y. S., Visual Reasoning and Design Processes, *Int'l. Conf. on Engineering Design (ICED),* Aug. 28-31, 2007.
[24] Burvil C, Field B, Lewis W, Weir J, 2005, A quantitative study of ideation, visualization and graphical skills in engineering design, ICED 2005, Melbourne, Australia, July.
[25] Ullman, D., Wood, S., and Craig, D., (1990), The Importance of Drawing in the Mechanical Design Process, *Computers and Graphics*, Vol. 14, No. 2, pp. 263-74.
[26] McKoy, F., Vargas, 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.
[27] Larkin, J., Simon, H., (1987), Why a Diagram is (Sometimes) Worth Ten Thousand Words, *Cognitive Science*, Vol. 11, pp. 65-99.
[28] Tovey, M., 1986, Thinking Styles and Modeling Systems, *Design Studies*, Vol. 7, No. 1
[29] Cross, N., Cross, A. (1996), Winning by Design: The Methods of Gordon Murray, Racing Car Designer, Design Studies, Vol. 17, No. 1, pp. 91-107.
[30] Lawson, B., (1994), Design in Mind, Butterworth Architecture Press, Oxford, UK.
[31] Roy, R., (1993), Case Studies of Creativity in Innovative Product Development, *Design Studies*, Vol. 14, No. 4, pp. 423-43.
[32] Verstijnen, I., Van Leeuwen, C., Goldschmidt, G., Hamel, R., and Hennessey, J., (1998), Sketching and Creative Discovery, *Design Studies*, Vol. 19, No. 4, pp. 519-46.
[33] Akin O, 1999, Architecture students spatial reasoning with 3D cube arrangements, Draft paper, www.andrew.cmu.edu/~oa04/Papers
[34] Perry W, Forms of intellectual and ethical development in college years: A scheme, San Francisco, Jossey Bass, 1970 and 1998.
[35] King P, Kitchener K, Developing reflective judgment: Understanding & promoting intellectual growth and critical thinking in adolescents and adults, San Francisco, Jossey Bass, 1994.
[36] Jung C, 1971, Psychological type in Collected works of CG Jung 6, Princeton.
[37] Wilde D, 1999, Design Team Roles, ASME Design Theory & Meth. Conference, Las Vegas
[38] Guildford J., 1967, The nature of human intelligence, McGraw-Hill, New York
[39] Torrance, E, Role of Evaluation in Creative Thinking, B of Educ research, U Minnesota, 1964.
[40] Mednick, S., (1962), The Associative Basis of the Creative Process, *Psychological Review*, 69.
[41] Lohman D, 1993, Spatial ability and G, 1st Spearman seminar, Univ. Plymouth, UK, July.
[42] Eliot & Smith, 1983, An international directory of spatial tests, NFER-Nelson, Windsor, England:
[43] Johnson-Laird, P. N. (1983). Mental models: Towards a cognitive science of language, inference, and consciousness. Cambridge, MA: Harvard University Press.
[44] Shah J.,(2005) Identification, Measurement & Development of Design Skills In Engineering Education, Int Conf on Eng Design (ICED05), Melbourne, Australia, August, 2005.
[45] Orhun E, Orhun D, Cognitive tools for assessing design knowledge and skills, IEEE 2004.
[46] Finke R A, Ward T B and Smith S M Creative Cognition: Theory, Research, and Applications MIT Press, Cambridge MA , 1992.
[47] Jansson, D.G., and Smith, S.M., (1991), Design Fixation, *Design Studies*, 12 (1), 3-11.
[48] Goff, K. and Torrance, E.P. (2002). Abbreviated Torrance Test for Adults. Bensenville, IL: Scholastic Testing Services, Inc.
[49] Meeker, M. & Meeker, R. (1982) Structure of Intellect Learning Abilities Test: Evaluation, Leadership, and Creative Thinking. El Segundo, CA: SOI Institute
[50] Meeker SOI checklist, http://www.soisystems.com/
[51] Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity–intelligence distinction.* New York: Holt, Rinehart, & Winston
[52] Guilford, J.P. (1950). Creativity. *American Psychologist, 5,* 444-454.
[53] http://asudesign.eas.asu.edu/testsportal/index.php
[54] Shah, J. J., Smith, S. M., Vargas-Hernandez, N., (2003), Metrics for Measuring Ideation Effectiveness, *Design Studies*, vol. 24, no. 2, pp. 111-134.
[55] Goldschmidt, G., 1992, Serial Sketching: Visual Problem Solving in Designing, Cybernetics and Systems, Vol. 23, No. 2, pp. 191-219.

[56] Suwa, M., Tversky, B., What Do Architects and Students Perceive in their Design Sketches? A Protocol Analysis, Design Studies, vol. 18, no. 4, 1997, pp. 385-403.
[57] McGown, A., Green, G., Rodgers, P., Visible Ideas: Information Patterns of Conceptual Sketch Activity, Design Studies, vol. 19, no. 4, 1998, pp. 431-53.
[58] CEEB Special aptitude test in spatial relations, College Entrance Examination Board, USA, 1939.
[59] Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie* 28: 215–285
[60] Ekstrom, R.B., French, J.W., Harman, H.H., Dermen, D. (1976). Kit of Factor-Referenced Cognitive Tests. Princeton, NH: Educational Testing Service
[61] Vandenberg, S.G. and Kuse, A.R., 1978. Mental rotation, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47,* pp. 599–604.
[62] Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. Psychological Review, 92, 137-172.
[63] Seddon, G.M., Eniaiyeju, P.A., & Chia, L. H. L. (1985). The factor structure for mental rotations of three-dimensional structures represented in diagrams. Research in Science & Technology Education, 3(1), 29-42.
[64] Shubbar, K. E. (1990). Learning the visualization of rotations in diagrams of three dimensional structures. Research in Science & Technological Education, 8(2), 145-154.
[65] C. R., Pearson, N. A., & Voress, J. K. (*2002*). *Developmental Test of Visual Perception–* Adolescent and Adult. Austin, TX: Pro-Ed Publishing.
[66] Quaiser-Pohl, C. (2003). The mental-cutting test "Schnitte" and the Picture Rotation Tests - Two new measures to assess spatial ability. *International Journal of Testing, 3,* 219-231.
[67] Novak J, Gowin D, 1984, Learning how to learn, Cambridge U. Press, Cambridge.
[68] Mumford. M. D., Reiter-Palmon, R., & Redmond, M. R. (1994). Problem construction & cognition: Applying problem representations in ill-defined problems. In M. A. Runco (Ed.) *Problem-finding, problem-solving, and creativity*. Norwood, NJ: Ablex, 3-39.
[69] Christensen, B. T., & Schunn, C. D. (2007). The relationship of analogical distance to analogical function and pre-inventive structure: The case of engineering design. *Memory & Cognition*.
[70] Ball L, Ormerod T, Morley N, 2004, Spontaneous analogising in engineering design: a comparative analysis of experts and novices, Design Studies, V25, pp 495-508.
[71] Kline, P. (1993). *The handbook of psychological testing*. Routledge, London
[72] Cronbach L.J. (1990) Essentials of Psychological Testing (5th Edition), Harper & Row, NY.
[73] Rose, R. G. (1993). *Practical issues in employment testing*. Psych Assmntt, Inc., Odessa, FL

Contact: Jami J. Shah
Arizona State University
Mechanical & aerospace Engineering Department
MC 6106, Tempe, Arizona 85287
USA

E-mail: jami.shah@asu.edu
URL: http://asudesign.asu.edu

Jami Shah is Professor of Engineering and Director of Design Automation Lab at ASU. His research areas include: Cognitive studies of design ideation, geometric computing, AI & Knowledge based systems and Engineering Metrology. He is the founding chief editor of ASME Transaction, the *Journal of Computing & Information Science in Engineering(JCISE)*. He is Fellow of ASME.

Steve Smith is Professor of Psychology at Texas A&M. His current research examines memory, creative cognition, and meta-cognition, particularly fixation. He is the co-author of the book "Creative Cognition".

Jay Woodward is Lecturer in Dept. of Educational Psychology at Texas A&M University. He specializes in educational testing, creativity research and learning. In addition to his teaching responsibilities, he is the Director of YAP, a career exploration summer program for gifted youth.