View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DSpace@MIT

Res Eng Design (2013) 24:277–295
DOI 10.1007/s00163-012-0145-x

ORIGINAL PAPER

# Empirical evaluation of procedures to generate flexibility in engineering systems and improve lifecycle performance

**Michel-Alexandre Cardin · Gwendolyn L. Kolfschoten ·
Daniel D. Frey · Richard de Neufville ·
Olivier L. de Weck · David M. Geltner**

Received: 6 October 2011 / Revised: 4 September 2012 / Accepted: 1 October 2012 / Published online: 13 October 2012
© Springer-Verlag London 2012

**Abstract** The design of engineering systems like airports, communication infrastructures, and real estate projects today is growing in complexity. Designers need to consider socio-technical uncertainties, intricacies, and processes in the long-term strategic deployment and operations of these systems. Flexibility in engineering design provides ways to deal with this complexity. It enables engineering systems to change in the face of uncertainty to reduce impacts from downside scenarios (e.g., unfavorable market conditions) while capitalizing on upside opportunities (e.g., new technology). Many case studies have shown that flexibility can improve anticipated lifecycle performance (e.g., expected economic value) compared to current design and evaluation approaches. It is a difficult process requiring guidance and must be done at an early conceptual stage. The literature offers little guidance on procedures helping designers do this systematically in a collaborative context. This study investigated the effects of two educational training procedures on flexibility (current vs. explicit) and two ideation procedures (free undirected brainstorming vs. prompting) to guide this process and improve anticipated lifecycle performance. Controlled experiments were conducted with ninety participants working on a simplified engineering systems design problem. Results suggest that a prompting mechanism for flexibility can help generate more flexible design concepts than free undirected brainstorming. These concepts can improve performance significantly (by up to 36 %) compared to a benchmark design—even though users did not expect improved quality of results. Explicit training on flexibility can improve user satisfaction with the process, results, and results quality in comparison with current engineering and design training on flexibility. These findings give insights into the crafting and application of simple, intuitive, and efficient procedures to improve lifecycle performance by means of flexibility and performance that may be left aside with existing design approaches. The experimental results are promising toward further evaluation in a real-world setting.

**Keywords** Conceptual design · Design procedures · Empirical evaluation · Flexibility in engineering design · Real options

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00163-012-0145-x) contains supplementary material, which is available to authorized users.

---

M.-A. Cardin (✉)
Department of Industrial and Systems Engineering,
National University of Singapore, Block E1A, #06-25,
1 Engineering Drive 2, Singapore 117576, Singapore
e-mail: macardin@nus.edu.sg

G. L. Kolfschoten
Faculty of Technology, Policy, and Management,
Delft University of Technology, 5 Jaffalaan,
a2.270, 2628 BX Delft, Netherlands

D. D. Frey
Department of Mechanical Engineering,
Massachusetts Institute of Technology, 77 Massachusetts
Avenue, 3-449D, Cambridge, MA 02139, USA

R. de Neufville
Engineering Systems Division, Massachusetts Institute
of Technology, 77 Massachusetts Avenue, E40-245,
Cambridge, MA 02139, USA

O. L. de Weck
Department of Aeronautics and Astronautics,
Massachusetts Institute of Technology, 77 Massachusetts
Avenue, 33-410, Cambridge, MA 02139, USA

D. M. Geltner
Department of Urban Studies and Planning,
Massachusetts Institute of Technology, 77 Massachusetts
Avenue, W31-310, Cambridge, MA 02139, USA

## 1 Introduction

"We created a marvelous technological achievement. Then, we asked […] how to make money on it." (Leibovich 1999) The words from Iridium's CEO explained the bankruptcy of the largest commercial satellite communication system ever engineered. The 77 low earth orbit (LEO) satellite infrastructure developed for US $4 billions enabled phone calls anywhere on the planet. The design and management processes were centered on very optimistic demand projections. The technology was working beautifully. This led to the rapid deployment of the constellation between May 1997 and May 1998 (MacCormack and Herman 2001). This inflexible design and rigid deployment strategy—combined with underestimation of demand for land-based cell phone technology—are cited among the possible causes for this economic demise (de Weck et al. 2004).

This case demonstrates that the design and management of engineering systems today—for example, airports, communication infrastructures, real estate projects—need to go beyond technological considerations. Engineering systems are characterized by a high degree of technical complexity, social intricacy, and elaborate processes, aimed at fulfilling important functions in society (ESD 2011). Dynamic socio-technical elements like markets, operational environment, regulations, and technology play a significant role in their success (Minai et al. 2006). Crucial decisions have to be made in early conceptual design phases regarding the system's strategic and long-term evolution.

de Weck et al. (2004) revisited the Iridium case after its initial downfall and suggested a *flexible design concept*—a flexible *strategy* and *enabler* in design—that would have saved up to 20 % in expected development cost. The flexibility would have protected the organization from lower demand scenarios and losses by reducing the initial capital expenditure. It would have positioned the system to capitalize on high demand opportunities by enabling capacity expansion. The strategy involved a flexible staged deployment of the constellation, starting with fewer satellites, and deploying more *only if demand* reached a certain level. This approach, however, would require a different satellite design. Each satellite would be designed to change orbital configuration, enabling the constellation to reorganize and expand coverage area as demand grows. This approach contrasts to satellites designed to reach and stay in a specific orbital configuration, as in the Iridium case.

de Weck et al.'s (2004) analysis raises the question why this design strategy was not considered. A full answer to this question is beyond the scope of the paper. Cardin and de Neufville (2009) suggest, however, that the answer may be rooted in traditional engineering culture. Uncertainty is often considered only through sensitivity analysis after a design is selected. Designers often rely on high-fidelity (or exact) models, making flexibility analysis more difficult from a computational standpoint.

The case here illustrates the impact that a lack of flexibility had on the economic lifecycle performance of the system. This example motivates this paper, concerned with devising and evaluating simple design procedures to help designers consider uncertainty and flexibility more systematically in the early design phases. It aims to provide a better understanding of the training and creativity approaches that should be used for engineering education and practice.

There are other challenges to developing efficient procedures to support early generation of flexibility in engineering systems. The benefits of flexibility may be difficult to quantify relative to the additional costs and design efforts. Minai et al. (2006) outline the cultural issue that engineering thinking often relies on linear, deterministic projections of future operating conditions. The design is then optimized for a set of market scenarios, requirements, and constraints, even though those are prone to change (Eckert et al. 2009). The groupthink phenomenon may cause engineers to be so focused on a solution that group pressure may cause tunnel vision and critique to be collectively ignored (Janis 1972). Financial evaluation tools based on discounted cash flow (DCF) and net present value (NPV)—often used in engineering project evaluation—do not integrate adaptive management over time, assuming that all deployment decisions are made as of $t = 0$ (Trigeorgis 1996). Murman et al. (2002) identifies a silo culture in enterprises that can hinder flexibility thinking. Given there are typically many uncertainty sources, design variables, and parameters to consider, it is not clear where to focus the design effort.

The study is motivated by the above challenges. It investigates empirically the effects of simple, intuitive, and efficient collaborative procedures to guide and stimulate early generation of flexibility in engineering systems. Current educational training and professional experience as it relates to flexibility are compared to an explicit training session on this topic. Free undirected brainstorming is compared to a prompting ideation mechanism geared toward flexibility, supported by group support system (GSS) technology. These procedures are evaluated in a controlled experiment where participants tackle a simplified real estate design problem.

The paper addresses the following research question: "What are the main and interaction effects of the proposed procedures on the quantity of flexible design concepts, anticipated lifecycle performance (e.g., expected economic value of the system) of an engineering system, user satisfaction with the process and results, and anticipated quality of results?" The hypotheses are that explicit training combined with a prompting mechanism specifically geared toward flexibility will produce main effects and help generate more flexible design concepts. In turn, the concepts will improve anticipated lifecycle performance compared to current training and free undirected brainstorming. These procedures will improve user satisfaction with the process and results, and anticipated quality of results.

In the remainder of the paper, related work is presented in Sect. 2. The procedures are presented in Sect. 3. The experimental methodology is explained in Sect. 4. Results are presented and discussed in Sect. 5. Section 6 concludes and proposes future avenues for research on flexibility in engineering design.

## 2 Related work

### 2.1 Flexibility in engineering design/real options

Flexibility in engineering design enables a system to change in the face of uncertainty (Fricke and Schulz 2005). It is associated with the concept of a real option, providing the "right, but not the obligation, to change a project in the face of uncertainty" (Trigeorgis 1996). Real options exist "on" a project, involving higher-level managerial decisions like abandoning, deferring until favorable market conditions, expanding/contracting/reducing capacity, deploying capacity over time, switching inputs/outputs, and/or mixing the above (Trigeorgis 1996). Real options "in" project are technical engineering and design components enabling options in operations (Wang and de Neufville 2005). Real options are referred here interchangeably with *flexible design concepts*.

The real options analysis (ROA) literature focuses on the economic valuation of flexibility (Trigeorgis 1996). It builds upon work in financial options by Black and Scholes (1973) and Cox et al. (1979). Many studies have shown that flexibility can bring expected performance improvements ranging between 10 and 30 % compared to standard design and evaluation approaches (de Neufville and Scholtes 2011). Flexibility improves expected performance by affecting the distribution of possible outcomes. It reduces the effect from downside, risky scenarios while positioning the system to capitalize on upside, favorable opportunities. Examples in engineering systems design abound: development of innovative water technologies (Zhang and Babovic 2012), offshore oil platform design for future capacity expansion (Jablonowski et al. 2008), adaptive supply chain mechanisms for uncertain exchange rates (Nembhard et al. 2005), etc.

One example in real estate is the ability to expand a building vertically (Guma et al. 2009). The HCSC building in Chicago exploited the strategy to "build small and expand later *if needed.*" This strategy reduced exposure to losses because less capital was required upfront. It also gave access to more profits under favorable market conditions to build more offices, hire personnel, and ultimately generate more profits. This expansion strategy was carefully enabled in the design in the early 1990s (e.g., larger elevator shafts, stronger structure). The company exercised the flexibility a few years ago, with the expansion phase completed in 2011.

### 2.2 Concept and idea generation procedures

#### 2.2.1 *Definition and categories*

Concept and idea generation is a human process bringing designers together to develop both practical and unconventional design concepts in an engineering setting (Kurtoglu et al. 2009). Shah et al. (2000) classified idea generation (IG) techniques either as intuitive or logical. Free undirected brainstorming (Osborn 1957) is an example of intuitive germinal technique, while TRIZ (Altshuller 1973) is a logical, history-based approach. The review by Knoll and Horton (2010) shows that ideation mechanisms can be classified based on analogy, provocation, and random changes of perspective. Analogy uses knowledge in a similar domain or system setting to generate new ideas. Provocation challenges the underlying assumptions of the creative task. Random relies on external stimuli unrelated to the task.

#### 2.2.2 *Group support system (GSS) technology*

Group collaboration may put barriers to creativity, resulting in the productivity loss. Evaluation apprehension (fear of being judged), free riding (letting others do the work), and production blocking (losing an idea because someone else is talking) are among potential causes (Mullen et al. 1991).

GSS technology minimizes productivity loss and stimulates creativity in collaborative activities (Bostrom and Nagasundaram 1998). GSS is defined as "socio-technical systems consisting of software, hardware, meeting procedures, facilitation support, and a group of meeting participants engaged in intellectual collaborative work." (de Vreede et al. 2003) Because the productivity loss is reduced, brainstorming methods become more easily comparable. Using GSS, one can get a more direct recording of thinking patterns among group members and reduced bias from within-group dominance, pressure, or focus.

GSS technology has never been used in the context of analyzing engineering systems for flexibility. It was used here to stimulate creativity, record discussion content efficiently, structure the collaborative design process, and help with moderation.

### 2.2.3 *Flexible design concept generation*

This process starts from an initial design, obtained by means of an existing or standard design process. This initial step is crucial to constrain the design space, since it is intractable to consider all possible sources of uncertainty and flexibility from scratch. The design space is expanded by explicitly considering uncertainty, with the goal of improving anticipated lifecycle performance by means of flexibility. This study focuses in particular on "known unknowns." These are referred as uncertainty sources known to engineers to have significantly impact on anticipated lifecycle performance (e.g., market demand, price, cost). Bahill (2012) presents the results in a related study of "unknown unknowns" and their unintended consequences on the design process.

Flexibility generation involves (1) *generating* concepts in response to major uncertainty sources and/or (2) *identifying* areas where to embed flexibility in the design. Fricke and Schulz (2005) suggested changeability principles based on industry guidelines to generate new concepts (e.g., ideality, simplicity, modularity). Trigeorgis (1996) introduced general real option strategies applicable to engineering systems design: defer investment until favorable market conditions, stage asset deployment over time, alter production capacity, abandon a loser project, switch production output and/or input, and grow by investing in research and development (R&D). Other approaches involve customers and stakeholders directly in the design process to safeguard against requirement changes and misunderstandings (Boehm et al. 2001; Herder and Bruijn 2009; Gil 2007).

Enabler identification methods rely mostly on design structure matrix (DSM) and platform methodologies—see reviews by Browning (2001) and Simpson (2004). Suh et al. (2007) suggested change propagation analysis (CPA) to look for change multipliers as areas to embed flexibility. Kalligeros (2006) suggested the sensitivity DSM (sDSM) to identify design variables most sensitive to changes, as indicators of subsystem or components to insert flexibility. Martin and Ishii (2002) suggested the generation variety index (GVI) and coupling index (CI) to standardize and modularize designs, thus enabling switching flexibility between product variants. Sered and Reich (2006) improved this framework with the standardization and modularization driven by process (SMDP) method that reduces engineering efforts, integrates within the DSM framework, and accounts for uncertainty in the design process. Mikaelian et al. (2011) suggested a systematic approach based on the [type, mechanism] characterization of real options.

This overview suggests that there is no simple, efficient, and intuitive procedure to help designers generate flexibility early in engineering systems (Cardin 2011). The real option [type, mechanism] characterization by Mikaelian et al. (2011) is a good start in this direction. The creative steps to flexibility generation, however, are not systematically stated. Industry guidelines do not provide a setting and techniques to stimulate creativity and organize collaborative design activities. Modularization and standardization techniques based on DSM, GVI, CI, and SMDP help identify opportunities for flexibility already embedded within a pre-defined description of the system. They have been used mostly for product design, and it is unclear how they scale for engineering systems design. They require building a DSM and/or system model describing component interactions before opportunities for flexibility can be identified—a non-trivial and time-consuming task. They enable switching flexibility between product variants, but do not explicitly consider other flexibility strategies requiring careful design considerations, like phasing capacity deployment, deferring, abandoning, etc. Focusing on a pre-defined system description may hinder creativity, as observed in an oil platform design case study (Cardin and de Neufville 2009). Many of these issues are alleviated by the procedures introduced in this study. This is because they rely directly on the designer's expertise with the system as opposed to detailed modeling, before flexibility can be generated.

### 2.2.4 *Experimental evaluation*

Many studies have evaluated concept and idea generation procedures in an experimental setting. Kolfschoten et al. (2009) studied different moderation techniques to help generate ideas. Reinig et al. (2007) studied different invocation of social comparisons to stimulate creativity. Santanen et al. (2004) studied the effects of different prompting rates on ideation quality. Kurtoglu et al. (2009) evaluated an online design library procedure integrating artificial intelligence principles to support concept generation. Linsey et al. (2010) evaluated a procedure to mitigate fixation in design sketching. Shah et al. (2001) compared the performance of the C-sketch procedure to the Gallery and 6-3-5 methods. van der Lugt (2002) compared brainsketching to traditional brainstorming. M Yang (2009) studied correlations between concept quantity and quality for the brainstorming, morphology charts, and sketching concept generation procedures. Chulvi et al. (2012) studied the effects of TRIZ, SCAMPER, brainstorming, and no method on concept novelty and utility.

Other studies have focused on the effects of education and pedagogy on engineering design. Daly et al. (2011) looked at the effects of teaching pre-defined design heuristics on concept quality. White et al. (2010) studied how teaching "principles of historical innovators" affect student creativity. Eppinger et al. (1990) studied how an interdisciplinary classroom environment affects product design and development. Robie et al. (1992) observed how different teaching methods affect designers' ability to form design abstractions. Buchal (2002) compared the effects of teaching computer aided design (CAD) versus sketches on concept generation. Okudan et al. (2010) compared the effects of teaching TRIZ versus design sketching on concept generation. Bender and Blessing (2003) studied how teaching systematic design methodologies affect concept performance.

This overview demonstrates that contributions are needed to understand the effects of educational training and ideation procedures on flexible design concept generation. The studies above do not explore uncertainty scenarios explicitly and do not guide design thinking about flexibility. Many existing procedures can be used (e.g., free undirected brainstorming) or adapted (e.g., prompting) for this purpose, motivating the approach taken in this study.

### 2.3 Empirical procedure evaluation

Evaluation metrics in empirical studies are often qualitative and subjective in nature. They may not support well-quantitative assessment of anticipated lifecycle performance as needed here. For instance, Kurtoglu et al. (2009) introduced completeness, the level at which a concept variant addresses a subfunction depicted in the function structure. Shah et al. (2000) suggested quality, quantity, novelty, and variety to assess creativity. Nelson et al. (2009) integrated these metrics to assess the quality of design space exploration. Briggs et al. (2006) used user satisfaction with the process and results.

Evaluating flexible design concepts should be based on quantitative measurements of anticipated lifecycle performance, as done in the real options literature. The metrics above, however, rely on expert assessments (e.g., using a 1–10 scale), with weights assigned based on the importance of the concept and intended functions. These may not be well suited for the intended purpose here.

Assessing lifecycle performance of engineering systems concepts can be challenging, even for an expert. So, many design variables, parameters, decision rules, long-term strategies, and scenarios need to be considered. Metrics like cost and weight do not measure how a concept will perform in operations. Similarly, a concept can be rated as highly complete, feasible, novel, or of high quality, there is no guarantee it will perform well once launched. Although positive correlations have been found between outcome quantity and quality (Yang 2009), it is not clear whether high quantity and variety of concepts necessarily improve performance. It is not clear either whether procedures providing good user satisfaction with the process and results necessarily lead to better performance. Even though no study has yet shown correlations between anticipated and actual lifecycle performance measurements—partly because engineering systems are long-lived (+20 years), making them difficult to study—these considerations motivate the modeling approach used in this study, based on quantitative anticipated lifecycle performance measurements, as well as qualitative user impressions.

**Table 1** Setup for 2 × 2 design of experiment (DOE)

| Educational training on flexibility ($E$) | Ideation mechanism ($I$) | |
|---|---|---|
| | Brainstorming ($-1$) | Prompting ($+1$) |
| Current ($-1$) | Treatment 1 | Treatment 2 |
| Explicit ($+1$) | Treatment 3 | Treatment 4 |

## 3 Choice of procedures

Table 1 summarizes the four procedures—or treatments—evaluated experimentally. To craft simple, efficient, and intuitive procedures, two factors with two levels each were considered: educational training received on flexibility ($E$) and ideation mechanism ($I$) used to stimulate creativity:

### 3.1 Educational training ($E$)

Educational training and pedagogy play a role in the ability to generate design concepts. A short *explicit training* program may therefore help designers generate flexibility in engineering systems, captured by level $E = +1$. The treatment is a short 15–20-min lecture on flexibility in large-scale infrastructure systems—a class of engineering systems (ESD 2011). The lecture[1] is expected to help designers become more aware of the effects of uncertainty on lifecycle performance. It should open their mind to the potential of flexibility to deal with uncertainty. The lecture content describes generic sources of uncertainty affecting lifecycle performance, why flexibility can improve such performance and why it must be considered in the early phases of design. It also discusses what important elements form a complete[2] flexible design concept. It provides real-world example applications of these principles in the aerospace and oil industries.

In reality, designers may or may not have received explicit training on flexibility during their educational training and professional experience. This reality is more likely to represent the wider population of designers. It is captured by factor level $E = -1$ and is called *current training*. In experiments, this treatment leaves participants

---

[1] Interested readers may refer to the supplementary material.

[2] Definition is provided in Sect. 4.6.4.

address the design problem without particular emphasis on flexibility. Participants generate concepts with the goal of improving performance based on their background and experience. Given the wide range of participants' background and experience, if they were exposed to flexibility thinking and thought it could improve performance, this procedure assumes that they would incorporate it in their thinking.

### 3.2 Ideation mechanism ($I$)

Design is inherently a social, creative, interdisciplinary, and collaborative process (Warr and O'Neill 2005). Generating flexibility in engineering systems is challenging, however, and requires guidance (de Neufville and Scholtes 2011). A *prompting* procedure may help scaffold the thought process systematically, as captured by level $I = +1$. Prompting is simple, intuitive, and useful to stimulate creativity in collaborative activities by supporting generic directions of thinking (Santanen et al. 2004). Asking direct questions may trigger collective discussions more effectively than relying on industry or real option guidelines alone. Prompting is similar to the approach used by researchers working on flexibility analysis with senior engineers and decision makers, as in the studies by Suh et al. (2007) and Mikaelian et al. (2011). Example prompts[1] used in this ideation mechanism were "What are the major sources of uncertainty affecting the future performance of this system?", "What flexible strategies would enable the system to change and adapt if the uncertainty scenarios you just discussed occur during operations?", "How should you prepare, engineer, and design this system to enable the flexibilities just discussed?", or "How should you manage and decide when it is appropriate to use, or exercise, the flexibilities in this system?", supported by industry guidelines and general real option strategies.

Free undirected brainstorming is a simple and intuitive approach to stimulate creativity (Osborn 1957). It is widely used in US industry and academia (Yang 2007). Captured in experiments by level $I = -1$, it encourages designers to focus on quantity, welcome unusual ideas, avoid criticism, and combine ideas or improve existing ones. This procedure was chosen mainly because it is widely used in practice, intuitive, and has few simple rules for training purposes.

To alleviate the concerns about productivity loss explained in Sect. 2.2.2, both ideation mechanisms were supported by GSS technology. While the mechanisms are referred as brainstorming ($I = -1$) and prompting (level $I = +1$) throughout the study, the procedures are in reality a combination of GSS technology with prompting or free undirected brainstorming activities.

## 4 Experimental methodology

### 4.1 Overview

The methodology focused on evaluating quantitatively the flexible design concepts generated in experiments. The real options literature stresses the importance of doing this to decide whether flexibility is worth the additional cost and design effort. The methodology was inspired from case studies quantifying the anticipated lifecycle performance of flexible engineering system in economic terms, loosely involving the following steps:

1. Describe a design problem and initial benchmark/standard solution;
2. Interview and/or discuss with designers/engineers/managers to elicit major uncertainty sources and potential flexible design concepts;
3. Develop a computer/analytical model to quantify the anticipated lifecycle performance of flexible designs;
4. Compare between flexible design concepts and the initial benchmark design solutions to demonstrate whether flexibility improves performance.

The experimental approach enabled efficient replication of these steps in a controlled setting. The quantity of complete and good flexible design concepts was measured to assess creativity, together with their anticipated lifecycle performances. User impressions of satisfaction with the process and results, as well as anticipated quality of results were measured. These measurements helped determine whether the procedures were simple enough and user-friendly to favor dissemination in industry practice and engineering education. They also helped determine whether participants valued flexible solutions, unbiased by the type of training received.

### 4.2 Participants

Ninety participants were recruited from professional masters and doctoral programs in engineering systems, design, and management at a top US engineering institution. Table 2 summarizes their demographics. They were recruited via class and electronic email announcements at the beginning of two courses on systems design and engineering analysis. The announcement invited voluntary participation to an experiment on flexibility in engineering design, off regular class hours. Most participants were mature graduate students with training in engineering, science and/or management and many years of industry experience.

**Table 2** Participant demographics

| Group characteristics | Category | Percent (%) |
|---|---|---|
| Age | <25 | 14 |
| | 25–34 | 67 |
| | >35 | 19 |
| Highest education level | Bachelor | 49 |
| | Master | 49 |
| | PhD | 2 |
| Gender | Female | 19 |
| | Male | 81 |
| Work experience (years) | <5 years | 36 |
| | 5–9 years | 39 |
| | >10 years | 25 |

## 4.3 Design problem

A large-scale multi-family residential design problem was selected from the real estate sector. Engineering systems like this typically show high levels of technical complexity, social intricacy, and elaborate processes fulfilling important functions for society (ESD 2011). Large-scale real estate housing projects require technical expertise not only in engineering, but also in management of human assets for developing, financing, and maintaining such infrastructures given prevailing social and market forces. They are central to the development and planning of new cities. Their design and finance are intricate components of the world economy (HBS 2011; UN 2008). They face significant socio-technical uncertainties in demand, prices, materials/construction costs, technology, and regulations. Designing these systems for uncertainty and flexibility is a looking forward, bottom-up approach to dealing with these concerns.

Many of the flexibility principles applicable to this engineering system also apply to more complex systems. For instance, the phasing and capacity expansion strategies explored for the HCSC building (Guma et al. 2009) were also explored by de Weck et al. (2004) for the Iridium system. Given this multi-family residential design problem can capture many different kinds of flexibility strategies, it represents a good platform for testing the proposed procedures in an experimental setting.

The design problem consisted of developing and deploying 309 units of a multi-family residential development project over three phases (1 year/phase). Participants could decide what kind of unit to develop in each phase—either condo (short for condominium) or apartment. They could think about the number of units to deploy in each phase and whether to develop/sell these units in each phase. The distinction between unit types was explained as different levels of quality, prices, and construction costs. A condo is typically more luxurious, built from expensive materials, and targets business professionals. Sales price and construction costs are typically higher. An apartment unit is functional, less luxurious, and for students and middle-class families. The sales price and construction costs are lower. Design details, market assumptions, and preliminary benchmark NPV analysis shown to participants are available online.[1]

Even though the problem was simplified, there was ample room for creativity about the engineering design and development. Participants could discuss the unit type and allocation in each phase, the phasing strategy (e.g., horizontal vs. vertical), the engineering impact on the infrastructure (e.g., buying land versus building a stronger structure for vertical elevation), what infrastructures to share between units and buildings (e.g., electricity, water, heating systems, ventilation), what materials to use, etc. Different strategies would produce different cash flows in each phase and different NPVs depending on market conditions, as explained in Sect. 4.6.

The benchmark design was set as a condo-only residential project, with all 309 units developed in the first year (i.e., phase 1). There was no specification how this design was achieved (i.e., how many buildings and units per building, etc.), to minimize design fixation. This choice was justified by providing the highest NPV between a condo-only and an apartment-only project, based on deterministic market projections. This choice represents current best practices in the real estate industry (Geltner et al. 2010).

## 4.4 Experimental manipulations

### *4.4.1 Sessions structure*

Figure 1 shows the structure of each session. The moderator welcomed participants and described the design problem and context.[1] Teams were told to represent an internal consulting firm at a renowned multi-family residential real estate firm. Participants received a short training on free undirected brainstorming, described as an ideation method that encourages idea quantity, restrains criticism, welcomes unusual ideas, and brings new ideas by combining and building upon existing ones. Participants were also trained on how to use the GSS technology, introduced as an approach to stimulate creativity and enable efficient idea transcription. The task was assigned to brainstorm and suggest alternative design and development plans improving anticipated lifecycle performance compared to the benchmark design. There was no emphasis to complete this task by means of flexibility, although participants were aware that the experiment was about this topic.

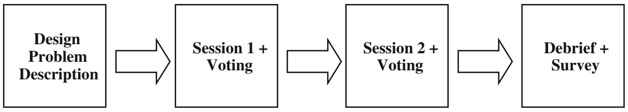


**Fig. 1** Pretest–posttest experimental structure used in experiments

Participants were given 25 min to generate concepts in session 1 under the procedures $E = -1$ and $I = -1$. Participants then had 5 min to vote on design concept quality, using a 1 (low) to 10 (high) Likert scale. When two seemingly opposite concepts were suggested, quality scores were used in coding analysis to discriminate between them.

Participants repeated the exercise in session 2 for 25 min under one of the treatments in Table 1 and then followed by another 5 min for voting. Treatments 1 and 3 used free undirected brainstorming. Treatments 2 and 4 used the prompting mechanism[1], allocating 5 min for discussions on each question. In treatments 3 and 4 only, the short lecture was presented before the session began. A debrief explained the purpose of the study after all experiments. Demographics as well as subjective impressions of satisfaction with the process, results, and quality of results were collected based on a questionnaire validated by Briggs et al. (2006).[1]

### *4.4.2 Control conditions*

Creativity: Each experiment was structured following a pretest–posttest design (Campbell and Stanley 1966) (Fig. 1), providing control safeguard for an inherent creativity variable. It was possible that some teams would be naturally more creative than others, thus making signal measurements more difficult compared to noise. Creative teams generating more ideas would contribute toward larger within-group variability (i.e., the noise or unexplained variability). This could wash out between-group variability (i.e., the signal or explained variability), thus making a small signal more difficult to detect compared to noise.

Flexibility training: Participants were screened to control for prior knowledge about flexibility. They could not participate if they attended a course on this topic within the last 5 years. This control ensured that participants would not have a biased view before experiments. This could have biased their contributions compared to other participants who did not have such training some time before experiments.

Location: All experiments were conducted in the same room to control for location effects, which may offer different lighting conditions, noise levels, etc. The large conference room had capacity for fifty people sitting around a U-shaped table and had a screen for computer projection. A different room was used three or four times due to logistical issues. At most nine people (three teams) could participate in an experiment simultaneously, although most often one or two teams participated together.

Procedure repeat: Providing the exact same content in all activities (i.e., introductions, training, task definitions, lectures, prompting sessions, debriefs, and surveys) controlled for information variability. Giving different information or formulating questions differently could have biased participants' views on the design problem (Morgan and Henrion 1990).

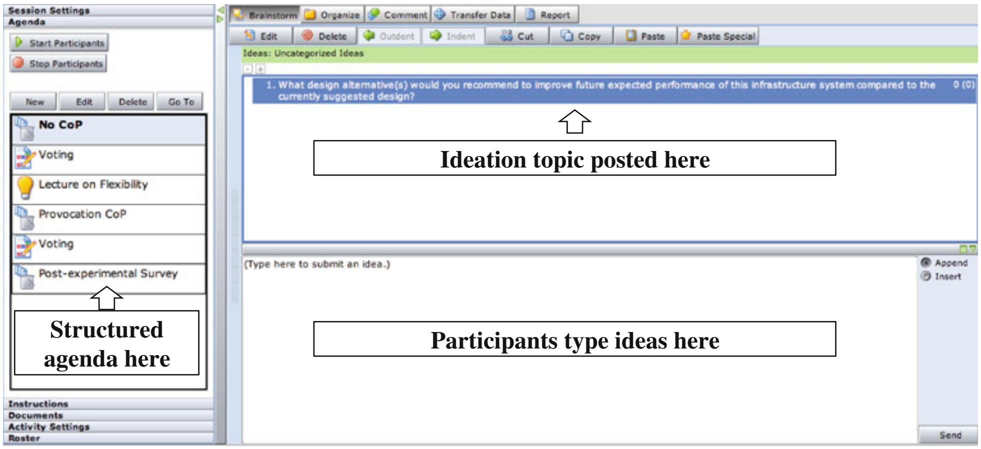
Team size: The same number of three participants was used in each experiment to control for the effect that team size might have on the creative process. A few last-minute cancellations forced six teams of two participants.

Time: The same time was allocated for each activity to control for possible effects on concept quantity and quality. Time possibly contributed to the non-zero baseline response observed for all dependent variables under the general linear model (GLM) (see the non-zero $\beta_0$ coefficients in Sects. 5.1–5.6). Modeling responses using GLM quantified this response separately from the main and interaction effects (coefficients $\beta_E$, $\beta_I$, and $\beta_{EI}$). This discriminated between the effects caused by giving considerable time to participants to think about the problem and the effects caused by the factors of interest $E$ and $I$.

## 4.5 Data collection

### *4.5.1 Online GSS interface*

GroupSystems' ThinkTank® online software was used as GSS technology (Fig. 2). It is an easy-to-use interface that enabled participants to type in real-time descriptions of their design solutions, similar to that of chatting software. It worked with any standard Internet browser. After the problem description, the moderator posted the ideation task of improving performance compared to the benchmark design. Each member described their solutions, which were displayed to other members to stimulate creativity and engage discussions. Each member could reply, comment, and append new ideas to a thread. The software also provided the voting module to rate concept quality. The interface provided an easy and efficient way to record participants' creative responses to the design problem—instead of manual transcription. The online feature allowed a few distance students to participate.



1. What design alternative(s) would you recommend to improve future expected performance of this infrastructure system compared to the currently suggested design?

(Type here to submit an idea.)

Ideation topic posted here

Structured agenda here

Participants type ideas here

**Fig. 2** Online GSS interface (adapted from ThinkTank® by GroupSystems)

### *4.5.2 Raw data description*

Raw data consisted of written descriptions of the design and development plans in a Word® document, quality scores for each idea/concept, and online survey results collected using LimeSurvey®. The GSS software produced the transcript automatically at the end of each session, ready for coding analysis. In treatments 2 and 4, written descriptions included answers to the specific prompts. These data were analogous to the raw interview data obtained after interviewing engineers, managers, and/or decision makers in car manufacturing and offshore oil platform case studies (Suh et al. 2007; Kalligeros 2006).

## 4.6 Computer model

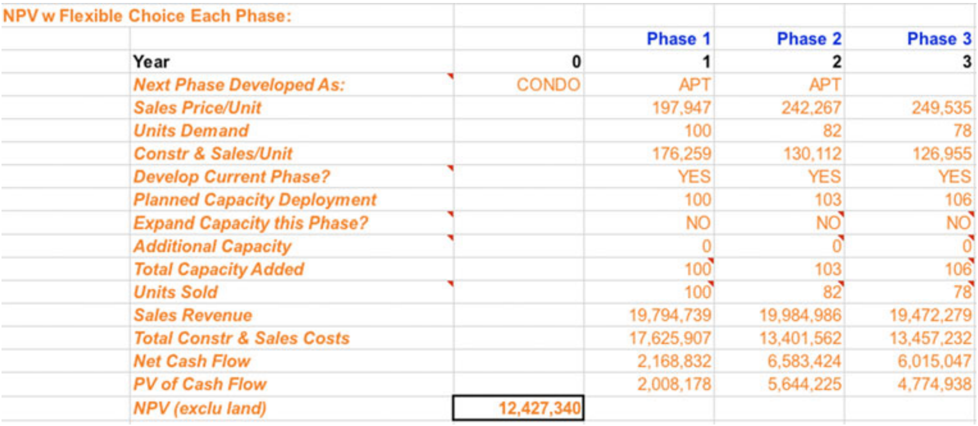
Figure 3 shows the computer-based Excel® DCF model used to quantify anticipated lifecycle performance in economic terms. A risk-neutral expected NPV (ENPV)—or average NPV—metric was used because it balances risk-seeking and risk-averse design decisions. Other metrics could have been used to suit different risk profiles and needs—for example, initial cost, 5th or 95th percentile response, standard deviation.

DCF valuation was motivated by the need to quantify the value of flexibility. Also, NPV is often used for large-scale engineering project evaluation (de Neufville and Scholtes 2011; Geltner et al. 2010). Excel® was chosen because it is transparent, ubiquitous, and a good communication tool for designers and decision makers.

Assumptions about the design, development plan, engineering, and market conditions are available online.[1] The effects of design and development plans were quantified at the conceptual phase using this model, by considering the revenues and costs generated in the future. No detailed embodiment of the concepts was necessary. Design and engineering trade-offs were in terms of the unit type (i.e., condo vs. apartment), deployment strategy (staged vs. all at once vs. deferred), and unit capacity allocation in each phase. These decisions affected the cash flows and NPV generated, discriminating between different design alternatives. For example, the decision to select condo versus apartments could affect the sales price and construction cost, as they were both higher for condo units than apartments. Also, discounting cash flows would imply that timing and unit allocation in each phase mattered from a managerial standpoint. Later cash flows would be more heavily discounted in the model and weigh less in the NPV. These decisions had engineering and cost implications affecting later phases of the design process, modeled and evaluated at the conceptual level.

### *4.6.1 Notation*

| | |
|---|---|
| $C_t$ | Total construction and sales cost at time $t$ |
| $\mathrm{CC}_t^S$ | Stochastic construction cost at time $t$ |
| $\mathrm{CF}_t$ | Cash flow at time $t$ |
| $\mathrm{d}Z_t$ | Standard Wiener process random variable at time $t$ |
| $D_t^S$ | Stochastic unit demand at time $t$ |
| ENPV | Expected net present value |
| $g_P$ | Projected annual growth rate for unit price |
| $g_{Pt}^S$ | Stochastic growth rate for unit price at time $t$ |
| $K_t$ | Planned capacity deployment at time $t$ |
| $K_t^{\mathrm{Tot}}$ | Total capacity deployment at time $t$ |
| $M$ | Maximum number of samples in Monte Carlo simulation |
| NPV | Net present value |
| $P_t$ | Unit price at time $t$ |
| $P_t^S$ | Stochastic unit price at time $t$ |
| $\mathrm{PV}_t$ | Present value of cash flow at time $t$ |
| $r$ | Discount rate or opportunity cost of capital |
| $R_t$ | Revenue at time $t$ |
| $\sigma_P$ | Uncertainty factor around annual unit price projections |
| $T$ | Maximum time value $t$ |
| $U_t$ | Number of units sold at time t |



| NPV w Flexible Choice Each Phase: | | | | |
|---|---|---|---|---|
| | | Phase 1 | Phase 2 | Phase 3 |
| Year | 0 | 1 | 2 | 3 |
| *Next Phase Developed As:* | CONDO | APT | APT | |
| *Sales Price/Unit* | | 197,947 | 242,267 | 249,535 |
| *Units Demand* | | 100 | 82 | 78 |
| *Constr & Sales/Unit* | | 176,259 | 130,112 | 126,955 |
| *Develop Current Phase?* | | YES | YES | YES |
| *Planned Capacity Deployment* | | 100 | 103 | 106 |
| *Expand Capacity this Phase?* | | NO | NO | NO |
| *Additional Capacity* | | 0 | 0 | 0 |
| *Total Capacity Added* | | 100 | 103 | 106 |
| *Units Sold* | | 100 | 82 | 78 |
| *Sales Revenue* | | 19,794,739 | 19,984,986 | 19,472,279 |
| *Total Constr & Sales Costs* | | 17,625,907 | 13,401,562 | 13,457,232 |
| *Net Cash Flow* | | 2,168,832 | 6,583,424 | 6,015,047 |
| *PV of Cash Flow* | | 2,008,178 | 5,644,225 | 4,774,938 |
| *NPV (exclu land)* | 12,427,340 | | | |

**Fig. 3** NPV model for the real estate development design problem

### *4.6.2 Mathematical formulation*

The DCF model in Fig. 3 is explained row by row from top to bottom. Each row is a vector $\boldsymbol{X} = [X_0, \ldots, X_T]$, where $X$ takes on numerical values for the variables above, or textual values, and $t = 0, 1, \ldots, 3$ years. The row "Next Phase Developed As": specified the unit type developed next phase, either "CONDO" or "APT." The decision rule for a *switching* flexibility could be implemented here. An example decision rule elicited by participants was "if current cash flows are higher for condos than for apartments, then develop next phase as condo, else develop next phase as apartments." Following Excel®'s programming language = *IF(logical test, value if true, value if false)*, this could be implemented as:

$$= \mathrm{IF}(\mathrm{CF}_{\mathrm{t-condo}} > \mathrm{CF}_{\mathrm{t-apt}}, \text{``CONDO''}, \text{``APT''})$$

The rows "Sales Price/Unit," "Units Demand," and "Construction & Sales/Unit" were modeled as random variables. The price, demand, and cost values were associated with the choices of unit developed as per the decision rule above. The row "Planned Capacity Deployment" represented the planned capacity deployment under the benchmark $\boldsymbol{K} = [309, 0, 0]$. Teams could modify the planned deployment to explore other phasing strategies over time (e.g., $\boldsymbol{K} = [100, 103, 106]$). Stochastic evolution of the random variables was modeled using geometric Brownian motion (GBM)[3]:

$$P_1^S = P_1 \sigma_P \mathrm{d}Z_1$$
$$g_{Pt}^S = g_P \mathrm{d}t + \sigma_P \mathrm{d}Z_t$$
$$P_t^S = P_{t-1}^S \left(1 + g_{Pt}^S\right)$$

The growth parameter $g_{Pt}^S$ was modeled according to the standard Itô's lemma (1951). The parameter $\sigma_P$ represented the uncertainty around annual price projections and d$t$ a small time increment of one period (e.g., $g_P = 3\,\%$, $\mathrm{d}t = 1$ year, $\sigma_P = 20\,\%$). The random variable $\mathrm{d}Z_t$ captured the standard Wiener variable modeling the stochastic error at time $t$ around the projected growth rate $g_P$. For simplification and computational efficiency, $\mathrm{d}Z_t$ was sampled from a uniform distribution $\sim U(-1,\ 1)$ instead of a standard normal distribution $\sim N(0,\ 1)$. Stochastic unit demand ($D_t^S$) and construction cost ($\mathrm{CC}_t^S$) random variables were modeled in a similar fashion.

The row "Develop Current Phase?" determined whether a phase would be developed or not. The values "YES" or "NO" could represent *deferral* or *abandonment* flexibility strategies if market conditions were unfavorable. A decision rule elicited was "if total construction cost per unit was lower than sales price in the current phase, develop (i.e., print "YES"), else do not develop (i.e., print "NO")." This could be implemented as:

$$= \mathrm{IF}(\mathrm{CC}_t^S < P_t^S, \text{``YES''}, \text{``NO''})$$

The row "Expand Capacity this Phase?" adjusted the number of units compared to planned capacity deployment. The values could be either "YES" or "NO," representing capacity *expansion*, *reduction,* or "just in time." For example, "if demand was higher than planned capacity, add more units to match exactly observed demand (i.e., print "YES"), else build according to planned capacity (i.e., print "NO")":

$$= \mathrm{IF}(D_t^S > K_t, \text{``YES''}, \text{``NO''})$$

The row "Additional Capacity" determined unit allocation in each phase. The decision rule above could be implemented as:

$$= \mathrm{IF}\left(\text{Expansion value} = \text{``YES''},\ K_t^{\mathrm{Tot}} = D_t^S, K_t^{\mathrm{Tot}} = K_t\right)$$

"Total Capacity Added" accounted for the planned capacity deployment, plus any additional unit added or removed within a phase. The row "Units Sold" determined how many units were sold within each phase. It was the minimum between total existing capacity $K_t^{\mathrm{Tot}}$ and demand $D_t^S$:

---

[3] $t = 2, 3$ for the last two equations.

$$U_t = \mathrm{MIN}\left(D_t^S, K_t^{\mathrm{Tot}}\right)$$

The sales revenue ($R_t$), total construction and sales cost ($C_t$), net cash flow ($\mathrm{CF}_t$), present value of cash flow ($\mathrm{PV}_t$), and "NPV (excluding land cost)" were calculated as:

$$R_t = U_t P_t^S$$

$$C_t = K_t^{\mathrm{Tot}} \mathrm{CC}_t^S$$

$$\mathrm{CF}_t = R_t - C_t$$

$$\mathrm{PV}_t = \mathrm{CF}_t/(1+r)^t$$

$$\mathrm{NPV} = \sum_{t=0}^{T} \mathrm{PV}_t$$

#### 4.6.3 Anticipated performance measurements

Anticipated lifecycle performance was measured using ENPV for each flexible design concept by simulating stochastically $M = 2{,}000$ combined scenarios of price, demand, and cost. Each scenario combination produced different cash flows based on the flexibility strategy and decision rules implemented, and one NPV measurement. ENPV was calculated as:

$$\mathrm{ENPV} = E[\mathrm{NPV}] = \frac{1}{M}\sum_{m=1}^{M} \mathrm{NPV}_m$$

Each run took ~1 s on a standard 2.4 GHz Intel Core 2 Duo MacBook with 4 GB of RAM, running Excel® 2004 on Mac OS X version 10.6.

### 4.7 Analysis

#### 4.7.1 Coding analysis

Ideation transcripts were analyzed to extract *complete* flexible design concepts using a well-established coding procedure described by Strauss and Corbin (1990). A design concept was considered *complete* if it contained coherent information about the following elements (using the switching example):

1. An *uncertainty source* affecting anticipated performance.
    - For example, unit demand.
2. A *flexible strategy* to adapt to the above uncertainties in design and operations.
    - For example, switch between condo and apartments.
3. A conceptual but concrete description of the *flexibility enabler*, considering engineering design, legal, management, and/or financial aspects.
    - For example, design each unit as empty shells to be finished later as condo or apartments.
4. A *decision rule*, "if" statement, or "trigger mechanism" based on observations of the uncertainty sources, determining when it is appropriate to exercise the flexibilities.
    - For example, if demand is higher for apartments than condos, switch to finishing and selling units as apartments, if not finish and sell as condos.

The switching strategy above contrasts with the benchmark inflexible plan where all units are developed at once as condos. It requires developing units as empty shells to be finished later, different than designing all units as condos. It is not clear at the conceptual stage what design and decision rules are most profitable given the uncertainties, hence the need for explicit modeling.

Other examples of complete concepts from ideation transcripts are available.[1] They exemplify strategies evaluated using the computer model: phase the development and deploy capacity over time, expand or reduce unit capacity in each phase whenever appropriate, temporarily abandon the project if market conditions were not suitable, and do not develop a phase if market conditions are unfavorable.

Two independent treatment blind coders reviewed each ideation transcript in a randomized order to extract and count complete concepts, with 95 % average inter-rater agreement. The inter-rater agreement was the average percentage agreement between raters on all thirty-two ideation transcripts. Concepts retained for implementation, evaluation, and statistical analysis, were the ones agreed upon by both reviewers.

#### 4.7.2 Dependent variables

The null hypothesis of no main and interaction effect of factors $E$ and $I$ was tested on the following dependent variables:

1. Quantity of *complete* flexible design concepts generated ($C$);
2. Quantity of *good* flexible design concepts generated ($G$);
3. Anticipated lifecycle performance of flexible design concepts (ENPV);
4. Subjective impressions of satisfaction with the procedures/processes (PS);
5. Subjective impression of satisfaction with the results (RS);
6. Subjective anticipated quality of results (i.e., quality assessment) (QA).

A response $\Delta y = y2 - y1$ was measured for each experiment, where $y1$ is the response of interest in session 1 only and $y2$ is the response of interest in both sessions combined. For instance, if one complete concept was generated in session 1 ($C1 = 1$) and two *new* concepts were generated in session 2 ($C2 = 3$), then $\Delta C = 3 - 1 = 2$.

Each complete concept was implemented using the computer model. Partially complete concepts were coded but not implemented because they did not provide enough information for computer evaluation. A complete concept was good *if and only if* the ENPV was higher than the benchmark under simulations (i.e., ENPV = \$9.3 million). For example, if one good design concept was generated in session 1 ($G1 = 1$), but only one of the two new complete concepts was higher than \$9.3 million ($G2 = 2$), then $\Delta G = 2 - 1 = 1$.

Only the best combinations of good flexible design concepts producing the highest ENPV values were considered ΔENPV measurements. Because some concepts could interact (e.g., early abandonment prevents future capacity expansion), it was not possible to simply add up the ENPV obtained for each independent concept. It was necessary to run the simulations with each possible combination of concepts generated over sessions 1 and 2. For example, in session 1 of a particular experiment, no complete concept was generated (ENPV1 = \$0 as compared to the benchmark). A switching strategy in session 2 generated value \$10.5 million (ENPV2 = \$10.5 − 9.3\$ = \$1.2 million). This led to measurement ΔENPV = \$1.2 − \$0 = \$1.2 million. In another experiment, a switching concept led to ENPV1 = \$3.0 million. Two new concepts generated in session 2 and combined with switching led to ENPV2 = \$3.9 million. Therefore, ΔENPV = \$0.9 million.

### 4.7.3 Survey analysis

Survey responses were analyzed to measure improvements in ΔPS, ΔRS, and ΔQA. Responses recorded the differences in user impressions between sessions 1 and 2, using a discrete 7-scale Likert mechanism. Each construct was evaluated using five or six questions (maximum score 35 or 42). A positive (negative) score meant improvement (worsening) from session 1 to 2. For example, an individual scoring PS1 = 27/35 and PS2 = 28/35 in session 2 meant ΔPS = +1. Responses ΔRS and ΔQA were measured similarly. Survey questions are available online.[1]

### 4.7.4 Statistical analysis

Each response $\Delta y$ was modeled based on the GLM. Coefficient $\beta_0$ approximated the total (i.e., baseline) mean response, $\beta_E$ and $\beta_I$ modeled the main effects, while $\beta_{EI}$ modeled the first-order two-way interaction between factors $E$ and $I$. Variable $\varepsilon$ accounted for the mean experimental error:

$$\Delta y = \beta_0 + \beta_E E + \beta_I I + \beta_{EI} EI + \varepsilon$$

Standard least square regression was used to calculate the main and interaction effects, as well as $p$ values. The null hypothesis was $H_0$: $\beta_E = \beta_I = \beta_{EI} = 0$. The $p$ values of the main and interaction effects were calculated using a nonparametric permutations test (i.e., randomization or exact test) (Welch 1990). Because $\Delta y \geq 0$ for $\Delta C$, $\Delta G$, and ΔENPV, sample distributions were truncated about zero. This could not satisfy normality assumptions for standard parametric tests. The permutations algorithm was programmed using basic MATLAB® functions, including the “regress” function.[4] All statistical results were corroborated using Excel®’s data analysis toolkit.

## 5 Results and discussion

### 5.1 Improvement in complete concepts ($\Delta C$)

Figure 4 shows the mean plots for $\Delta C$ under four treatment conditions. The ideation mechanism ($I$) produced a significant main effect ($\beta_I = 0.75$, $p = 0.00$). This means that prompting helped participants generate more new complete concepts after two sessions than free undirected brainstorming.[5] It shows that the form of the questions and the concepts referred to in prompting—even though fairly abstract—were useful and effective. The GLM response was:

$$\Delta C = 1.25 + 0.25E + 0.75I$$

For example, treatment 1 generated an improvement of $\Delta C = 2$ concepts among eight teams/replicates. This corresponds to the mean value 0.25 in Fig. 4 and to the value obtained in the GLM equation if one inserts variable value $E = -1$ and $I = -1$. The mean values for the remaining treatments were calculated accordingly.

Current flexibility training and free undirected brainstorming produced the lowest mean response $\Delta C = 0.25$. This shows however that some teams thought naturally and explored flexibility without guidance. This is consistent with the real-world observation that flexibility is not necessarily prevalent in practice, but is sometimes exploited in the real estate industry (Guma et al. 2009).

It is surprising that explicit training did not produce any effect compared to current training. The cognitive network model of creativity by Santanen et al. (2004) suggests that explicit training might have caused information overload,

---

[4] The permutations code is available in Appendix K of Cardin (2011).

[5] The positive sign implied prompting was responsible for the main effect, while a negative sign implied that free undirected brainstorming was responsible for the effect.

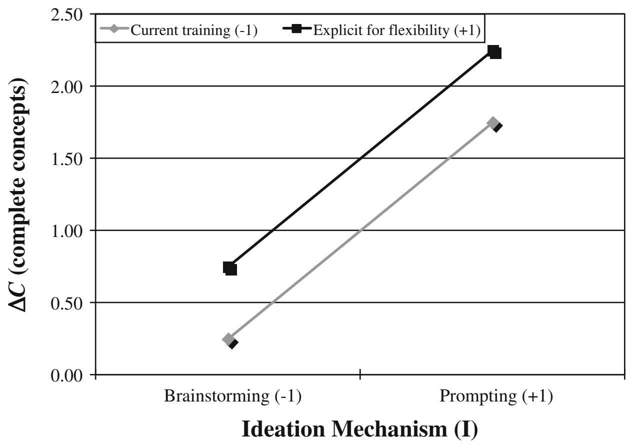


**Fig. 4** Mean plots for $\Delta C$ (complete concepts). The *lower curve* shows mean results for current training on flexibility ($E = -1$). The *upper curve* shows results for explicit training on flexibility ($E = +1$)

which reduced positive effects on creativity. It is also possible that information was not presented in a retention-maximizing sequence. In contrast, prompting stimulated frame activation consistently, helped structure the thought process, and guided participants throughout each session.

A crucial argument put forward here is that complete concepts can improve anticipated lifecycle performance. This section demonstrated that prompting effectively helps designers generate more complete concepts. The following two subsections demonstrate that these concepts in turn helped improve lifecycle performance.

### 5.2 Improvement in good concepts ($\Delta G$)

Figure 5 shows that prompting ($I = +1$) also had a main effect on $\Delta G$ ($\beta_I = 0.59$, $p = 0.00$). The mean values were typically lower than for $\Delta C$, since a few complete concepts could not be counted as good (i.e., ENPV < \$9.3 million). The GLM response was modeled as:

$$\Delta G = 1.09 + 0.22E + 0.59I - 0.03EI$$

A complete concept did not improve performance when the decision rule was too conservative, or value destructive. In the former case, the flexibility strategy would never be exercised, so the infrastructure would behave the same as the inflexible benchmark design. In the latter, the ENPV would be lower than the benchmark's. For example, a team suggested deferring development if market demand went 80 % below projections. Given the problem constraints, this would never occur ($\sigma_D = 20$ %). Another team suggested expanding capacity when construction costs would increase, inevitably leading to value destruction. These examples showed the importance of implementing the decision rules using the computer model. Relying purely on qualitative judgment for concept evaluation would have made more difficult catching such conceptual flaws.

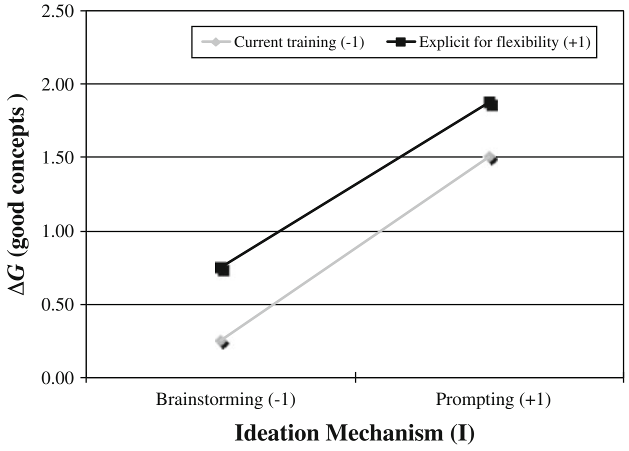


**Fig. 5** Mean plots for $\Delta G$ (good complete concepts)

Prompting was crafted around the four criteria of a complete concept. It seems natural that it helped generate more flexible design concepts than any other procedure. The nuance, however, is that more concepts did not guarantee better cash flows and improved ENPV. The cash flows generated by the flexible strategies had to be better on average than that of the benchmark, which could only be evaluated quantitatively via computer modeling.

### 5.3 Improvement in ENPV (ΔENPV)

Figure 6 shows that prompting had a significant main effect on ΔENPV ($\beta_I = 0.98$, $p = 0.00$). The procedure improved anticipated lifecycle performance compared to the benchmark by up to 36 % percent (i.e., \$3.34 million/ \$9.30 million) in combination with the lecture. These results are consistent with those in the real options literature (Trigeorgis 1996; de Neufville and Scholtes 2011). The GLM response was:

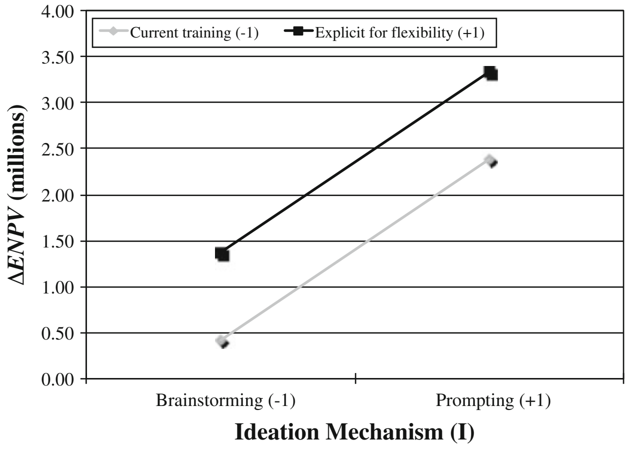


**Fig. 6** Mean plots for ΔENPV (anticipated lifecycle performance)

$\Delta\text{ENPV} = 1.88 + 0.48E + 0.98I$

Flexibility improved expected lifecycle performance by acting explicitly on the distribution of outcomes. For example, capturing whichever condo or apartment markets with highest demand helped capitalize on better profit opportunities. Deferring the first phase until favorable market conditions emerged reduced the impact of loss-generating scenarios. Expanding unit capacity within phases increased profits, while reducing it avoided spending resources on units that may not sell.

The following observations are at the heart of this paper. Results show that the procedures suggested here may help designers effectively improve anticipated lifecycle performance by means of flexibility. They show that current engineering training and free undirected brainstorming may help, but may not be sufficient to generate valuable flexibility. Such approaches may leave aside significantly valuable design alternatives. Participants under treatment 1 ($E = -1,\ I = -1$) generated $\Delta\text{ENPV} = \$0.41$ million under current training and free undirected brainstorming. This is a ~5 % improvement compared to the \$9.3 million benchmark ENPV, a significant amount for a real estate project. Treatment 4 ($E = +1,\ I = +1$), however, produced $\Delta\text{ENPV} = \$3.34$ million, about 36 % improvement. This shows there is an opportunity to improve current training and widely used concept generation procedures by integrating a simple, efficient, and intuitive lecture/prompting package as suggested.

### 5.4 Improvement in process satisfaction (ΔPS)

Figure 7 shows that explicit training on flexibility had a significant main effect on ΔPS ($\beta_E = 1.35, p = 0.06$). This suggests that participants receiving the lecture were significantly more satisfied with the process than participants who did not. There was also a significant interaction ($\beta_{EI} = 1.37, p = 0.05$) where participants using prompting were significantly more satisfied with the process when they received the lecture. There was almost no satisfaction improvement when prompting alone was used. The GLM response was given by:

$\Delta\text{PS} = 2.40 + 1.35E + 0.28I + 1.37EI$

These findings demonstrate that the lecture was crucial to the acceptability of the prompting procedure. The interaction effect also supports this view. Prompting alone led to almost zero ΔPS improvement, clearly below the results from other treatments. Perhaps because participants were being taught a new skill, they did not see the purpose of the prompts. This observation brings empirical support to an observation often made in GSS research, but not well documented. Participants will not appreciate an intervention—and sometimes refuse to participate—if they are not explained the rationale behind a given procedure.

### 5.5 Improvement in results satisfaction (ΔRS)

Results in Fig. 8 show that explicit training on flexibility ($E = +1$) had a significant main effect on ΔRS ($\beta_E = 2.33,\ p = 0.00$). Participants were more satisfied with the results after the short lecture on flexibility:

$\Delta\text{RS} = 4.19 + 2.33E - 0.38I + 1.08EI$

The results concur with those for process satisfaction, although there was no significant interaction effect here. Using prompting without the lecture also produced near-zero improvement in results satisfaction. Participants were significantly more satisfied with prompting after the lecture, perhaps because they could appreciate more the benefits of flexibility.

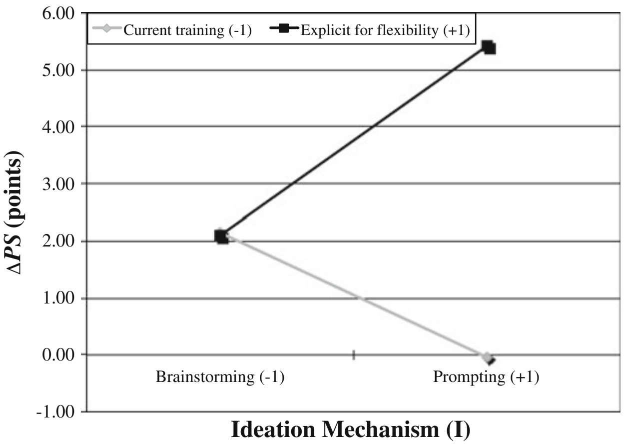


**Fig. 7** Mean plots for ΔPS (process satisfaction)

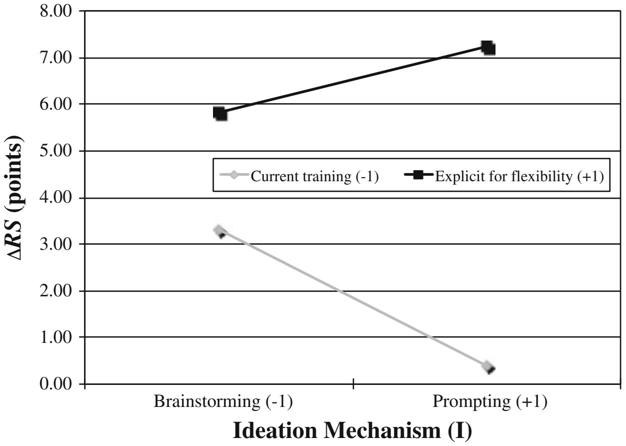


**Fig. 8** Mean plots for ΔRS (results satisfaction)

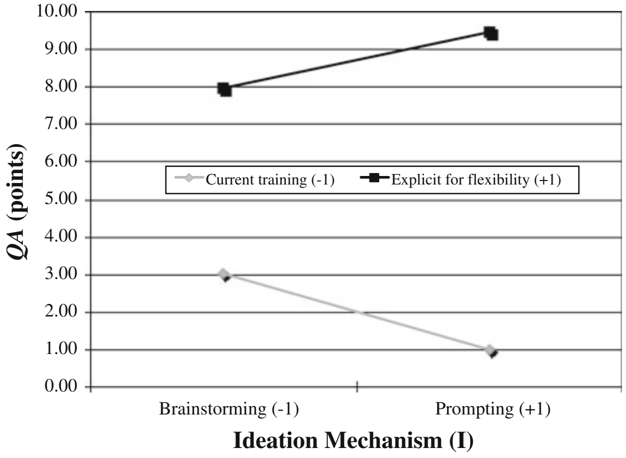


**Fig. 9** Mean plots for ΔQA (results quality assessment)

## 5.6 Improvement in quality assessment (ΔQA)

Figure 9 shows a significant main effect of explicit training on ΔQA ($\beta_E = 3.34$, $p = 0.00$).[6] Participants expected better quality of results after the lecture on flexibility. This may be because they felt more committed or believed more in the quality of results once exposed to the ideas of flexibility. The GLM response was given by:

$$\Delta\text{QA} = 5.36 + 3.34E - 0.14I + 0.89EI$$

Even if prompting demonstrably improved anticipated lifecycle performance, it is interesting that participants did not think it would generate better results quality. This may be because they could not measure in real-time the NPV impact of their ideas, since all concepts were evaluated after experiments. It could also be because they had no clear quality criteria in mind when they evaluated their own ideas, while the group who received the lecture did.

## 5.7 Remarks

The results above suggest that evaluating procedures based solely on quantitative performance or qualitative user impressions may not highlight all the strengths and weaknesses of procedures supporting design activities. One procedure may very well improve lifecycle performance, but may be too cumbersome for use in practice. Equally, a procedure may generate user satisfaction, but not improve lifecycle performance. Empirical studies based on both criteria should provide better grounds to identify promising procedures for dissemination in industry and engineering education. This is in line with the call for more empirical studies of design procedures by Frey and Dym (2006) and Reich (2010).

## 5.8 Results validity and study limitations

### 5.8.1 *Internal validity*

One threat to internal validity of results is whether prompting was a valid design procedure, as it may seem like answers were given away to participants. Here, researchers knew *of* flexible strategies potentially improving performance, but did not know *of all* the possible strategies participants could generate. The design problem left enough room for creativity, and participants identified strategies researchers did not think of. One team suggested developing units "just in time," which these authors had not considered. On the other hand, it was necessary to create and implement some flexible design solutions a priori to test and validate the computer model. This did not imply that all possible flexible design concepts were identified and packaged implicitly in the prompting mechanism. The researchers did not know ahead of time what strategies, design enablers, and decision rules participants would formulate. In addition, experimental results presented here are only valid for risk-neutral design decision making. Using a different performance metric other than ENPV may lead to different set of results and conclusions altogether.

The prompting mechanism was crafted to be general enough to be usable directly for a different engineering system in a different domain. Interested readers may consult the actual used as posted online.[1] Prompting provided some level of direction, but not complete direction as to give away answers. Figure 10 depicts conceptually a spectrum of "amount of direction," ranging from no direction at all—for example Santanen et al. (2004)—to complete direction—for example Santanen and de Vreede (2004). This prompting mechanism lies somewhere in-between, closer to complete direction. Prompting challenged the underlying assumptions of the design problem and induced participants to consider alternatives they may not have otherwise considered. This is in line with Knoll and Horton's (2010) change of perspective approach of stimulating creativity.

In terms of measuring the effects of training, current training about flexibility was favored over placebo training

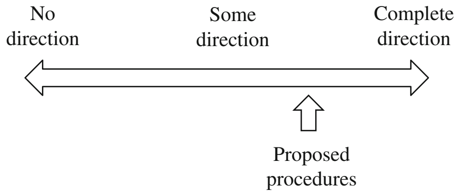


**Fig. 10** "Amount of direction" spectrum for prompting-based ideation mechanisms

---

[6] Results refer to the anticipated quality of overall results obtained from the validated survey.[1] They do not refer to the quality of individual concepts scored during voting sessions.

(i.e., training with general content minimally influencing knowledge) because a) it is difficult to determine the nature of what general training is in this context and b) it could have disengaged participants if they realized there was no particular message intended in a placebo lecture. Given the idea of placebo involves some notions of deception, it would have been difficult to provide any training without participants thinking about its purpose. Knowing the experiment was about flexibility, they could have wondered about the point of the message, increasing the risk of disengagement, possibly causing the deception to be too obvious and hence losing its purpose.

#### *5.8.2 External validity*

Threats to external validity included the fact that participants were graduate students, as opposed to practicing engineers. The controlled environment did not fully represent the realities of daily industrial activities. Participants' self-selection may have biased results. It may not reflect the natural resistance arising when new procedures are introduced in real-world organizations. Also, the design problem did not capture the full complexity of a real-world engineering system.

Experimental conditions and the sample population nonetheless represented some of the realities of engineering practice. Participants were mature graduate students with many years of experience in industry. They represented the next best sample available, given the difficulty of conducting controlled experiments with practitioners—whose time and availability are limited. They represented a wide array of industries, expertise, and educational backgrounds. This led to cultural and personality clashes in experiments, also seen in practice. Even if the engineering system problem was simplified, it was modeled through close interactions with real estate experts. The benchmark solution represented some of the best practices in this field.

#### *5.8.3 Measurements reliability*

Two aspects contributed to measurement reliability in $\Delta C$ and $\Delta G$. Two independent reviewers analyzed ideation transcripts using the same systematic coding procedure. Transcripts were analyzed in a random order, and coders did not know what treatment from what team they were working on.

The inter-rater agreement showed that flexible concepts were the same for 95 % of all flexible concepts extracted independently. Under the same set of assumptions and decision rules, and for several runs of 2,000 simulations, the computer model generated ENPV values always falling within the same 95 % confidence interval. The same market and stochastic parameter assumptions were used to evaluate all flexible design concepts, to ensure they were all compared on an equivalent basis. Building upon a survey already validated experimentally by Briggs et al. (2006) enhanced response reliability for ΔPS, ΔRS, and ΔQA. Cronbach $\alpha$ values were measured for each session, treatment, and dependent variable. The twenty-four $\alpha$ values measured between 0.91 and 0.99—with maximum possible value $\alpha = 1$—showed that survey items measured reliably the constructs within and across participants (Cortina 1993).

## 6 Conclusion

This paper presented the results of an empirical evaluation of procedures to generate flexibility in engineering systems. The effects of educational training about flexibility (current vs. explicit) and ideation procedures (free undirected brainstorming vs. prompting) were studied. These procedures were chosen because of their relative ease of use and simplicity. Measurements involved the quantity of flexible design concepts generated, anticipated lifecycle performance improvements compared to a benchmark design, user satisfaction with the process and results, and quality of results. Current training assumed that participants may or may not have received training on flexibility. Explicit training consisted of a short lecture on the topic of flexibility. Free undirected brainstorming represented a simple and intuitive approach to generate ideas (Osborn 1957) widely used in US engineering practice (Yang 2007). Prompting stimulated creativity and structured the thought process to generate valuable flexibility. The procedures were evaluated experimentally by having ninety participants tackle a simplified design problem in real estate development—an example of engineering systems (ESD 2011).

Results showed that prompting helped participants generate more valuable flexible design concepts. In turn, the concepts improved anticipated lifecycle performance significantly more than free undirected brainstorming, by up to 36 %. Users, however, did not expect prompting to improve results quality. Providing explicit training on flexibility improved user satisfaction with the process, results, and results quality.

Results show that the proposed lecture/prompting procedure package may be an effective and complementary toolkit to help designers improve lifecycle performance of an engineering system by means of flexibility. They also suggest that current engineering training and a widely used concept generation procedure like free undirected brainstorming may not be sufficient to do this effectively in early design phases. Design alternatives significantly improving lifecycle performance may be left aside using such approaches. Even though some teams generated flexibility using these techniques, the new procedures generated even more flexible concepts, which in turn improved lifecycle performance even more. Also, explicit training on flexibility proved crucial for the acceptability of the prompting procedure. It provided participants with improved satisfaction with the process and results, as well as anticipated quality of results. This is essential for future use in engineering education and industry practice.

This study showed that design procedures should be evaluated using both qualitative user impressions and quantitative performance measurements. A procedure providing good user impressions does not guarantee the design concepts will perform well. Similarly, users may not appreciate a procedure that is too complicated or cumbersome, even if it improves performance. Measuring both types of effects provides better grounds for wider dissemination and real impact.

### 6.1 Future work

Much work remains to develop and evaluate useful and efficient procedures to generate flexibility in engineering systems design. Many opportunities exist by addressing the limitations of this work. The relationships between anticipated and actual lifecycle performance of the concepts should be studied, even though this is difficult for engineering systems with typically long lifecycles (+20 years). A follow-on study may look into the effects of using different approaches to elicit probability distributions and model the main uncertainty sources, or the impact of using different numbers of participants per team. One can validate further the experimental methodology by conducting more experiments using a different design problem, computer model, and sample population. Cultural effects could be studied using the same approach as here, but with a sample population in a different country or culture. Procedures could be evaluated using different performance metrics (e.g., initial cost, 5th or 95th percentile), recognizing the one studied here is not the only possibility.

External validity of results can be improved by evaluating the procedures in a real-world setting by working with industry practitioners. This input may help improve further the prompting and training platforms presented here. For instance, one study could focus on understanding the best sequence for presenting information in the lecture so as to maximize retention (Deese and Kaufman 1957). The procedures could be used to help a major infrastructure company identify flexible design alternatives to a system currently being designed. These alternatives could be modeled and compared explicitly based on costs and lifecycle improvements to the design concepts currently on the table.

**Acknowledgments** The authors would like to thank Rhonda Jordan, Alexander Verbraeck, Maria Yang, and Delft University of Technology students at the Systems Engineering Section for their intellectual support. We are grateful for the financial support provided by the National University of Singapore (NUS) Faculty Research Committee via MOE AcRF Tier 1 grant WBS R-266-000-061-133, the Engineering Systems Division, Systems Design and Management program, and Center for Real Estate at the Massachusetts Institute of Technology (MIT), the MITRE Corporation, the National Science and Engineering Research Council of Canada, the Fonds Québécois de la Recherche sur la Nature et les Technologies, the MIT Portugal Program, the Singapore-MIT International Design Center, the Cambridge-MIT Institute, and the Fondation Desjardins.

## References

Altshuller G (1973) The innovation algorithm. Technical Innovation Center, Worcester

Bahill AT (2012) Diogenes, a process for identifying unintended consequences. Syst Eng 15(3):287–306

Bender B, Blessing LTM (2003) Re-Interpretation of Conceptualization: A Contribution to the Advance of Design Theory. In: Lindemann U (ed) Human behaviour in design: individuals, teams, tools. Springer, Berlin/Heidelberg

Black F, Scholes M (1973) The pricing of options and corporate liabilities. J Polit Econ 81(3):637–654

Boehm B, Gruenbacher P, Briggs RO (2001) Developing groupware for requirements negotiation: lessons learned. IEEE Softw 18(3)

Bostrom RP, Nagasundaram M (1998) Research in creativity and GSS. In: Paper presented at the proceedings of the 31st annual Hawaii international conference on system sciences, Kohala Coast, HI, USA

Briggs RO, Reinig BA, de Vreede G-J (2006) Meeting satisfaction for technology-supported groups: an empirical validation of a goal attainment model. Small Group Res 37(6):585–611

Browning TR (2001) Applying the design structure matrix to system decomposition and integration problems: a review and new directions. IEEE Trans Eng Manag 48(3):292–306

Buchal RO (2002) Sketching and computer-aided conceptual design. Paper presented at the 7th international conference on computer supported cooperative work in design

Campbell DT, Stanley JC (1966) Experimental and quasi-experimental designs for research. Houghton

Cardin M-A (2011) Quantitative performance-based evaluation of a procedure for flexible design concept generation. Doctoral Dissertation in Engineering Systems, Massachusetts Institute of Technology, Cambridge, MA, USA

Cardin M-A, de Neufville R (2009) A direct interaction approach to identify real options ‘in’ large-scale infrastructure systems. Paper presented at the Real Options Conference, Braga (Portugal), Santiago (Spain)

Chulvi V, González-Cruz MC, Mulet E, Aguilar-Zambrano J (2012) Influence of the type of idea-generation method on the creativity of solutions. Res Eng Des (in press)

Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. J Appl Psychol 78:98–104

Cox JC, Ross SA, Rubinstein M (1979) Options pricing: a simplified approach. J Financ Econ 7(3):229–263

Daly SR, Christian JL, Yilmaz S, Seifert CM, Gonzalez R (2011) Teaching design ideation. Paper presented at the 118th ASEE annual conference and exposition

de Neufville R, Scholtes S (2011) Flexibility in engineering design, engineering systems. MIT Press, Cambridge de Vreede G-J, Vogel D, Kolfschoten G, Wien J (2003) Fifteen years of GSS in the field: a comparison across time and national boundaries. Paper presented at the 36th annual Hawaii international conference on system sciences, Big Island, HI, USA

de Weck OL, de Neufville R, Chaize M (2004) Staged deployment of communications satellite constellations in low earth orbit. J Aerosp Comput Inf Commun 1:119–136

Deese J, Kaufman RA (1957) Serial effects in recall of unorganized and sequentially organized verbal material. J Exp Psychol 54(3):180–187

Eckert CM, de Weck OL, Keller R, Clarkson PJ (2009) Engineering change: drivers, sources and approaches in industry. Paper presented at the 17th international conference on engineering design, Stanford, CA, USA

Eppinger SD, Fine CH, Ulrich KT (1990) Interdisciplinary product design education. IEEE Trans Eng Manag 37(4):301–305

ESD (2011) Engineering systems division strategic report. Massachusetts Institute of Technology, Cambridge

Frey DD, Dym CL (2006) Validation of design methods: lessons from medicine. Res Eng Design 17:45–57

Fricke E, Schulz AP (2005) Design for changeability (dfc): principles to enable changes in systems throughout their entire lifecycle. Syst Eng 8(4):342–359

Geltner D, Miller N, Clayton J, Eichholtz P (2010) Commercial real estate analysis and investments, 2nd edn. Cengage Learning Inc., USA

Gil N (2007) On the value of project safeguards: embedding real options in complex products and systems. Res Policy 36:980–999

Guma A, Pearson J, Wittels K, de Neufville R, Geltner D (2009) Vertical phasing as a corporate real estate strategy and development option. J Corp Real Estate 11(3):144–157

HBS (2011) Executive education in real estate management program—new conditions, new leaders. Harvard Business School. http://www.exed.hbs.edu/programs/rem/print.aspx. Accessed 6 Oct 2011

Herder PM, Bruijn HD (2009) System and actor perspectives on sociotechnical systems. IEEE Trans Syst Man Cybern 39(5):981–992

Itô K (1951) On stochastic differential equations. Mem Am Math Soc 4:1–51

Jablonowski C, Wiboonskij-Arphakul C, Neuhold M (2008) Estimating the cost of errors in estimates used during concept selection. SPE Projects Facilities Constr 3:1–6

Janis IL (1972) Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes. Houghton Mifflin Company, Boston

Kalligeros K (2006) Platforms and real options in large-scale engineering systems. Doctoral Dissertation in Engineering Systems, Massachusetts Institute of Technology, Cambridge, MA, USA

Knoll SW, Horton G (2010) Changing the perspective: improving generate thinklets for ideation. Paper presented at the 43rd annual Hawaii international conference on system sciences, Koloa, Kauai, HI, USA

Kolfschoten G, Briggs RO, Duivenvoorde GPJ, de Vreede G-J (2009) Practitioners vs. facilitators a comparison of participant perceptions on success. Paper presented at the 42nd annual Hawaii international conference on system sciences, Big Island, HI, USA

Kurtoglu T, Campbell MI, Linsey JS (2009) An experimental study on the effects of a computational design tool on concept generation. Des Stud 30:676–703

Leibovich M (1999) A dream come back to earth; missteps, shortfalls, glitches have iridium scaling back expectations for its satellite phone service. The Washington Post, May 24

Linsey JS, Tseng I, Fu K, Cagan J, Wood JL, Schunn C (2010) A study of design fixation, its mitigation and perception in engineering design faculty. J Mech Des 132(4):1–12

MacCormack A, Herman K (2001) The rise and fall of iridium. Harvard Business School, Cambridge

Martin MV, Ishii K (2002) Design for variety: developing standardized and modularized product platform architectures. Res Eng Des 13:213–235

Mikaelian T, Nightingale DJ, Rhodes DH, Hastings DE (2011) Real options in enterprise architecture: a holistic mapping of mechanisms and types for uncertainty management. IEEE Trans Eng Manag 54(3):457–470

Minai AA, Braha D, Bar-Yam Y (2006) Complex Engineered Systems: A New Paradigm. In: Minai AA, Braha D, Bar-Yam Y (eds) Complex engineered systems: science meets technology. Springer, Netherlands

Morgan MG, Henrion M (1990) Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, United Kingdom

Mullen B, Johnson C, Salas E (1991) Productivity loss in brainstorming groups: a meta-analytic integration. Basic Appl Soc Psychol 12(1):3–23

Murman EM, Allen T, Bozdogan K, Cutcher-Gershenfeld J, McManus H, Nightingale DJ, Rebentisch E, Shields T, Stahl F, Walton M, Warmkessel J, Weiss S, Widnall S (2002) Lean enterprise value: insights from MIT's lean aerospace initiative. Palgrave, New York

Nelson BA, Wilson JO, Rosen D (2009) Refined metrics for measuring ideation effectiveness. Des Stud 30:737–743

Nembhard HB, Shi L, Aktan M (2005) A real-options-based analysis for supply chain decisions. IIE Trans 37:945–956

Okudan GE, Ming-Chuan C, Schmidt LC, Hernandez NV, Linsey JS (2010) A pilot exploration of systematic ideation methods and tools on design learning. Paper presented at the 9th international conference on information technology based higher education and training

Osborn AF (1957) Applied imagination: principles and procedures of creative thinking, 1st edn. Scribeners and Sons, New York

Reich Y (2010) My method is better! [editorial]. Res Eng Des (April)

Reinig BA, Briggs RO, Nunamaker JF (2007) On the measurement of ideation quality. J Manag Inf Syst 23(4):143–161

Robie BH, Kicher TP, Wolfe D (1992) Factors influencing individual results in concept generation. Paper presented at the 4th international conference on design theory and methodology

Santanen EL, de Vreede GJ (2004) Creative approaches to measuring creativity: comparing the effectiveness of four divergence thinklets. Paper presented at the Hawaii international conference on system sciences, Los Alamitos, CA, USA

Santanen EL, Briggs RO, de Vreede G-J (2004) Causal relationships in creative problem solving: comparing facilitation interventions for ideation. J Manag Inf Syst 20(4):167–197

Sered Y, Reich Y (2006) Standardization and modularization driven by minimizing overall process effort. Comput Aided Des 38:405–416

Shah JJ, Kulkarni SV, Vargas-Hernandez N (2000) Evaluation of idea generation methods for conceptual design: effectiveness metrics and design of experiments. J Mech Des 122:377–384

Shah JJ, Vargas-Hernandez N, Summers JD, Kulkarni S (2001) Collaborative sketching (C-Sketch)—an idea generation technique for engineering design. J Creat Behav 35(3):168–198

Simpson TW (2004) Product platform design and customization: status and promise. Artif Intell Eng Des Anal Manuf 18:3–20. doi:10.1017/S0890060404040028

Strauss A, Corbin J (1990) Basic of qualitative research: grounded theory procedures and techniques. Sage Publications, Newbury Park

Suh ES, de Weck OL, Chang D (2007) Flexible product platforms: framework and case study. Res Eng Des 18:67–89

Trigeorgis L (1996) Real options. MIT Press, Cambridge

UN (2008) The real estate and financial crisis: causes, effects, and impacts on development. United Nations. http://www.un.org/News/Press/docs/2008/note6186.doc.htm. Accessed 6 Oct 2011

van der Lugt R (2002) Brainsketching and how it differs from brainstorming. Creat Innovation Manag 11(1):43–54

Wang T, de Neufville R (2005) Real options 'in' projects. Paper presented at the real options conference, Paris, France

Warr A, O'Neill E (2005) Understanding design as a social creative process. Paper presented at the proceedings of the 5th conference on creativity and cognition, London, UK

Welch WJ (1990) Construction of permutation tests. J Am Stat Assoc 85:693–698

White C, Talley A, Jensen D, Wood K, Szmerekovsky A, Crawford R (2010) From brainstorming to C-sketch to principles of historical innovators: ideation techniques to enhance student creativity. Paper presented at the 2010 ASEE annual conference and exposition

Yang M (2007) Design methods, tools, and outcome measures: a survey of practitioners. Paper presented at the ASME international design engineering technical conferences and computers and information in engineering conference, Las Vegas, NV, USA

Yang M (2009) Observations on concept generation and sketching in engineering design. Res Eng Des 20:1–11

Zhang SX, Babovic V (2012) A real options approach to the design and architecture of water supply systems using innovative water technologies under uncertainty. J Hydrodyn 14(1):13–29