Research in Engineering Design (2006) 17: 45–57
DOI 10.1007/s00163-006-0016-4

ORIGINAL PAPER

Daniel D. Frey · Clive L. Dym

Validation of design methods: lessons from medicine

Received: 17 March 2005 / Revised: 15 January 2006 / Accepted: 10 April 2006 / Published online: 25 May 2006

Abstract This paper discusses the validation of design methods. The challenges and opportunities in validation are illustrated by drawing an analogy to medical research and development. Specific validation practices such as clinical studies and use of models of human disease are discussed, including specific ways to adapt them to engineering design. The implications are explored for three active areas of design research: robust design, axiomatic design, and design decision making. It is argued that medical research and development has highly-developed, well-documented validation methods and that many specific practices such as natural experiments and model-based evaluations can profitably be adapted for use in engineering design research.

Keywords Validation · Design methodology · Robust design · Design decisions

1 Introduction

The validation of design methods is important for the continuing advancement of both design theory and the professional practice of engineering. Researchers in design theory require validation processes to guide the development and evaluation of new methods. Professional practitioners need validation processes to determine which methods to employ, as well as when and how to employ them. The latter is especially important in large, complex organizations such as automobile and airplane manufacturers and in their government counterparts (e.g., NASA). Early-stage design decisions are characterized by uncertainty and ambiguity and, especially in large organizations, later design decisions are made through processes involving teams with a variety of experiences, skill, and information. In such circumstances, it can be challenging to establish the benefits of new methods.

This paper explores the topic of validation through an analogy with medical research and development. The principal audience is design researchers although it is hoped that policy makers and industry practitioners may also find the discussion useful. Section 2 reviews the literature on validation in the context of design methodology. Section 3 introduces an analogy between *design methods* and *medical treatments*. Section 4 considers this analogy in the context of three active research topics in engineering design, and some concluding remarks are presented in Sect. 5.

D. D. Frey (✉)
Department of Mechanical Engineering and Engineering Systems Division, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
E-mail: danfrey@mit.edu
Tel.: +1-617-3246133

C. L. Dym
Department of Engineering, Harvey Mudd College, 301 Platt Boulevard, Claremont, CA 91711–5990, USA
E-mail: clive_dym@hmc.edu
Tel.: +1-909-6218853

2 Literature review

2.1 Validation of design knowledge

This paper is concerned with the validation of claims to knowledge in and about engineering design. This subject can be viewed as a specialized topic within epistemology—the branch of philosophy concerned with the nature of knowledge, the justification of knowledge, and the nature of rationality. There are three prominent contemporary views of the justification of knowledge claims (Audi 1995):

- Foundationalism holds that some instances of knowledge are basic and that the remaining instances are justified by relating them to basic beliefs (e.g., by deduction from axioms);
- Relativism argues that knowledge cannot be validated in an objective way and that individual, subjective preferences and rules of fraternal behavior among

scientists must be considered a part of validation processes; and
- Naturalistic epistemology promotes empirical study of how subjects convert sensory data into theories.

Schön suggested that engineering design and related professions such as architecture and management have created a demand for an epistemology of practice (Schön 1983). Schön conducted field studies of engineers and other professionals and noted that skilled practitioners frequently rely on *tacit knowledge* that cannot easily be codified. Dym (1994a, b) has made similar arguments about the difficulty of representing and articulating design knowledge. Argyris called the set of tacit knowledge driving one's professional work a theory-in-action, which, he noted, often failed to match the espoused theory of the discipline or practice (Argyris 1991). Schön and Argyris (1975) proposed a framework for evaluating theories related to professional practice that included checks on: (1) internal consistency, (2) congruence with the espoused theory, (3) testability of the theory, and, ultimately, (4) effectiveness of the theory. They further argue that a theory-in-action is:

- Testable if one can "specify the situation, the desired result, and the action through which the result is to be achieved," and
- Effective when "action according to the theory tends to achieve its governing variables."

Pedersen et al. (2000) proposed a similar framework in which design theories are subjected to a validation square of four quadrants, each representing one of the following design dimensions:

1. Theoretical structural validity.
2. Empirical structural validity.
3. Empirical performance validity, and
4. Theoretical performance validity.

Thus, Pedersen et al. (2000) and Argyris (1991) and Schön (1983) all suggest a balanced approach that includes the evaluation of internal consistency and effectiveness. One key difference is that the framework of Pedersen et al. is more nearly aligned with relativist epistemology: they define scientific knowledge within the field of design as socially justifiable belief. The validation square does balance this relativistic definition of knowledge with use of empirically-based notions of validity. An objective of the present paper is to explore a more objectivist foundation for validation of design methods.

A distinctly different framework was suggested by Simon, who proposed that "human rational behavior is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon 1990). Simon emphasized the fit between problem-solving behaviors and the problem environment, rather than the internal consistency of the behaviors. This concept is sometimes called ecological rationality because it implies that knowledge—as inferred from behavior—should be judged by a fit with the environment (Todd and Gigerenzer 2003). Simon's view of rationality is generally aligned with naturalistic epistemology since it has a basis in empirical studies of agents.

2.2 Validation of methods

Engineers frequently seek to evaluate a design method or software tool for a specified use or range of uses. The Institute of Electrical and Electronics Engineers (IEEE 1998) defines validation as the "confirmation by examination and provision of objective evidence that the particular requirements for on intended use are fulfilled." As will be discussed in Sect. 3, this paper will seek to build upon the IEEE definition showing it is closely related to that of the Food and Drug Administration.

Olewnik and Lewis (2005) propose an alternative definition of validation (applicable to decision support methods and design methods) wherein, for a method to be valid, it must:

1. Be logical;
2. Use meaningful reliable information; and
3. Not bias the designer.

One desirable property of this definition is that it reveals the ways that some design methods impose preferences on the designer. However, a potential drawback of Olewnik and Lewis' definition is that invalidity "does not imply that the methods are ineffective." By contrast, the IEEE definition emphasizes a link between validation and an assurance of effectiveness for its specific intended uses.

Todd and Gigerenzer (2003) have proposed still a different way of validating decision methods (generally aligned with naturalistic epistemology) comprised of the following steps:

1. Proposing computational models of candidate methods that are realistically based on human competences, and testing whether they work via simulation;
2. Mathematically analyzing when and how the methods work with particular environmental structures; and
3. Experimentally testing when people use these methods.

This approach to validation was applied to many decision scenarios resulting in the conclusion that "there is a point where increasing information and information processing can actually do harm." One specific example is that a "Take The Best" heuristic equals or outperforms any linear decision strategy because decision cues are frequently non-compensatory, that is, the potential contribution of each new cue falls off rapidly so that

combinations of later cues cannot outweigh earlier ones (Todd and Gigerenzer 2003). This empirically observed effect, sometimes called the "less is more" effect, is particularly of interest in light of Olewink and Lewis' proposed criteria of validity (Olewnik and Lewis 2005). Todd and Gigerenzer (2003) have provided well-documented examples in which inclusion of valid information into a decision process causes worse decision outcomes rather than better ones.

2.3 Validation of models

Engineers frequently seek to evaluate and apply a model for a specified use or a range of uses. The American Institute of Aeronautics and Astronautics (AIAA 1998) defines model validation as "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model." Much work has been done to practically implement a system of model validation consistent with the AIAA definition. For example, Hasselman has proposed a technique for evaluating models based on bodies of empirical data (Hasselman 2001; Hasselman et al. 1998).

Hazelrigg (2003) proposed an alternative definition of model validity from the perspective of decision theory—a model is valid to the extent that it supports the conclusion that "design point O will produce an outcome that is preferred to the outcome that would be produced by design point C with essentially probability 1." This definition of validity prizes resolution over accuracy. One desirable property of Hazelrigg's definition is that a relatively inaccurate model may be viewed as valid for making choices among alternatives when one of the alternatives is vastly preferred to the others. Hazelrigg's view of model validation is, in some ways, foundationalist, since it seeks to maintain traceability to basic knowledge, in this case to classical decision theory. However, the framework also embraces some elements of subjectivity since, according to Hazelrigg, "a model is valid when, in the mind of the decision maker, it's up to the task." The AIAA definition, by contrast, emphasizes an objective correspondence of a model with data.

McAdams and Dym (2004) have also discussed the validation of models in engineering design, drawing on analogies and procedures used in mathematical modeling (Reich 1994), while also observing that "design models operate on information to produce information." This, of course, leads one to ask, "What is information?", a question to which many answers have been offered (e.g., see McAdams and Dym 2004). Hazelrigg defined information as "what [a] decision is made on," and noted that it can be quantified in terms of probabilistic outcomes of specific decisions. But, as McAdams and Dym noted, there are many aspects of design (e.g., concept generation and synthesis) that are poorly modeled when modeled only as decisions (McAdams and Dym 2004).

2.4 Validation of design research methodology

The engineering design research community also has a need to evaluate research programs and their results. Therefore critical analysis of research methodology has periodically been pursued and has been a healthy part of our community's development. Reich (1994) analyzed the status of research methodology in artificial intelligence for engineering design. His conclusion was that the status is "poor" partly because "it makes claims far beyond what other disciplines dream of making" and that "the need for reflection is not taken seriously in AI." As an antidote to these ills, Reich proposed a layered model of research methodology. The first layer differentiates research methods according to their metaphysical stance. The two options in this layer were *scientism* which is based on an objectivist epistemology and *practicism* which is based on a subjectivist epistemology, tempered to avoid misuse. The second layer concerns research heuristics that characterize different communities, such as cognitive science, decision science, and software engineering. The third layer concerns the most specific issues of hypothesis evaluation including statistical testing and assessment of parsimony. The goals of the present paper are quite similar to those of Reich (1994) but the position taken here and to be developed subsequently is substantially different. Reich says of the scientist/objectivist view "we have witnessed its demise in philosophy." Our view is that it is premature to dismiss an objective basis for evaluating design research, methods, or models. At the same time, the concept of layered evaluations and many of the specific suggestions Reich makes are strongly endorsed, especially "if the purpose is improving practice, then a study that illustrates such improvement should be furnished" (Reich 1994).

2.5 Needs analysis

A review of the literature reveals a substantial range of opinion regarding validation in design theory. On the other hand, engineering's professional societies have been fairly consistent in other matters in seeking validation based on the provision of objective evidence. However, these same professional societies have not yet explicitly applied these concepts to design methods and theories. If the engineering profession does choose to extend an objective concept of validation to design methods and tools, it will need a supporting set of practices and standards for the provision of evidence. Questions such as the following will have to be answered:

- Can theoretical arguments alone serve to validate a design method or theory, or must they ultimately be validated experimentally?
- What kinds of experiments provide valid evidence of effectiveness for an intended use?

- Can design method validation be made economically viable by improvements in speed and efficiency?

In connection with these questions, it is worth noting the recent study by Dorst and Vermass (2005), which analyzes Gero's function–behavior–structure model. Dorst and Vermass (2005) point out that many difficulties arise when one tries to match empirical (or experimental) results against models and theories, including the lack of clear criteria for matching the "quality" of models with "validating" empirical data, and indeed, whether models can be defined or specified sufficiently to be observed in the laboratory (or in practice.)

The principal goal of this paper is to provide some practical answers to these (and other) questions by drawing on the experience of another profession, medicine, which has faced similar questions and answered them with a substantial degree of success. In subsequent sections, it is argued that medical research and development has highly-developed, well-documented validation methods and that many specific practices such as natural experiments and model-based evaluations can profitably be adapted for use in engineering design research.

3 The medical treatment–design method analogy

3.1 Introducing the analogy

The professional community involved in medical research and development has developed a set of well-documented validation processes. The processes used in medicine are far from perfect and many significant mistakes are still made, both by releasing unsafe or ineffective drugs or by withholding effective treatments too long. Still the practices applied in medicine possess many positive attributes from which much can be learned. In an effort to draw out those lessons, this section introduces an extensive analogy between medical research and development, on the one hand, and design theory and methodology, on the other, here called the medical treatment–design method analogy.

The primary goal of medical research and development is to develop treatments to be administered to human patients. The purpose of the treatment is to achieve clinical outcomes related to improved health (e.g., lowering blood pressure). The treatments are developed through academic research, or at pharmaceutical companies, or both. Patients seek advice about treatments from medical professionals and often gain access to such treatments exclusively through them. Before a claim can be made about a treatment's effects, the 1962 amendment of the Food, Drug, and Cosmetics (FDC) Act requires provision of "evidence consisting of adequate and well controlled investigations . . . that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof" (U. S. Federal Food, Drug, and Cosmetic Act, Chapter 9.V, Sec. 355(d), http://www.access.gpo.gov/uscode/title21/chapter9_.html).

The primary goal of design research is to develop design methods to be learned and used by designers to create engineering artifacts, usually within commercial engineering enterprises. The purpose of a design method is to achieve specific design outcomes (e.g., improved profitability or better product quality and reliability). The methods are developed by academic researchers, industry practitioners, or by both. Engineering companies look to a variety of professionals (e.g., engineers. statisticians, and managers) for advice about design methods and for help in implementing them. Although lacking the force of law, the IEEE definition of validation entails "confirmation by examination and provision of objective evidence that the particular requirements for an intended use are fulfilled" (Institute of Electrical and Electronics Engineers 1998).

Table 1 summarizes the analogy just outlined, based on the fact that both areas include professionals seeking outcomes by applying treatments or methodologies that might require validation. An important caution arises from inspection of Table 1. If medical treatments are to be compared with design methods, the most salient differences between the two must be openly acknowledged. Design methods almost invariably require substantial interpretation and judgment in their application. This is true to a lesser extent with medical treatments. Many drugs, for example, are either administered according to the suggested dosage or not at all, leaving little room for interpretation of what administration of the treatment entails. Surgical procedures, on the other hand, may vary substantially depending on the judgment of the practitioner implementing them. Design methods, it must be admitted, lie at an extreme of the spectrum: the application of design methods depends

Table 1 An analogy between medical research and development and design theory and methodology

	Medical research and development	Design theory and methodology
What is validated	Medical treatments	Design methods
Entity affected	Human patient	Engineering organization
Outcomes evaluated	Health, side effects, etc.	Quality, time to market, profitability, etc.
Developers	Academic researchers, pharmaceutical companies, etc.	Academic researchers, industry practitioners, consultants, etc.
Professions involved	Medical doctors, nurses, technicians	Engineers, statisticians, managers
Standards for validation	Food, Drug, and Cosmetics Act, and so on	IEEE definition of validation, and so on

strongly on judgment of the designer who applies them, and their effectiveness is almost assuredly compromised by poor implementation.

While the association among the entities compared is not perfect, it is worth continuing the effort to extend the medical treatment–design method analogy to compare validation of medical treatments and validation of design methods. Validation requires that evidence be provided. The types of evidence provided in medical research and development are rich and varied. Table 2 (below) provides a partial list of the types of evidence in used in medical research. The list is roughly structured downward from the most comprehensive evidence—clinical trials—through varying layers of supporting evidence, each requiring more assumptions and abstractions than the one above. All of the layers of evidence are useful, but as will be discussed, there is a hierarchy to the levels of evidence. The next five subsections discuss each layer in detail and flesh out the analogy with design theory.

3.2 Clinical trials

A clinical trial is the ultimate standard for validating the effectiveness of medical treatments. The U.S. Food and Drug Administration (FDA) has developed detailed guidance for industry on the conduct of such trials (Department of Health and Human Services—Food and Drug Administration 1996, 1998a). Human subjects are identified and allocated to different medical treatments including the specific treatment to be studied and a comparator (sometimes a placebo). The medical treatments are administered, the effects are monitored, and the outcomes are recorded. To avoid bias, the administration of the treatments is usually blinded, often double blinded so that neither the patient nor the physician-researcher knows which treatments are given to which subjects. Further, even though patients may volunteer for or asked to be included in clinical trials, an increasingly common experience in cancer treatments wherein there are other ethical issues whose presence must be acknowledged, the subjects in most clinical trials do not get to choose their treatment(s). The health of the subjects in the clinical trial is affected by a multitude of factors that may not be controlled and/or may not be monitored. For this reason, the data from clinical trials is subject to careful statistical analysis (Department of Health and Human Services—Food and Drug Administration 1998b), and adequate sample sizes are required to reach conclusions with a satisfactory degree of confidence. It should be noted that the requirement for clinical trials does have negative consequences, such as when treatments are withheld from patients that might otherwise benefit from them. The point is not that the methods are perfect nor that they are perfectly executed, only that they are highly-developed and that much can be learned from them.

Table 2 Types of evidence used to develop and validate medical treatments and design methods

Evidence used to develop/validate medical treatments	Evidence used to develop/validate design methods
Clinical trials	Controlled field evaluation of design methods
Natural experiments	Studies of industry practice
In vitro experiments	Laboratory experiments with human subjects
Animal models	Detailed simulations of design methods
Theory (organic chemistry, genetics, molecular biology, cellular biology, etc.)	Theory (probability, decision science, cognitive science, organizational behavior)

Now, to explore the medical treatment–design method analogy for validation, a "clinical trial" might proceed as follows. Design problems would be identified, perhaps as benchmark design problems, and then allocated to different design methods or tools, including the specific design method to be studied and a comparable tool. (It is hard to identify a design equivalent of a placebo here, although perhaps there is a design equivalent of the medical admonition to do no harm that could be brought into play). The design methods are applied, the results are reviewed, and the design outcomes are recorded. To avoid bias, the application of the design methods ought to be done in some blinded fashion, and while the "patient" or the design problem can be oblivious to the "treatment" or the design method being applied, the "physician-researcher" or designer would have to know which method she was applying, else she could not apply it! (On the other hand, whereas subjects in most clinical trials do not get to choose their treatment, designers—and analysts—often recognize that some problems "beg for" certain solutions or treatments.) The outcomes or solutions to the design problem in the clinical trial may be affected by other factors that cannot be readily controlled, although it is more conceivable that problems and their solution methods can here be more readily isolated. Nevertheless, adequate sample sizes would still be required to reach conclusions with a satisfactory degree of confidence.

Thus, a clinical trial for design methods could be an experiment in which different methods are allocated to organizations or to specific tasks therein. The resultant effects on quality, profitability, or time to market might be monitored and analyzed statistically. Such studies are sometimes conducted and a recent example (Kunert 2004) will be discussed in Sect. 4.1. However, as foreshadowed just above, analogies of clinical studies for engineering design raise several questions, including:

- How can appropriate (benchmark) design problems be identified, articulated and represented. Moreover, who should do that?
- Since designers—unlike their physician-researcher counterparts—must be familiar with the design

method they are implementing, does it matter that the process of blinding is likely impossible in this context? If it does matter, how much?
- Since many engineering firms are unlikely to accept and implement a design method not of their own choosing, does it matter if the allocation of treatments or design methods cannot be randomized? If it does matter, how much?
- Can the costs of doing such research—over many companies, designers, methods and products—be kept sufficiently bounded that adequate sample sizes are enabled?

The last question suggests that ways to make clinical validation more efficient should be sought. One way to reduce the cost—or shorten the time—in a clinical trial is to use surrogate variables, that is, parameters that are observed in place of the relevant clinical outcomes when it is difficult to make a direct observation of effectiveness. For example, it might take an overly long time to demonstrate clinically that a drug reduces the long-term risk of heart disease. Therefore, drugs are sometimes approved if they can be shown to have an effect on variables thought to be related to risk of heart disease (e.g. blood cholesterol levels). However, "there have been many instances where treatments showing a highly positive effect on a proposed surrogate have ultimately been shown to be detrimental to the subjects' clinical outcome" (U. S. Department of Health and Human Services - Food and Drug Administration, 1998, Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products, http://www.fda.gov/cder/guidance/idex.htm). Surrogate variables can be misleading indicators of clinical effectiveness and no general procedure for evaluating surrogate variables has yet been broadly accepted in the medical community.

Similarly, surrogate variables may also be somewhat helpful in validating engineering design methods. For example, Olewnik and Lewis' (2005) proposed criteria for evaluating decision support methods could be viewed as surrogate variables. It seems reasonable that a method is more likely to be effective if it is logical, uses meaningful, reliable information, and does not bias the designer. However, as Olewnik and Lewis (2005) acknowledged, these criteria do not *ensure* effectiveness in practice.

3.3 Natural experiments

One alternative to a clinical trial, which is a controlled experiment, is the so-called natural experiment in which the consequences of different treatments are observed—but without controlled administration of the treatments. For example, in the first major study linking smoking and lung cancer, 1,500 patients who had been diagnosed with lung cancer and a similar-sized sample of patients not diagnosed with lung cancer were surveyed about their smoking habits. The result of the study was that "smoking is a factor, and an important factor, in the production of carcinoma of the lung" (Doll and Hill 1950). There was subsequent debate over the validity of the study: since it was a natural—rather than controlled—experiment, correlation had been statistically established, but causation was less certain. The choice to smoke or not was made by the subjects, and those who choose to smoke are not an unbiased, random sample of the population at large. The influence of such considerations on inference should not be underestimated. They prompted the statistician R. A. Fisher to conclude—regarding the cancer research—that "it is more likely that a common cause supplies the explanation" (Fisher 1958). Subsequent studies have substantially removed these doubts, establishing the causal link between smoking and lung cancer without actually resorting to a clinical study in which human subjects were forced to smoke. Today, most medical scientists accept that causation can be established from natural experiments by including (Hill 1966):

- An analysis of temporal relationship;
- Plausibility based on prior knowledge; and
- Coherence with other known facts.

It is interesting to consider what "natural experiments" might entail in design methodology. Every time a new method becomes widely adopted by industry, there is an opportunity for the design research community to learn about its effectiveness and its side effects. For example, many U.S. companies adopted design methods from Japanese companies in the 1980s (e.g., Quality Function Deployment aka, QFD). Many believe that these companies derived benefits in quality, cost, and profitability. A careful study of this "natural experiment" might have more clearly established the effects of this "treatment" on the relevant outcomes. This idea will be explored more fully in Sect. 4.3.

A useful augmentation to natural experiments and clinical trials is meta-analysis: the combination of data from several studies to produce a single estimate. Meta-analysis employs statistical, multi-factorial methods in which the treatment is one predictor variable and the study is another predictor variable (Bland 1987). A principal difficulty with meta-analysis is so-called publication bias. Studies which produce significant differences are more likely to be published than those which do not (Easterbrook et al. 1991). Also, unfavorable results are left unpublished more frequently than favorable results. Thus, in medical research it is advised that all studies be used in any meta-analysis, both published and unpublished, demanding that researchers seek out studies through personal knowledge and intensive investigation.

Design methods can also, in principle, be studied by meta-analysis. As will be discussed in Sect. 4.2, meta-analysis is used in design research, but more extensive studies may be possible. Applications of different design methods are frequently published and they could be

collected and subject to statistical analysis. However, just as in medical research, publication bias is a real concern. Many case studies are developed to illustrate the successful use of a design method. This is a reasonable practice as new methods are much more useful to practitioners when accompanied by examples illustrating successful outcomes. Publication bias is a natural consequence of the dynamics of the publication process and generally not evidence of deliberate attempts at deception. Nevertheless, and while difficult to accomplish, adequate safeguards against bias must be made in meta-analysis of case studies of engineering design methodology. In most cases, it is not possible to collect all of the published and unpublished applications of a design method. Further, if a subset of unpublished cases is sampled, it would be difficult to avoid bias since engineering firms will justifiably be concerned about disseminating evidence of unsuccessful outcomes.

3.4 In vitro experiments

In many cases, applying an experimental medical treatment to a human subject is not justifiable, but it may be possible to apply the treatment to human tissues and cells outside the human body. This type of experiment is described as in vitro as opposed to in vivo. This approach avoids the risk of harming subjects, and sometimes enables closer observation of the effects and mechanisms of the treatment. However, because the tissues or cells are outside the normal context of their existence, great care must be exercised in drawing inferences about clinical effectiveness.

In design theory, human subjects are sometimes used as subjects in laboratory experiments (e.g., Chakrabarti et al 2004; Cacciabue and Hollnagel 1995). These are analogous to in vitro experiments in medicine in the following ways:

- The subject of the experiment is removed from the usual context, in this case, the corporation where most authentic engineering practice takes place.
- Cooperation of an entire engineering enterprise is generally not needed. The human subjects can volunteer individually.
- Closer observation and control of experimental conditions may be possible.

Laboratory experiments with human subjects are currently providing insights into engineering design. However, there is a risk in extending results from laboratory experiments to make inferences about engineering practice. Macrocognition is a term that has been coined to describe the cognitive functions performed in natural—versus artificial, laboratory—settings (Klein et al. 2003). Real-world settings require such activities as problem setting, attention management, planning, and adaptation/re-planning in ways that laboratories can rarely simulate. As a result, tools developed based on laboratory research, especially decision support tools, often degrade performance, rather than improve it (Klein et al. 2003).

3.5 Animal models

Animal models are another important tool in medical research and development. An animal model is an organism—often a mouse—selected or developed to bear specific physiological similarities to humans with specific medical disorders. A variety of animal models exist for studying disorders of the heart, blood, brain, eyes, and so on (Center for Modeling Human Disease, http://www.cmhd.ca). The precise form of an appropriate animal model is strongly determined by the medical disorder being modeled. For example, to study attention deficit hyperactivity disorder (ADHD), mice lacking a specific gene encoding a mechanism regulating brain chemistry, a dopamine transporter or DAT, have been developed, thus the mice are called DAT knockout or DAT-KO mice: "The preponderance of common symptomatologies between the DAT-KO mice and individuals with ADHD suggests that these mice may... serve as a useful animal model and as resource to test new therapies" (Gainetdinov et al. 1999).

The use of animal models entails uncertainty due to the inevitable differences between the animal and the human. For example, of the DAT-KO mice, it is noted that "despite the similarities between the mutant mice and humans with ADHD-HKD, it is unlikely that their phenotypes are completely identical" (Gainetdinov et al. 1999). In this regard, an animal model for medical research and development is similar to an engineering model for use in design. The model is intended to represent another entity, but does not represent that entity in all regards. In fact, there are no universally accepted procedures for validating animal models of human disease because the validity of the models is so closely linked to the investigations to be conducted (S.L. Adamson, personal communication).

Simon proposed a way to go about validation of design methods that is similar to the practice of using animal models in medicine (Simon 1996):

> ". . . there exist today a considerable number of examples of actual design processes, of many different kinds, that have been defined fully . . . in the form of running computer programs . . . Because these computer programs describe complex design processes in complete, painstaking detail, they are open to full inspection and analysis, or to trial by simulation. They constitute a body of empirical phenomena . . . There is no question, since these programs exist, of the design process hiding behind the cloak of "judgment" or "experience." Whatever judgment or experience was used in creating the programs must now be incorporated in them and hence be observable."

In effect, Simon is proposing that a computer simulation of a design process stand in for the actual design process. To be more precise, the argument is that a computer simulated design scenario to which design methods would be applied is a more apt analogy for an animal model as used in medical research. In fact, the animal model analogy suggests that a full correspondence between the simulation and the reality being modeled is not required. Animal models such as DAT-KO mice bear only a few selected similarities to humans with ADHD. Similarly, simulations used in engineering design may bear only certain key similarities to real-world design and still reveal interesting insights. This practice is not uncommon for research in engineering design; many papers use an approach in which computer simulations of design processes are used as a source of epirical knowledge about design (Moss and Cagan 2004). Such studies do not necessarily require precise models of cognitive processes of engineering designers.

Schön has proposed another means to test and use methods, a *practicum*, rather than a computer simulation (Schön 1983):

> ". . . a practicum ... is really a virtual world. A virtual world in the sense that it represents the world of practice, but is not the world of practice. . . "

Just as Simon proposed the computer simulation as a model of real design, Schön has proposed another kind of entity a bit closer to professional design practice. A key difference in Schön's practicum is that an actual person has to carry out the design. Therefore a practicum can assess a design method and the degree to which it fits human cognitive and psychological attributes. Using practica (or something very similar) is common in research in engineering design; for example, researchers often use classroom settings to evaluate design methods (e.g., Wilkening and Sobek 2004; Reich et al. 2006).

3.6 Theory

The underlying theories of medical science such as chemistry and biology play a pivotal role in the development of new treatments. Because basic research catalyzes new innovations, the developers of medical treatments pay careful attention to research and even make substantial investments in privately funded basic research. However, in validation of medical treatments, theory plays a surprisingly small role. Drugs are often administered widely without full understanding of their underlying mechanisms—as long as their effectiveness can be established clinically. For example, it is known that psycho-stimulants have a calming effect on humans with ADHD. This effect appeared so logically inexplicable that medical researchers referred to these calming effects as *paradoxical*: seemingly self-contradictory, yet nonetheless true (Gainetdinov et al 1999). Because the value of the treatments was established clinically, the prescription of psycho-stimulants to ADHD patients continued despite the paradox. Meanwhile, animal models were (and still are) being used in detailed investigation of the effects of psycho-stimulants on the brain functions of DAT knockout mice, which may "provide insights into the basic mechanisms that underlie the etiology of this and other hyperkinetic disorders" (Gainetdinov et al. 1999). These investigations may eventually lead to drugs that more specifically target the relevant mechanisms to ADHD and therefore are more effective and/or have fewer side effects.

The phenomenon of practical uses preceding underlying theory is also common in engineering. The founding of thermodynamics as a field was substantially accelerated by the study of working steam engines. The ability to make workable steam engines preceded understanding of thermodynamics, not the other way around as it is frequently assumed. Similarly, many of the most useful innovations in design methodology including robust design, lean manufacturing, and QFD emerged from industry practice. Only later were these practices studied by theorists. In some cases, the innovations were refined based on the theory, but just as often, theories had to be extended or even revolutionized to accommodate the understanding of the practice. The interplay of theory and application—with applications often leading the way—is long known, but too often forgotten. Design researchers should certainly keep in mind the approach of medical researchers. If theoretical investigations of a design method seem to conflict with field reports about that method, then the design methodology may still be valid and perhaps the theory should be extended to include a richer variety of considerations.

4 Implications of the analogy

This section considers validation in three different areas of design theory and methodology—robust design, axiomatic design, and decision-based design—in order to explore more thoroughly the medical treatment–design method analogy.

4.1 Robust design

Robust design is a set of design methods in which products and processes are made less sensitive to manufacturing variations, customer use conditions, and degradation over time. The techniques were first pioneered by Taguchi who, through his extensive work with Japanese industry, developed a methodology (Taguchi 1987; Phadke 1989). Later, researchers and practitioners—especially in statistics—developed new approaches intended to be more tightly linked to mathematical theory, and especially to design of experiments (DOE).

Design of experiments is a body of knowledge and techniques for planning a set of experiments, analyzing

the resulting data, and drawing conclusions from the analysis (Wu and Hamada 2000; Box et al. 1978). DOE in general—and fractional factorial design in particular—have been key theoretical foundations for development of robust design methods. Both Taguchi methods and newer techniques use fractional factorial experiments, but in different ways. Recent theoretical developments have led statisticians to discourage the use of crossed arrays and encourage the use of a single array including both control and noise factors (Wu and Hamada 2000; Borror and Montgomery 2000). The justification is that ". . . some of the single arrays . . . are uniformly better than the cross arrays in terms of the number of clear main effects and two-factor interactions" (Wu and Hamada 2000).

In robust design, the "clinical endpoint"—what the engineering organization wants—is improved system performance in the presence of noise. Single-array designs have been justified on the basis of the number of clear effects and of control by nose interactions. One may say that these properties are being used as surrogate variables, and that the clinical endpoints should be evaluated in a clinical trial. Such a trial was recently conducted by Kunert et al. (2004). They used crossed-array designs and single-array designs to improve the consistency of a sheet metal spinning process. In effect, this was a paired comparison experiment where the experimental "treatment" was the design method employed. The result of the experiment was that the crossed-array method led to process settings with a more consistent profile of the sheet metal parts when compared with the single-array method.

The "clinical trial" by Kunert et al. provides some objective evidence validating the crossed-array method and disconfirming the single-array approach to robust design. However, this evidence is not conclusive. First of all, the experiment included only one replication of the paired comparison. If the exact same experiment were repeated, the single array might have beaten the crossed array simply due to random variations. Further, if the same two methods were applied to some other engineering system, say a lithography process rather than a sheet metal spinning process, the single array may have prevailed over the crossed array. A paired comparison in a single instance of application does not provide as much information as one would like to validate a design method.

Now, let us consider how validation methods analogous to those in medicine can be applied in robust design. As an example of meta-analysis, Li and Frey (2005) conducted a study of experiments carried out in many fields of engineering to ascertain the degree that system regularities such as effect sparsity, hierarchy, and inheritance were evident in the resulting data. Frey and Li (2004) then conducted a model-based investigation of crossed-array and single-array designs using a computer simulation of robust design on a large number of simulated responses drawn from a third order polynomial model of the system's response. A crossed-array design resulted in a standard deviation of the response substantially smaller than that provided by a single-array design, although both methods were much better than a "placebo" (selecting control factor settings at random) (Frey and Li 2004). These model-based results tend to corroborate Kunert's "clinical" study showing that the single instance in the field was probably typical of the population of realistic scenarios in the field.

The results in this section suggest that it is a mistake to infer too much about robust design methods based solely on mathematical theory. A similar conclusion has been voiced by Box and Liu (1999). In recent decades, the DOE research community has emphasized the development of mathematically optimal experimental designs which tend to be "one-shot" procedures. Such designs are optimal within a formal axiomatic framework, but Box has argued that, in practice, they undermine the experimenter's need to alternate between forming hypotheses and conducting experiments (Box and Liu 1999). Box further argued that this resulted in less improvement of systems than would have been achieved by iterative procedures (even though they are partly heuristic). Box's findings, if accepted, provide an instance of overemphasis on mathematical theory leading to ineffective professional practice due to neglect of relevant human factors.

4.2 Axiomatic design

Axiomatic design posits two "axioms": the independence axiom and the information axiom (Suh 1990). The full theory of axiomatic design, including theorems and corollaries, is based on theorems whose validity, it is said, relies on the validity of the axioms. Thus, axiomatic design theory employs foundationalist epistemology. As long as the axioms are valid and the other elements of the theory are deduced from the axioms, it is proposed that the theory is sound and useful for professional practitioners. In particular, it is argued that the theory is useful in the design of large scale systems along almost all dimensions of their design and operation. That is, axiomatic design claims to improve design, construction, operation, modification, maintenance, documentation, and diagnosis of system failures (Suh 1998).

Suh (1990) defines axioms as "truths that that cannot be derived or proved to be true except that there are no counter-examples or exceptions". This suggests that the means to validate the axioms is to look for counter-examples. However, Suh's axioms are in the imperative form. The *independence axiom* is "maintain the independence of functional requirements." It is not clear what a counter-example to an imperative might entail; the axiom is not testable—as defined by Schön (1983) and Argyris (1991) and described in Sect. 2.1—because it does not "specify the situation, the desired result, and the action through which the result is to be achieved."

The clinical trial part of the medical treatment–design method analogy suggests one testable hypothesis about

axiomatic design. If training in axiomatic design is viewed as a treatment, the hypothesis might be put forth that providing training in axiomatic design to a group of engineers will enable them to produce designs having a higher "probability of success" than a group that receives training in another method (a comparator). By devising a study to test this hypothesis, axiomatic design might be objectively assessed. It might also enable the assessment of how well axiomatic design theory can be learned and implemented in practice. Such an investigation would be very costly.

The natural experiment part of the medical treatment–design method analogy suggests another approach. A software product, Acclaro, has been developed to support axiomatic design and is currently being marketed to engineering companies (Axiomatic Design Solutions Inc. website http://www.axiomaticdesign.com/). In effect, a natural experiment is currently underway. The company that sells the software lists among its clients Ford Motor Company, Lockheed Martin, Hewlett Packard, Saab Rosemount, and others. Also listed is ASML, which acquired Silicon Valley Group Incorporated—a company that was among the first to use the software. A study of these companies might reveal how this treatment has affected various measures of their performance, such as product quality ratings, time to market, profitability, market share, and so on. Such a study could not provide a definitive evaluation of the theory, but is essential to forming a complete assessment.

The in vitro experiment part of the medical treatment–design method analogy suggests still another means of evaluating Axiomatic Design. A simulation of a design process could be created to test a hypothesis such as, "if a system design does not maintain the independence of functional requirements, then the set of requirements cannot be satisfied." In fact, just such an investigation was carried out. Human subjects were asked to satisfy a set of n functional requirements using a set of n design parameters for both uncoupled systems and coupled systems of modest size (from $n = 2$ to $n = 5$) (Hirschi and Frey 2002). The experiment showed that human subjects could succeed in simultaneously satisfying all of the requirements of both uncoupled and coupled systems, but that the task completion time scaled linearly for uncoupled systems and geometrically for coupled systems. This investigation does not support the idea that coupling must be avoided, but it does provide empirical support for the idea that coupling has some negative implications in system design.

The results in this section suggest that an axiomatic approach to design theory development will, in some instances, lead to logical or practical difficulties. If axioms are stated regarding design, they may not be testable unless they are stated as empirical hypotheses. As theorems are derived from the axioms, one must ensure not only consistency with the axioms, but also consistency with every fact of reality that bears on the theorem, including human cognitive capabilities. Practices from medical research and development may help avoid such difficulties. Data should be collected from ongoing natural experiments and laboratory experiments should be used to further investigate the implications of the theory.

4.3 Making design decisions

Designers frequently find themselves having to rank objectives against one another, choose among alternative means for achieving functions, select from alternative design options, and generally make a wide variety of design decisions (Dym and Little 2004). Many of the tools used to help designers make such decisions have become widely used, as well as often criticized. For example, QFD is a method frequently used in the early phases of product design (Hauser and Clausing 1988). QFD was first developed at the Kobe ship yard in Japan, then adopted by major Japanese companies and subsequently by a broad variety of companies in North America and Europe. One primary goal of QFD is to help companies align their designs with the "voice of the customer." One claim often made about QFD is that the method encourages teamwork among engineers and marketing professionals by providing immersion in the specifications prior to design efforts and by setting appropriate technical targets for mature products (Dym and Little 2004). There is some empirical evidence that suggests that QFD has not been correlated in the short term with quality improvements, but that users nevertheless feel it has longer term benefits (Griffin 1989).

Meanwhile, QFD and similar decision-making tools such as the pairwise comparison chart (PCC) and the analytical hierarchy process (AHP) have been taken to task by some mathematicians and design theorists. Saari (2001) has investigated the mathematics of different forms of voting and of related decision support procedures. These investigations have brought to light undesirable properties the methods exhibit under certain circumstances. Hazelrigg (1998, 1999) has argued that all of these methods are simply wrong because they violate tenets of decision theory. One specific conclusion drawn by Hazelrigg is that industry should stop using QFD. However, the framework for validation used in medicine suggests that QFD should not be abandoned so easily because the existing field data show some positive effects of QFD (Kuppuraju et al. 1985). Also, it is not obvious that decision theory is an adequate model for many forms of design decision making (Dym et al. 2002). Rejecting QFD because of Arrow's theorem in decision theory is akin to telling patients to reject medicines whose effects cannot be explained using current knowledge of chemistry and biology. No medical treatment can work except by a means that may ultimately be subject to scientific explanation. However, in practice, many effective medical treatments are discovered before their relevant mechanisms are understood. Under currently accepted medical practices in the developed world,

it is acceptable to market a treatment without a full understanding of its underlying mechanisms as long as the treatment's safety and effectiveness can be clinically established. Design researchers might conclude, at least tentatively and cautiously, that QFD and similar tools are currently the subject of field evaluation. At the same time, design researchers, in collaboration with design practitioners, should continue to investigate the mechanisms that provide perceptible benefits and simultaneously use theory to seek improvements. Reports from industry practice suggest that the long term value of QFD derives from factors such as cross-functional coordination and information flow (Griffin 1989). The medical analogy suggests that models should be developed that adequately represent these mechanisms, and then these models should be used to test and refine our decision making methods.

Although the medical analogy seems to suggest QFD may continue to be used, the analogy also raises an important issue concerning labeling. Over-the-counter medications must carry carefully defined instructions for their use. Without such labels, medicines would be less useful and more dangerous. Such labeling has been suggested for QFD through research in engineering design. Dym et al. (2002) suggest that QFD and related methods can be used cautiously at early stages of design, but should not be used as the sole means to establish a single numerical rating that can then be used to choose a single alternative. Further, specific cautions are made against over-interpretation of numerical differences because so much of design knowledge is not amenable to straightforward mathematical modeling (Dym and Little 2004; Dym et al. 2002). This paper suggests that researchers in engineering design should pay close attention to such instructions since they are an essential part of the "treatments" being developed.

5 Closing remarks

The principal argument of this paper is that many of the highly-developed validation techniques found in medicine can profitably be used in engineering design research. Validation of design methods is an important topic for engineering design theory and practice. Every engineer that designs products or processes uses some design method or composite of methods. The design methods now in use have an impact on all aspects of success, including quality, performance, cost, and time to market. Therefore, it seems reasonable to seek validation procedures for design methods that are as good as the ones applied in developing medical treatments. Inasmuch as medical treatment is so widely viewed as important, the associated validation procedures are well-documented, objective, and evidence-based. New treatments must provide proof of their effectiveness before they can be deployed widely. This proof is built up of many layers, including basic theory in chemistry and biology, in vivo experiments with animal models, clinical trials with human subjects, and natural experiments.

Model-based validation is one of the techniques we most strongly encourage design researchers to adopt. In medicine, tremendous investments are made in creating animals that reasonably model aspects of human disease. Similar efforts might be made to create models of engineering design processes for the explicit purpose of evaluating design methodologies. As discussed in Sect. 4.1, some small scale examples have been carried out in robust design, but much more may be possible with concerted effort.

Another practice in medicine worthy of consideration is the careful collection and analysis of field data as new treatments are introduced. Similar natural experiments in engineering design are ongoing every time a new method is adopted widely in the field. It is essential that design researchers work to collect data and include analysis of such data in evaluations of methods such as QFD, Pugh controlled convergence, Taguchi methods, and Axiomatic Design. All of these methods have seen substantial use in the authentic context of industry practice. At this point, any broad statements about either their validity or inherent flaws must be interpreted in light of what has actually transpired.

In addition to these concrete suggestions for design research practice, a more philosophical point may also be made regarding the role of foundations in our field. In both medicine and in engineering design it has frequently been observed that as the many layers of evidence are traversed, surprising discoveries are made. Developments based on theory alone may prove to be ineffective in practice. Despite this fact, the field of design research has, at times, taken an exaggerated stand regarding theoretical foundations—that as long as consistency with first principles can be maintained, then good methodology is likely to result. Leonard Savage, in developing an axiomatic basis for statistics and decision making, provided a caution against this approach and suggested a more balanced view of the interactions between foundations and professional practice:

> "It is often argued academically that no science can be more secure than its foundations, and that, if there is controversy about the foundations, there must be even more controversy about the higher parts of the science. As a matter of fact, the foundations are the most controversial part of many, if not all, sciences... As in other sciences, controversies about the foundations of statistics reflect themselves to some extent in everyday practice, but not nearly so catastrophically as one might imagine. I believe that here, as elsewhere, catastrophe is avoided, primarily because in practical situations common sense generally saves all but the most pedantic of us from flagrant error... Although study of the foundations of a science does not have the role that would be assigned to it by naïve first-things-firstism, it certainly has a continuing importance as the science develops, influencing, and

being influenced by, the more immediately practical parts of the science (Savage 1954).''

It is hoped that this discussion of validation frameworks in medical research will help the design research community continue and strengthen its approach to validation of design methods. Perhaps the practices from medicine can help design researchers avoid the pitfalls warned against by Savage and encourage a healthy interplay between scientific foundations and professional practice.

Acknowledgments D. D. Frey gratefully acknowledges the financial support of the National Science Foundation (award #0448972) and the Ford/MIT Alliance. The extensive comments offered by Yoram Reich have been very beneficial to the authors in completing this manuscript, as have the suggestions of the anonymous reviewers.

References

American Institute of Aeronautics and Astronautics (1998) Guide for the verification and validation of computational fluid dynamics simulations, AIAA G-077–1998

Argyris C (1991) Teaching smart people how to learn. Harvard Business Review, Reprint No. 91301

Audi R (ed) (1995) The Cambridge dictionary of philosophy. Cambridge University Press, Cambridge

Bland M (1987) An introduction to medical statistics. Oxford University Press, New York

Borror CM, Montgomery DC (2000) Mixed resolution designs as alternatives to Taguchi inner/outer array designs for robust design problems. Qual Reliability Int 16:117–127

Box GEP, Hunter WG, Hunter JS (1978) Statistics for experimenters: an introduction to design, data analysis, and model building. Wiley, New York

Box GEP, Liu PTY (1999) Statistics as a catalyst to learning by scientific method. J Qual Technol 31(1):1–29

Cacciabue PC, Hollnagel E (1995) Simulation of cognitive applications. Expertise and technology: cognition and human computer cooperation. Lawrence Erlbaum Associates, Mahwah, pp 55–73

Chakrabarti A, Morgenstern S, Knaab H (2004) Identification and application of requirements and their impact on the design process: a protocol study. Res Eng Des 15(1):22–39

Doll R, Hill AB (1950) Smoking and carcinoma of the lung: preliminary report. Br Med J 2:739–748

Dorst K, Vermass PE (2005) John Gero's function–behavior–structure model of designing: a critical analysis. Res Eng Des 16(1):17–26

Dym CL (1994a) Representing designed objects: the languages of engineering design. Arch Comput Methods Eng 1(1):75–108

Dym CL (1994b) Engineering Design: A Synthesis of Views. Cambridge University, New York

Dym CL, Wood WH, Scott MJ (2002) Rank ordering engineering designs: pairwise comparison charts and boorda counts. Res Eng Des 13:236–242

Dym CL (2004) Principles of mathematical modeling, 2nd edn. Elsevier, Boston

Dym CL, Little P (2004) Engineering design: a project-based introduction, 2nd edn. Wiley, New York

Easterbrook PJ, Berlin JA, Gopalan R, Mathews DR (1991) Publication bias in clinical research. Lancet 337:867–872

Fisher RA (1958) Cigarettes, cancer, and statistics. Centennial Rev II(2):151–166

Frey DD, Li X (2004) Validating Robust parameter design methods, DETC2004-57518. In: ASME design engineering technical conference, September 28 to October 2, Salt Lake City, Utah

Gainetdinov RR, Wetsel WC, Jones SR, Levin ED, Jaber M, Caron MG (1999) Role of setotonin in the paradoxical calming effect of psychostimulants on hyperactivity. Science 283:397–401

Griffin A (1989) Evaluating development processes: QFD as an example. Marketing Science Institute Report, pp 91–121

Hasselman TK (2001) Quantification of uncertainty in structural dynamics models. J Aerospace Eng 14(4):158–165

Hasselman TK, Anderson MC, Gan W (1998) Principal components analysis for non-linear model correlation, updating, and uncertainty evaluation. In: Proceedings of the 16th IMAC, Bethel, pp 644–651

Hauser J, Clausing DP (1988) The house of quality, Harvard Business Review

Hazelrigg GA(1998) A framework for decision-based engineering design. ASME J Mech Des 120:653–658

Hazelrigg GA (1999) An axiomatic framework for engineering design. ASME J Mech Des 121:342–347

Hazelrigg GA (2003) Thoughts on model validation for engineering design, DETC 2003/DTM48632. In: ASME design engineering technical conference, September 2–6, Chicago

Hill BA (1966) The environment and disease: association or causation?. Proc Roy Soc Med 58:295

Hirschi NW, Frey DD (2002) Cognition and complexity: an experiment on the effect of coupling in parameter design. Res Eng Des 13(3):123–131

Institute of Electrical and Electronics Engineers (1998) IEEE Standard for Software Verification and Validation, IEEE Std 1012–1998

Klein G, Ross KG, Moon BM, Klein DE, Hoffman RR, Hollnagel E (2003) Macrocognition. IEEE Intell Syst 18(3):81–84

Kunert J (2004) A comparison of Taguchi's product array and the combined array in robust-parameter-design. In: Eleventh annual spring research conference (SRC) on statistics in industry and technology, Gaithersburg, May 19–21

Kuppuraju N, Ittimakin P, Mistree F (1985) Design through selection: a method that works. Des Stud 6(2):91–106

Li X, Frey DD (2005) A study of factor effects in design of experiments. DETC2005-85486. In: Proceedings of ASME design engineering technical conferences, Long Beach, September 24, pp 1–10

McAdams DA, Dym CL (2004) Modeling and information in the design process. In: Proceedings of the 2004 ASME design engineering technical conferences, Salt Lake City, September 2004

Moss J, Cagan J (2004) Learning from design experience in an agent-based design system. Res Eng Des 15(2):77–92

Olewnik AT, Lewis KE (2005) On validating engineering design decision support tools. Concurr Eng Res Appl 13(2):111–122

Pedersen K, Emblemsvag J, Bailey R, Allen JK, Mistree F (2000) Validating design methods & research: the validation square. DETC2000/DTM14579. In: Proceedings of the ASME design engineering technical conference, Baltimore

Phadke MS (1989) Quality engineering using robust design. PTR Prentice-Hall, Inc., A Simon & Schuster Company, Englewood Cliffs

Reich Y (1994) Layered models of research methodologies. Artif Intell Eng Des Anal Manuf 8(4):263–274

Reich Y, Kohlberg E, Levin I (2006) Designing contexts for learning design. Int J Eng Educ (in press)

Saari DG (2001) Decisions and elections: explaining the unexpected. Cambridge University Press, New York

Savage LJ (1954) The foundations of statistics. Dover Publications Inc., New York

Schön DA (1983) The reflective practitioner: how professionals think in action. Basic Books, New York

Schön DA, Argyris C (1975) Theory in practice: increasing professional effectiveness. Jossey-Bass, San Fransisco

Simon HA (1990) Invariants of human behavior. Annu Rev Psychol 41:1–19

Simon HA (1996) The sciences of the artificial, 3rd edn. MIT Press, Cambridge

Suh NP (1990) The principles of design. Oxford University Press, Oxford

Suh NP (1998) Axiomatic design theory for systems. Res Eng Des 10:189–209

Taguchi G (1987) System of experimental design: engineering methods to optimize quality and minimize costs. Translated by Tung LW, QualityResources: a Division of the Kraus Organization Limited, White Plains, and American Supplier Institute, Inc. Dearborn, vol 1, pp 1–531

Todd PM, Gigerenzer G (2003) Bounding rationality to the World. J Econ Psychol 24:143–165

U. S. Department of Health and Human Services—Food and Drug Administration (1996) Good clinical practice: consolidated guidance. http://www.fda.gov/cder/guidance/idex.htm

U. S. Department of Health and Human Services—Food and Drug Administration (1998a) Providing clinical evidence of effectiveness for human drug and biological products, http://www.fda.gov/cder/guidance/idex.htm

U. S. Department of Health and Human Services—Food and Drug Administration (1998b) Statistical principles for clinical trials, http://www.fda.gov/cder/guidance/idex.htm

Wilkening S, Sobek DK (2004) Relating design activity to quality of outcome: a regression analysis of student projects. In: ASME international design engineering technical conferences. Salt Lake City, Sept 28–Oct 2

Wu CFJ, Hamada M (2000) Experiments: planning, analysis, and parameter design optimization. Wiley, New York