Explore, Explain, Design

Andrew S. Gibbons
Brigham Young University, Provo, Utah, USA

C. Victor Bunderson
The EduMetrics Institute, Provo, Utah, USA

Glossary

design experiments Exhibit all aspects of a design study, except that, in seeking explanatory and design theories, reliance on narrative methods is supplemented with invariant measurement of the growth or change constructs spanning the domain. The measurement instruments evolve over the cycles of design; they implement, evaluate, redesign, and come to embody an increasingly adequate descriptive theory of the processes operative in the domain. In addition, the technological devices designed to introduce and control the treatment effects are forthrightly described using the emerging layers and languages of technology in that domain. In experimental terminology, design experiments are quasi-experiments, but may include mini-randomized experiments within a larger cycle.

design research Includes design studies and design experiments, both of which build domain-specific descriptive theories as well as design theories. Design research also includes research on design methods or design theory, as applied across two or more domains.

design studies Seek two kinds of theoretical knowledge: first, a descriptive explanation of the processes operative in a domain, and second, technological or design knowledge about how to create and implement the tools—both measurement instruments and the treatment control technologies. These studies are attempts to discover new artifact- and intervention-related principles or to improve the effectiveness of existing artifacts or intervention plans. Design studies take place in live settings, and are iterative, cyclical applications of a process of principled design, implementation, evaluation, and redesign. Design studies often aid in exploring a domain and possible treatments, and thus may be largely qualitative, producing narrative accounts. These accounts may not provide adequate warrant for causal claims, nor do they fully rule out alternate explanations.

invariant measurement A careful and extended design, development, and validation process may obtain technological devices (measurement instruments) that produce measures approaching the ideal of invariance. The measures produced are invariant to the sample of subjects or the set of tasks selected, have an agreed-upon zero point and a standard unit used by many practitioners and scientists across national boundaries, and have a widely accepted interpretive framework based on a theory of the operative constructs of change or growth being measured in the domain spanned by the instrument's scales.

natural history A knowledge-producing process in which natural phenomena are observed, described, measured, and collected to amass a body of facts from which patterns and trends can be detected through study of the facts in the presence of each other. This type of knowledge seeking is equated here with exploratory research. A benefit of attention to this type of knowledge seeking is the growth of collections, which can be studied and restudied as evidence to support or question new models and hypotheses as they are advanced.

science A knowledge-producing process in which questions are asked, primarily of an explanatory nature, and research is carried out to answer them. The questions asked seek to describe authoritatively the nature of underlying operations that lead to observed phenomena. This type of research attempts to discover the single best coherent description of observed phenomena that is consistent with all observations; here, explanatory research and science are equated.

technology A knowledge-producing process in which questions are asked, primarily to learn principles for connecting human intentions with the form and function of human-made artifacts, and research is carried out to answer these questions. The questions asked seek ways of structuring

Encyclopedia of Social Measurement, Volume 1 ©2005, Elsevier Inc. All Rights Reserved.

time and space with information, forces, and materials in such a way that a particular prespecified outcome is achieved. Here, this activity is placed under the heading of design. Rather than a single explanation of observed phenomena, this type of knowledge seeking attempts to discover efficient structuring principles and processes that will produce a variety of solutions to a problem, from which the most suitable may be selected on the basis of problem-specific criteria.

The explore, explain, and design concept designates three synergistically related knowledge-producing enterprises. "Explain" denotes the familiar goal of scientific research, which is to explain why and explain how. In this type of research, the single best explanatory principle is sought. "Design" denotes the goal of design research, which is to discover and apply structuring and synthesizing principles in order to satisfy a set of criteria. In this type of research, classes of increasingly desirable artifacts (or structuring principles) are sought so that they may be used to satisfy target criteria. "Explore" denotes a type of research aimed at producing observations that can lead to category formation and formation of hypotheses of relationships relevant to both of the other research enterprises. Within emerging domains of human knowledge, the questions concern what is there and what are possible groupings and relationships among what is there.

Distinguishing Explore, Explain, and Design

The three knowledge-producing enterprises—explore, explain, and design—are necessary conditions for each other, each producing results that become mutually and self-inputting, providing a continuous stream of research questions in all three areas. The three enterprises are discriminated on the basis of the kinds of questions they address and the types of knowledge they produce, but not definitively on the research techniques employed. The focus here is to place these three knowledge-producing activities into context with each other, and to highlight design research. Throughout this encyclopedia, the word "design" appears in only a handful of titles, referring, for instance, to experimental design, sampling design, survey design, and longitudinal cohort design. These are methods of setting up data- gathering opportunities for exploration- and explanation- centered observations, but design as a synthetic process by which knowledge can be gained is not featured. This is the domain here. Moreover, though measurement foundations and models are discussed in the literature, little attention is paid to the fact that the practice of research through synthesis requires measurement, and that measurement instruments are technological products—tools both for research and for practical purposes. These tools, however, have unique properties of scientific interest. Even prior to the existence of adequate theory, the development and use of new measurement instruments has repeatedly, in the history of science, led directly to discovery and then to new theory. As instruments and theory evolve together, the instruments are used to test, confirm, or discredit and replace theory with better theory. Measurement instruments are thus inextricably linked as products of—and as precursors to—the advancement of theory.

Social measurement requires extensive and disciplined design and development to design, produce, validate, and maintain scientifically, socially, and economically valuable measurement instruments. Examining the relation of explore, explain, and design and their relationship to measurement is important because valuable measurement instruments produce social, economic, and scientific capital. Producing this capital is the work of natural history and technology as much as it is the work of science, a distinction elaborated herein. Explore, explain, and design activities are mutually sustaining, and no one of them can proceed far without producing questions to be answered by the other two. Figure 1 illustrates this relationship. The two-way arrows in Fig. 1 indicate that influence between activities can act in either direction. In the graphical depiction, the explore activity is situated above the explain and design activities, indicating that exploration often tends to provide a starting point, stimulating the other two activities.

New concepts about invariant and computer-administered instrumentation make possible research methods in the social sciences that offer new and promising opportunities for all three knowledge-producing activities. Computers and networks can support introduction of both treatments and metrics into live settings where real phenomena of interest are found, though it requires far greater engineering and scientific investment. Extended efforts to achieve broad acceptance (efforts far beyond what is usually attempted in the social

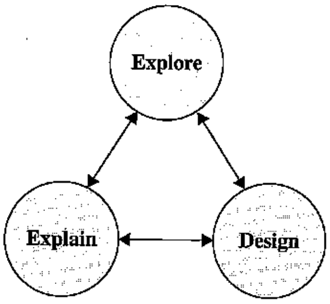

Figure 1 Mutually sustaining relationship between three knowledge-creation activities: explore, explain, and design.

sciences) can then approach the ideal of invariant measurement, providing comparability from occasion to occasion and from group to group, so that unprecedented progress can be made. However, progress will continue to be limited in the human sciences and in design so long as the metrics are incommensurable from study to study.

Explore

Early Scientific Exploration—Natural History

Before it is possible to create theories, we must have some description of nature and the content to which theories pertain. Natural history has been described as a type of research involving the collection and cataloguing of specimens as an inventory of "what we've got." This becomes a register of facts and a compilation of the contents of the world. This way of knowing is contrasted with experimental inquiry into cause, which originated as natural philosophy, i.e., an explanation of causes rather than an inventory.

Natural history studies have historically embraced anything that can be named and collected, though it is a common misperception today that naturalistic studies are confined to animals and plants. One of the most famous collectors, the founder of the British Museum, Sir Hans Sloane (1660–1753), collected thousands of plant and animal specimens, at the same time amassing collections of coins, medals, spoons, goblets, rings, minerals, weaponry, combs, and a huge library of manuscripts. Today, similar collections of human-made objects from around the world compete in size with those of naturally occurring phenomena. Sloane's collections, donated to government, formed an important core of the British Museum. Sloane considered his major contribution to science the "collection and accurate arrangement of these curiosities."

Natural history explorations have a long history of rich collectors assembling collections of oddities, gathered mainly as conversation pieces and enhancements to social status. However, natural history research expanded greatly from approximately 1800 to 1900 and changed in character. More serious and systematic collectors emerged to gather and catalogue with the purpose of arranging specimens in orderly tableaux that emphasized gradation of similarities, suggested regular underlying patterns, and identified gaps that supplied targets for further collection. The method of natural history research became wide search, collection, description, cataloging, preservation, preparation for display, and archiving. This, of course, required that systems of measurement be devised for describing the magnitude of the different qualities of specimens. The search of natural historians became guided by empty positions in collections: places where something ought to be. The sense-making of natural historians was to find the patterns and gradations of difference that might give hints about underlying relationships. These motives continue in today's natural history research. Researchers today work with all possible speed to catalogue Brazilian tribes still untouched by modern civilization, dying native languages, mutant viruses, and evolving forms of computing machinery. Though the subjects of observation, collection, measurement, and inventory have changed, as have the means of measuring, what has not changed is the necessity of measuring observed phenomena. The formation of categories that result from this type of research depends on ordering along presumed dimensions, leading to measurement. The constant refinement of categories goes hand in hand with the constant refinement of all parts of measures: constructs, dimensions, observable indicators, rubrics, scales, instruments, and inference/analysis methods.

Many Exploratory Methods Exist and Often Lead to Explanation

Natural history collections lead to the formation of named categories, and subsequently to order relationships. In the mathematical theory of measurement, this empirical achievement provides the basic definitions needed to construct theorems of representation from the empirical world into the real number system. Theorems of representation must be anchored in the empirical world. First, the categories of objects are defined as sets, and order relations among them are defined as relations on those sets. When these sets and relations meet certain axioms, it can be shown that a mathematical structure (usually the real number system with its many desirable properties) can be used to represent or model the real-world observations. Mathematical models yield predictions of consequences (often otherwise unexpected) that can be tested using methods much stronger than natural history or other largely exploratory methods. Useful explanatory theories can result, with warrant to claims that the causal connections proposed in the theory operate as claimed, and warrant to claims that other competing explanations are not the causal agents. This path from exploration through category formation through measurement, which enables crucial and rigorous testing, is typical in science. Discussion of theorems of representation, based on the fundamental work of David Krantz and co-authors, in three volumes published in 1971, 1989, and 1990, shows the fundamental contribution of exploration to further progress toward explanation. Exploration leads to the formation of the categories (sets) of empirical observations, and relations among them. These lead in turn to measurement, and then to strong tests of causal predictions and relations among them.

The intent of some methods is initially exploratory, but may lead to hypotheses testable by stronger methods. Exploratory narrative methods include case studies, field and observational studies, naturalistic inquiry, some of the cognitive methods, focus groups, participant observation, unobtrusive methods, and others. In all these cases, experts in these methods point out a linkage to theories in their domains of interest, which increasingly guide and shape the plans for their otherwise exploratory observations. Thus, it may perhaps more properly be said that a method is exploratory when it is used in an early phase of observation in a relatively new area. In a later phase, when constructs and relations have been observed, named, and documented, beginning explanatory theories come into being and guide explanatory research. The example of the "empty space" on the shelf in a natural history collection is a case in point. Even the beginnings of relational constructs can direct the search for interesting observations. A method used early to explore a new domain may be used later to investigate a theory-based prediction.

In addition to narrative methods, quantitative methods are also used in exploratory phases. Exploratory factor analysis, cluster analysis, latent class analysis, and Q methods produce sets of possible groupings—possible constructs with varying explanatory value and potential utility. After further development, the quantitative methods give way to confirmatory analyses, structural equation models, hierarchical models, and causal models. Other more quantitative exploratory methods include data mining, computerized record linkage, and statistical matching.

Exploratory Research Leads Also to Design Research

Just as exploratory research leads naturally to research into causes, it leads also to research into outcomes and the means of predictably producing them. Exploration leads to measurement instruments, which are technological devices, the products of design. "Design" is a term that describes intentional structuring of artifacts and intervention plans to bring about predictable outcomes. Design is both a subject of research and a method of research and knowledge production. As a subject of research, it poses many interesting questions:

- How can knowledge about how things work be converted into principles for making things work in a particular way to reach specified goals?
- What generalizable structuring principles guide synthesis and dimensioning of artifacts?
- What generalizable principles guide the formation of intervention plans using artifacts?
- How can practices that currently produce desirable results be refined to meet ever higher, ever more specific criteria?
- How can new measurement constructs, measurement instruments, and constructs be created that permit better measurements of artifact properties, intervention points, intervention progress, and attainment of criterial end points?

As a method of research, design involves principles of search, analysis, problem re-representation, discovery of abstract operational principles, principled proliferation of configurations, means-end matching, optimizing, and testing—all of which requires constant, continuous measurement. Herbert Simon has attributed recent rapid growth of interest in design research to "the fact that all who use computers in complex ways are using computers to design or to participate in the process of design. Consequently, we as designers or as designers of design processes, have had to be explicit as never before about what is involved in creating a design and what takes place while the creation is going on." The domain of design can be illustrated as in Fig. 2. As shown, natural processes involving the transfer and translation of energy and information proceed independently of human agency. Even matters as complex as human learning and forgetting are natural processes, and either one may proceed in a natural but less desirable path than if instructional interventions are introduced. Natural processes are the subject of much exploratory (natural history) and explanatory (scientific) research. Humans, in the exercise of technology designs, identify measured intervention points at which they impress energy (force) or information (structure) on natural processes, using artifacts and intervention plans in a way calculated to deflect the natural processes toward the achievement of desired, measured outcomes. Attaining the desired outcomes may require multiple interventions, multiple intervention artifacts, and multiple, often continuous, measurements.

Intervening in different processes at different measured points using different artifacts is an important

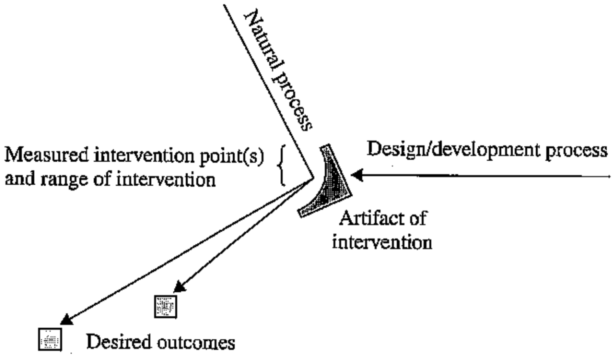

Figure 2 The domain of design.

method of knowledge production through design. It can rely on random permutation, which results in occasional success, or it can proceed deliberately, according to generative design principles that promise a higher rate of success and predictability. Of interest here are the many ways in which the technology of measurement is not just involved in, but makes possible, other technologies. Measurement is therefore a central topic in design research. Design research produces knowledge that is nontrivial, and according to Walter Vincenti, consists of several distinct types:

1. Fundamental design concepts—principles for structuring designs.
2. Criteria and specifications—means of expressing problems and designs.
3. Theoretical tools—concepts that relate scientific to technological knowledge and that support the formulation of design theories.
4. Quantitative data—data on properties and ranges, constants, data for prediction.
5. Practical considerations—how things work in real and messy settings.
6. Design instrumentalities—design process reasoning and procedural knowledge.

These categories interact with knowledge from other sources, especially from science, to enable the design of new things, processes, and events. However, it is not always practical to separate out sources of knowledge, because science and design operate hand in hand in everyday problem solving, Vincenti underscores the importance of beginning to see design research as a knowledge-producing activity on a par with explanatory research.

Design

What Is Design?

Design research is historically the least well known and least well understood of the triad in Fig. 1, but interest in design research is increasing. Tough-minded experimental, sampling, and survey designs are used, and other aspects of the design of measurement instruments, including new types of items or questions, and computer-administration, are increasingly common, but the bulk of these works take their perspective from the framework of scientific research; the focus here is on discussing design research in more detail to provide an alternative viewpoint. Design research is emphasized here as a distinct knowledge-producing activity that discovers processes, principles, and structural concepts essential for the production of the technological tools and devices used in explorational research, explanatory research, and in design research.

For a long time, design has been deemphasized among other ways of knowing. Aristotle, in *Nichomachean Ethics*, described five kinds of "intellectual virtues," for which he used the term knowledge. Considering the first three of these, scientific knowledge is analytical, deductively provable, and teachable, whereas art and design, or "making," were regarded as chancy, variable, and intuitive, thus not teachable. Practical knowledge, including ethics, was also variable, not deductively demonstrable, and a virtue, not an art. Aristotle dealt much with practical knowledge or political science, which he claimed was that kind of knowledge most related to the most good for the most people.

The axiomatic method of deductive proofs developed and stressed by Aristotle became associated with science, but modern writers such as Herbert Simon and Buckminster Fuller have reasserted it and other mathematical methods in efforts to establish the foundations of a design science. It is through mathematical, logical, and conceptual rigor that Simon hopes that design will be accepted within academic circles: "In terms of the prevailing norms, academic respectability calls for subject matter that is intellectually tough, analytic, formalizable, and teachable. In the past, much, if not most, of what we knew about design and about the artificial sciences was intellectually soft, intuitive, informal and cookbooky."

The term "design" here thus refers to knowledge-producing studies and experiments into (1) the act of designing, (2) the design processes and generative principles for design, and (3) study of the structural properties of designed things. Design consists of the structuring of time and/or space in order to achieve specified purposes within the bounds of given constraints and to the level of given criteria. This involves the arrangement of materials and/or events to transfer or translate energy or information with the intention of producing a specific, measured result. The varieties of design knowledge previously described are produced through design research. Simon emphasizes the importance of giving design research equal standing with explanatory, or scientific, research: "The proper study of those who are concerned with the artificial is the way in which that adaptation of means to environments is brought about—and central to that is the process of design itself. The professional schools will reassume their professional responsibilities just to the degree that they can discover a science of design, a body of intellectually tough, analytic, partly formalizable, partly empirical, teachable doctrine about the design process."

Some social scientists have puzzled over why many graduate students and advisors prefer to engage in hypothesis testing rather than instrument design and development. Warren Tryon, in 1996, noted that such investigators may recognize that defective instrumentation limits their inferences, but that nevertheless, they continue to engage in hypothesis testing without prior instrument development. Tryon explains this phenomenon by pointing out that test development is not as scientifically respected as hypothesis testing is, even though test development entails construct validation, which is a highly theoretical enterprise. The neglect of design as a knowledge-producing activity has been changing slowly. An Internet search on the term "design science" shows that it has been incorporated forthrightly in fields as disparate as engineering design, public administration, and education.

Establishment of a Science of Design

Simon lists the "tough-minded, analytic, formalizable, and teachable" subjects he thought might become a part of a design science curriculum. He encourages rigor and discipline by associating design research with techniques for utility theory and statistical decision theory for evaluating designs; linear programming, control theory, and dynamic programming for choosing optimal alternatives; imperative and declarative logics for a formal logic of design; heuristic search for design alternatives, theory of structure, and organization of designs; and alternate representation of design problems. The term "design science" has more often been used in disciplines influenced by his work than has the term "artificial science." Simon emphasizes that natural science is knowledge about natural objects and phenomena, so an "artificial" science would contain knowledge about artificial objects and phenomena—and the disciplined activities by which artifacts are designed and developed. Simon believed that design is central, and should be at the core of every liberal education: "If I have made my case . . . the proper study of mankind is the science of design, not only as a technical discipline but as a core discipline for every liberally educated person."

Another influential 20th-century design scientist, Buckminster Fuller, promulgated the concept of a "comprehensive anticipatory design science" to seek solutions for global problems and to preserve life and progress on "spaceship earth." As does Simon, Fuller points to rigorous theory of structure and organization at the heart of design. Unlike Simon, he invented synergetic geometry as the heart of his theory of structure. To Fuller, the idea was not to set out to invent or design something. As quoted by a colleague and interpreter, Amy Edmonson, in 1987, he "did not set out to design a geodesic dome." Rather, he sought to "tap into the exquisite workings of nature," to "discover the principles operative in universe," and then to apply "generalized principles to produce artifacts useful to humankind." Fuller's observations of nature and explorations into a geometry of space different from what he (and we) learned in school reveal a geometry of spatial forms that is practical and empirical. Fuller's attention to "universe" and its general principles and constraints on what can exist therein guides his designs, and has been amplified in the *Design Science Series* of books on natural structural principles in design. Many other writers besides Simon and Fuller provide ample evidence that design is not restricted to the chancy, the variable, and the intuitive as Aristotle stated, nor is it unteachable.

Relationships among Explore–Explain–Design

It is customary for scholarly writers to end their works with the statement that more research is needed to answer questions generated by the results of the research being reported. Figure 1 illustrates the relationship of the three knowledge-producing activities—explore, explain, and design—as a set of two-way arrows, to show that the results of any one type of research can lead to questions answerable by any of the other types. A few examples are given at this point to show that this relationship has important implications for research of each type.

Exploratory Research Leads to Explanatory Research

As the activities of natural history measure and catalog natural phenomena, patterns become evident, requiring explanations of causal relationships, origins, and interdependencies. For example, when paleontological research on both sides of the Atlantic revealed the types of prehistoric animal and plant life that had once inhabited those regions, a pattern of relationship became evident to the scientist Alfred Wegener, and to others, that ran directly contrary to the prevailing explanatory theories of the time regarding the origin and history of the continents. To Wegener, the only explanation that fit all of the observations was that the separate continents had been joined at one point in the past but had drifted apart. Though his explanatory theory of continental drift was dismissed by opinion leaders, additional evidence that supported Wegener's theory appeared many years later when the Atlantic sea floor was being mapped for the first time. Clear signs of sea-floor spreading gave a new relevance to Wegener's theory. What is important here is not that Wegener's theory triumphed, but that it was the description—the sea-floor mapping—of natural phenomena that led to the ultimate reconsideration of Wegener's theory.

Exploratory Research Leads to Design Research

When humans find patterns in nature's mechanisms that they can copy, they do so even if the details of the pattern's internal operations are not fully understood. For thousands of years of recorded history, humans have created and used artifacts and intervention processes that mimic natural phenomena, with little or no prior research that could produce an explanation of the effect. The discovery of medicines and medical procedures through the ages has tended to follow this route. Even today, folk remedies and natural healing concoctions derived from locally found substances supply the beginning point for the refinement of pharmaceuticals. Until recently, serendipity, or dogged persistence in trial-and-error, was a major source of new drug discovery. Interestingly, this is being replaced with principled anticipatory design research.

Early in the industrial revolution, some manufacturers proceeded by trial-and-error in the absence of direction from explanatory science. The excellence of early Wedgwood china traces its origins to tireless design experiments involving clay and various minerals from Wedgwood's own farm. Writer Jenny Uglow has ascertained that "ceramics has always been a mix of science, design, and skill, and every good potter was in a sense an experimental chemist, trying out new mixes and glazes, alert to the impact of temperatures and the plasticity of clay." This, along with various patterns of aging, working, and firing using a variety of glazes, ultimately produced formulas that looked attractive and survived in their environment of use better than did the products of competitors. In some cases, then, exploratory research turns up useful structures and processes that become fundamental technologies. For example, the transistor effect was discovered by chance because of studies on the impurities in crystal substances.

Explanatory Research Leads to Exploratory Research

As explanatory theories gain support through scientific research, inferences using the theory lead to the expectation of finding as-yet undetected natural phenomena. The early stages of success of a theory constructed and initially supported through explanatory research often require additional exploratory research to develop further support. Armed with predictions of theory, researchers can go in search of specific phenomena that are predicted but as yet not observed. Exploratory research motivated by the explanatory theory of Mendeleyev's periodic table of the elements led to the discovery of several elements, most notably the trans-uranium elements, one of which was named after Mendeleyev, in honor of his discovery. Likewise, improved astronomical instruments developed decades after Einstein's relativity theory allowed astronomers to observe the phenomenon of gravitational lensing, just as Einstein's theory had predicted.

Explanatory Research Leads to Design Research

As descriptive theories gain support through scientific research, using principles from the theories to exercise control to produce specific outcomes becomes a possibility. Though the ideal of science is often expressed as seeking knowledge for the sake of knowledge, it is also true that the foundation of our technology-based economy relies on the ability to turn principles into products. Examples of this transfer are common. One of the most interesting and little recognized examples, however, is the basing of the modern technology of surgery on John Hunter's realization that the human body's own healing powers could be relied on following surgery, or what Sherwin Nuland has described as a "controlled injury" to the body. By determining what injury to inflict on the self-healing system, surgeons today recruit the body's natural processes to correct painful and deadly conditions. The study of surgery can be seen in this light as the design of injuries to inflict for given unhealthy body conditions.

The pattern of transfer from explanatory research to design research is so common that what is being described here as design research is often referred to as "applied science." It is the unquestioning acceptance of this belief and a lack of awareness of the complexities of bridging the science–technology gap that has, until recently, given technology an image subordinate to, and as simply a receiver of, the benefits of science. One of Simon's main points is that the image of applied science has obscured the knowledge-producing nature and complexities of design research and has masked the full range of technological knowledge that is required for the design of artifacts and intervention plans. Historically, this has placed design research in the shadow of science, rather than giving it equal regard and proportionate study.

Design Research Leads to Explorational Research

Design research leads to exploratory research when a design study or a design experiment produces results that are unexpected and even undescribed. The first objective after such an event is to explore the region around the undescribed phenomenon, in search of an entire class of similar new phenomena. This occurred in the 1980s when chemists examined the residue produced by intense laser heating of small sections of graphite, or pure carbon. What was found was a family of pure carbon compounds containing up to 60 or more atoms per cluster in the shape of a sphere—a form of carbon never before recognized. Philip Ball has described the process of discovering the compound's shape: "Smalley [one of the chemists] located a book on Buckminster Fuller's work, *The Dymaxion World of Buckminster Fuller* by Robert W. Marks, which he took home on the Monday evening to ponder. It is remarkable to find that, despite what we like to think of as the sophistication of modern science, some of the most significant insights can still be obtained by sitting down with a can of beer and fiddling with cardboard or ball-and-stick models. Yet this is how the structure of C_{60} was deduced." Once the structure of sixty-carbon compounds had been recognized and confirmed, and the C_{60} cage of hexagonal and pentagonal sides with 60 vertices was named "Bucky Ball," in Fuller's honor, more exploratory research was launched. An entire family of pure carbon compounds containing geometrically determined numbers of carbon atoms had to be explored. This included the "stable and cagelike" clusters of C_{32}, C_{50}, and C_{70}. This additional period of exploratory research was initiated by the unexpected discovery of previously unrecognized design research.

Design Research Leads to Explanatory Research

Design research leads to explanatory research when phenomena result from design studies or design experiments that cannot be explained. As methods for producing the buckminsterfullerines (as the class of chemicals became named) in quantity were perfected and published, the question of their chemical behavior became paramount, particularly the manner in which they encapsulate other atoms within a spherical cage, and the arrangements they take on under different conditions of creation. According to Ball, an entire subdiscipline of fullerene chemistry has formed just around the study of one class of these forms, the nanotubes.

Design Studies and Design Experiments

Definitions and a Brief History

The design study/design experiment movement has been growing rapidly in educational research, and the examples that follow will largely be drawn from that domain, but surely Simon was correct in stating that design methods cut across the professional schools of many disciplines. Education is but one of many social science and engineering professions that can use design studies and design experiments, and it is the one closest to the experience of the present authors. In the glossary at the beginning of this article, brief definitions were given of the terms "design studies" and "design experiments." More extended definitions are given here.

Design Studies

Two kinds of theoretical knowledge are sought in a design study; the first is a descriptive explanation of the processes operative in a domain (e.g., learning, growth, reversal of a disease state, improvement of an environmental condition), and the second is technological or design knowledge about how to create and implement the technological tools that control the treatment effects. Design studies take place in live settings, and are iterative, cyclical applications of a process of design, implement, evaluate, redesign. Design studies often aid in exploring a domain and possible treatments, and thus may be largely qualitative, producing narrative accounts of intense, iterative, often ideographic observations over each cycle. These accounts may perform the functions of exploration discussed in this article, and may begin to use theory to guide both the design interventions and the explanations of the processes operative in the domain. However, they may not provide adequate warrant for causal claims, nor do they fully rule out alternate explanations. For these research goals, design experiments are needed.

Design Experiments

All aspects of a design study are shared by a design experiment, except that heavy reliance on narrative methods in seeking the explanatory and design theories is supplemented with research using (and improving) the invariant measurement instruments designed to assess and track the learning or growth constructs spanning the domain. Because measurement in live settings has typically been very costly in time and attention, design experiments are becoming practical primarily because of new adaptive measurement technologies that can be implemented in live settings and can measure progress unobtrusively, seemingly as a fully integrated part of learning and instruction, not as separate events. Moreover, the perception of measurement as a separate accountability activity for grading, sorting, rewarding, and punishing can change to a useful helping function owned by the participants in pursuing valued attainment goals. All the time, the measurements are being used for research, and the teachers and others are participant evaluators in this research, not unknowing subjects. They are the ones who, in concert with researchers and developers having additional expertise, will participate in the redesigns, then will implement them during the next cycle. The measurement instruments evolve over the cycles of design, implement, evaluate, redesign, and come to embody an increasingly adequate descriptive theory of the processes operative in the domain. In addition, the technological devices designed to introduce and control the treatment effects are forthrightly described using the emerging layers and languages of technology in that domain. In addition, both quasi-experimental designs and randomized experiments may be used to warrant causal claims and rule out alternative explanations.

Some History

The roots of the idea of a design experiment in education can be traced back to the time before the accelerating growth of interest in the previous decade; documentation of this was provided by John Dewey in 1916 and

Robert Glaser in 1976, for example. Despite this longer history, most proponents writing today cite Ann Brown's 1992 paper on design experiments and Allan Collins' 1992 paper on a design science of education; the term "design experiment" was introduced in the latter paper. Both of these authors, as well as Glaser, cite Herbert Simon's work. There has been a substantial growth of interest in, and publications dealing with, design studies and design experiments since the 1990s. An entire issue (January/February 2003) of the *Educational Researcher* has been devoted to the design research movement.

How a Design Study or Experiment Proceeds

A study begins by selecting one or more live settings in which the phenomenon of interest may be observed. The setting may be one or more classrooms, or nonconventional learning settings, or it may be a medical treatment facility studying treatment options and ways to implement them, or a management initiative in a group setting, repeated with different groups at later times. A repeating cycle is selected. A semester or block schedule may be used for education, an end point in a medical treatment, or natural assessment points for the management activity. The repeating cycles are essential, because the earlier cycles serve as controls for later ones. (There may also be additional control or treatment groups in rigorous design experiments.) The goals of the research depend on the stage of maturity of the study in the domain of interest. The researchers may seek exploratory, explanatory, or design knowledge in any mixture. Good research design will, as usual, set forth the questions in advance. To be called a design study or experiment, design knowledge will be central among the research questions. In exploratory studies, the questions concern what design interventions exist already and which seem to be working well. As three examples of explanatory studies, the questions may concern the nature of the descriptive progression of learning or growth (in education), the progression of the disease state (in medicine), or the propagation of employee behaviors and associated results (in management). In all design research studies, design knowledge is sought, along with the implementation conditions that enable the designed treatments to yield their best results.

The procedure is to design four aspects of the study: (1) the measurement instruments to assess the progression (of learning, disease states, effective employee performance, and results), (2) the treatment properties (technology), (3) the evaluation/research design for the next cycle, and (4) the implementation plan for the next cycle. Once the measurement instruments are in place, even before a new treatment is introduced, a cycle can be run to collect baseline data. When all four designs are in place, subsequent cycles can be run, involving implementation of the latest version of the designs, then evaluation, followed by seeking improvements in the next cycle through design. Anything may be improved, including the measurement instruments, the treatment, the implementation methods, or the evaluation/research plan. If measurement instruments are changed, equating must be performed to assure comparability with measures taken during previous cycles, or a new series of cycles launched. Because new technologies are making possible continuous measurements all along the way—that is, along the explanatory theory of progression of learning, disease progression or improvement, or employee performance (keeping to our three examples)—evaluation and research data collection are also continuous, and their effects (good and bad) can be attributed to stages along the path of progression. The outcomes of a series of cycles investigated in this type of research are answers to the knowledge-seeking questions stated at the outset of each study, whether explore, explain, or design. An outcome of substantial importance is a documented argument for the validity of a total treatment method, including the technology used and "best practices," i.e., the manner of implementing it to get the best results. When design experiments are used, this validity argument may include evidence to back causal claims for the efficacy of the treatment in leading to desirable outcomes.

Terminology is still in flux. The term "design experiment" has often been appropriated by those interested in intuitive and reflective types of action research in education, in which the teacher takes the lead role, rather than in a partnership role with researchers. The teacher journals personal reflections on possible designs and outcomes as a way to improve practice. Theory, measurement, and the validity of causal inferences are not often important in such studies, but are very important issues in research generalizable beyond single practitioners. For this reason, the term "design studies" can be used to avoid the term "experiment" so as not to confuse these studies with randomized trials in social and behavioral science. This usage was suggested by Richard Shavelson and co-authors, in the previously mentioned issue of *Educational Researcher* focusing on design research, to distinguish single-practitioner studies from true experiments. Influential writers on experimental methods define and emphasize two main types of experiments, i.e., the randomized experiment and the quasi-experiment. Both types of experiments are designed to test descriptive causal hypotheses about manipulable causes. Thus, when the definition of experiment or quasi-experiment holds in a design study, it may justly be called a design experiment.

Design Experiments and Experimental Design

It is useful to consider the nature of experimental design to appreciate fully what can be accomplished with a principled design experiment using invariant measurement scales. In publications extending back to 1963, Donald Campbell and Julian Stanley, and later Thomas Cook, William Shadish, and many others, have given guidance much used over the years in how threats to the validity of causal inferences can be reduced by employing good experimental design. Classically, three pre-experimental designs, three true experimental designs, and 10 quasi-experimental designs have been discussed. There has been a shift away from significance testing in fairly recent times in favor of effect sizes, but the need to consider threats to the validity of inference has not abated. The simplest inference is that the introduction of treatment X did indeed cause the change in observation (measure) O. The emphasis here is that there is a need for design disciplines to assure both that measurement O and treatment X do indeed involve the constructs of learning progression along the pathway to greater expertise. Design theories to guide the development of instructional treatments can succeed far better to the extent that they have a descriptive account of the progress that learners typically follow in moving to higher levels of knowledge and expertise. Figure 2 depicts a design intervention into a natural process. In learning and instruction, the most useful explanatory account would describe the sequence of progressive attainments in the particular learning domain of interest. This descriptive knowledge guides the design of both the measurement instruments, to determine outcome effects (O), and the instructional treatments (X). Instructional-design theories (a term elaborated in a series of books by Charles Reigeluth) are now available to guide the development of controlling technologies for the treatment. Designing X and O in a principled manner can help give this design the six aspects of construct validity discussed in numerous publications by validity theorist Samuel Messick. The descriptive account (or theory) of progressive attainments in learning the increasingly difficult tasks in a learning domain, along with construct-linked measurement scales of learning and growth, along with a validity argument, can assure that the O indeed measures valid levels of progress in the construct. For simplicity, designs that require adaptation to individual difference measures are not discussed here. It is sufficient to note that good quasi-experimental and experimental designs exist for examining hypotheses of how treatments might differ for subgroups of individuals receiving the treatment differentially, depending on their individual profiles.

To understand the sequential nature of a design experiment, and how results from prior cycles can serve as controls for later cycles, consider a simple pretest/post-test design, one of Campbell and Stanley's three true experimental designs. It can be diagrammed as follows, where R means random assignment to either group, and the sequence of events in a single row (e.g., $R \rightarrow O \rightarrow X \rightarrow O$ for row one) is the temporal sequence of events for the experimental group:

$R \quad O \quad X \quad O$ Experimental group
$R \quad O \quad \quad O$ Control group

A gain score can be calculated as $O_{\text{post-test}} - O_{\text{pretest}}$ if the measurement scale O supports equal intervals, and other precautions in using gain scores have been followed. Then we may graph the gain for the experimental group and compare it to the gain for the control group.

Consider a series of such experiments, each with a well-documented change in the design specifications for treatment X, or perhaps only changes in the implementation procedures for the treatment.

The design experiment in Table I assumes that all repeating cycles of principled design intervention, X_1–X_5, take place in the same class in an educational institution (or group of similar classes using the same treatments and outcome measures, X and O). The designers of the experiment in Table I altered the treatment condition X_i each time, being careful to conform each design change to a prescriptive theory and assuring that each version of X was well documented. No R for random assignment is shown in this table, because it is neither possible nor necessary for random assignment to occur in these classes. Assume that both the registration procedures of the educational institution and the ethical considerations bar such a practice. Thus we must depend on the sample-invariance and interpretive invariance properties of outcome measure O, and on repeated near-replications with other groups in additional studies, presumably from the same or similar population to substitute for random assignment. The construct-linked variable of interest, O, has the appropriate properties to reduce the risks to internal validity that randomization was assumed to reduce.

Table I Repeating Cycles of a Principled Design Experiment Using Invariant Scale O

Parameter	*Cycle* 1	*Cycle* 2	*Cycle* 3	*Cycle* 4	*Cycle* 5
Baseline measure O_0	X_1	X_2	X_3	X_4	X_5
	O_0	O_1	O_2	O_2	O_3
Control for cycles 1–5	Control for cycles 2–5	Control for cycles 3–5	Control for cycles 4–5	Control for cycle 5	Control for future cycles

Repeated measurements of subsequent groups over the cycles of a design experiment give further basis for ruling out the effects due to peculiarities of one group of students during one semester. But the students who flow through the classes of any one institution are unique and slanted in their own way, so population representativeness has not been obtained. For this function, randomization is a poor tool as well. To obtain evidence of generalizability to other groups, with other language, gender, racial, and special conditions, it is necessary that design experiments be set up in other locations wherein samples of students with subsets of these other characteristics abound. Nevertheless, it is no small benefit to causal inference to be assured that each semester's group was either equivalent on the highly interpretable pretest/post-test measure O, or of known deviation. It is of no small benefit to know that the use of a gain score is appropriate because we have achieved a close approximation to equal interval properties in the measurement scale(s) O. Finally, the quest for the validity of causal inference is, in the framework of a design experiment, set in a larger context similar to the quest for total quality management. After all, there are many aspects to treatment X and its procedures for implementation. To which specifically is the causal inference to be made? An outstanding result is a tribute to the entire group of people and the roles they assumed, the rules they followed, and the tools they used to administer the treatment. An outstanding result is evidence of the possibility that such results can be obtained by managing the implementation of the treatment well, and if in one group, why not in another?

In addition to substituting known starting positions on an invariant scale for random assignment, the design experiment in Table I also replaces the control group, using a series of comparisons to previous groups. Starting with the baseline measurement on O_0, each succeeding cycle can compare its outcome not only to this baseline measure, but also to each of the preceding outcomes, using invariant scales O_0–O_2 (the subscripts indicate that, in this experiment, it was necessary to modify the outcome measure only twice, by adding or deleting tasks, whereas during the same five cycles, the treatment X was modified five times, after each cycle and before the next one). By the specific objectivity property of scale invariance, we can add or subtract questions or tasks from the most recent version of the instrument each semester. Then, after assuring that the versions are equated to the same meaningful scale each semester, we can make inferences that the version of X used that semester was likely the main cause of whatever improvement in gain score was observed.

New guidelines to update good experimental design prescriptions must be developed for principled design experiments. Scientists writing in Campbell and Stanley's era could not have foreseen today's opportunities for unprecedented control of a complex interplay of many treatment variables, possible in live settings through interactive technology. Neither could they have envisioned sophistication of online measurement, nor new measurement methods that can provide close approximations to the invariance properties needed to realize the design experiment scenario. Certain of Campbell and Stanley's quasi-experimental designs are similar to parts of this design experiment scenario, but scientists of their day could not have anticipated the full extent of reliable replication (using technology) with well-documented design changes. Advances in the technology of learning, measurement, and management of complexity changes how well we can control each of the 12 common threats to validity, of concern to these pioneering scientists. Moreover, these advances both introduce new threats to validity and help to find the means to reduce their effects.

It is interesting that the use of the term "validity" by experimental design writers and validity concepts from psychometrics, such as Messick's unified validity concept, do not coincide. For example, the idea of "external validity" is used in entirely different ways. By bringing these two views together through validity-centered design and the methods of principled design experiments, perhaps we are taking a step toward bridging the gap between the "two disciplines of scientific psychology" described by Lee Cronbach.

Conclusion

The three knowledge-producing activities—explore, explain, and design—have been placed into a common context in this article in hopes of distinguishing among the different types of question each activity addresses; at the same time, it has been shown that it is not research methodology that allows them to be separated. Contributions of each form of knowledge-seeking to the other forms have also been described, and examples have been given to show that though exploratory research and explanatory research have longer formal histories than design research has, all three forms of research continue today to contribute to unanswered questions, and none has outlived its usefulness, especially not exploratory research, which continues today at what might be considered an accelerated, rather than a diminished, rate.

The types of research and knowledge-seeking have been described in proportion to the measurement theory and technique concerns addressed in this volume. In particular, measurement has been related to the health and progress of design research and exploratory research to correct what can be viewed as a current underemphasis in those areas caused by living in the shadow of science. The intent here has not been to isolate the three enterprises from each other, but to show their essential contribution to each other: each produces results in the form of answers that stimulate a continuous stream of research questions in all three areas. The three are discriminated on the basis of the kinds of questions they address and the types of knowledge they produce, but not definitively on the research techniques employed.

Progress will continue to be limited in the human sciences and in design as long as metrics used in research are incommensurable from study to study. In discussing design studies and design experiments, the emphasis here is that rigorous design experiments are now possible using well-designed instrumentation for administering both experimental treatments and measurements in live settings. This depends not only on development of commensurable measure-building techniques, but also on use of emerging technologies that give access to data collection and analysis tools previously not available.

See Also the Following Article

Research Designs

Further Reading

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *J. Learn. Sci.* **2**(2), 141–178.

Campbell, D. T., and Stanley, J. (1963). *Experimental and Quasi-Experimental Designs for Research.* Hougton-Mifflin, Boston, MA.

Collins, A. (1992). Toward a design science of education. In *New Directions in Educational Technology* (E. Scanlon and T. O'Shea, eds.), pp. 15–22. Springer-Verlag, New York.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *Am. Psychol.* **12**, 671–684.

Cronbach, L. (1975). Beyond the two disciplines of scientific psychology. *Am. Psychol.* **30**, 116–127.

Dewey, J. (1916). *Democracy and Education.* Free Press, New York. [Cited in Tanner, L. N. (1997) *Dewey's Laboratory School: Lessons for Today.* Teachers College Press, New York.]

Edmondson, A. C. (1987). *A Fuller Explanation: The Synergetic Geometry of Buckminster Fuller.* Barkhauser, Boston, MA.

Glaser, R. (1976). Components of a psychology of instruction: Toward a science design. *Rev. Edu. Res.* **46**(1), 1–24.

Hubka, V., and Eder, W. E. (1996). *Design Science: Introduction to Needs, Scope and Organization of Engineering Design Knowledge.* Springer, Berlin and New York.

Kelly, A. E. (ed.) (2003). Theme issue: The role of design in educational research. *Edu. Res.* **32**(1).

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement: Additive and Polynomial Representations.* Vol. 1., Academic Press, New York.

Luce, R. D., Krantz, D. H., Suppes, P., and Tversky, A. (1990). *Foundations of Measurement: Representation, Axiomatization, and Invariance.* Vol. 3., Academic Press, San Diego, CA.

Messick, S. (1989). Validity. In *Educational Measurement* (R. L. Linn, ed.), pp. 13–103. Macmillan, New York.

Messick, S. (1995). Validity of Psychological Assessment. *Am. Psychol.* **50**(9), 741–749.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicat. Res.* **45**, 35–44.

Pickstone, J. V. (2001). *Ways of Knowing: A New History of Science, Technology, and Medicine.* University of Chicago Press, Chicago, IL.

Reigeluth, C. M. (ed.) (1999). *Instructional-Design Theories And Models: A New Paradigm of Instructional Theory.* Lawrence Erlbaum Assoc., Mahwah, NJ.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Houghton-Mifflin, Boston, MA.

Simon, H. (1969). *The Sciences of the Artificial.* MIT Press, Cambridge, MA.

Suppes, P., Tversky, A., Krantz, D. M., and Luce, R. D. (1989). *Foundations of Measurement: Geometrical, Threshold, and Probabilistic Representations.* Vol. 2., Academic Press, San Diego, CA.

Tryon, W. W. (1996). Instrument-driven theory. *J. Mind Behav.* **17**(1), 21–30.

Vincenti, W. G. (1990). *What Engineers Know and How They Know It: Analytical Studies from Aeronautical History.* Johns Hopkins University Press, Baltimore, MD.