Experimental Research in Construction

Leonhard E. Bernold, M.ASCE[1]; and Tai Sik Lee, M.ASCE[2]

Abstract: Humans' unique use of tools originated with their capability to observe, to link causes with effects and to conduct trial and error experiments. Experimental research today is still using these same basic elements albeit augmented by sophisticated tools and methods. Researchers conducting scientific experiments in the construction arena, however, face a "harsh" environment to work in. This paper draws from 20 years of experimental work in construction and engineering education. It provides a short historic background before discussing a framework useful to categorize the various ways researchers conducted experiments that provided meaningful results. Many examples are used to underscore the most important points.

DOI: 10.1061/(ASCE)CO.1943-7862.0000085

CE Database subject headings: Experimentation; Research; Comparative studies; Construction management; Engineering education.

Author keywords: Experimental research; Controlled experiments; Comparative evaluations; Pilot study; Consent forms.

Introduction and Background

"The more I know, the more I know that I don't know," is a frustration uttered by many researchers especially those involved in experimental research. Nevertheless, one can safely assume that this kind of intellectual pursuit is as old as humans' capability to observe and to establish causal relationships. Observing does not simply mean watching a bird fly but to understand the phenomenon of flight that allows them to traveling long distances in the air. Discerning natural occurrences helped them to develop not one but various methods to start a fire or to develop a series of medicines to treat an illness. A major step to establishing "tighter" and much more reliable cause-effect associations was the emergence of mathematics. Ibn al-Haytham, an early Iraqi physicist who lived from 965–1039, established the first regimented approach to conducting research. Al-Haytham's recommended procedure listed seven steps of inquiry is still valid today: "(1) observation; (2) statement of problem; (3) formulation of hypothesis; (4) testing of hypothesis using experimentation, (5) Analysis of experimental results; (6) interpretation of data and formulation of conclusion; and (7) publication of findings" (Bradley 2006). While not the first in recognizing its importance, Ibn al-Haytham stipulated the use of experiments and the capturing of data as a means to prove or disprove a hypothesis.

The final breakthrough for scientific methods was paved by the uneducated Galileo Galilei. After a long standoff, he forced the establishment to accept "a posteriori knowledge" that had been discovered through the use of scientific methods but contradicted long established dogmas. Another "stepping stone" leading to modern experimental science was provided by a contemporary of Galilei namely the Frenchman Rene Descartes born in 1596. Descartes not only received a solid education but also served in the French military as an Ingenieur Civil. His most important contribution was the *Discourse on the Method of Rightly Conducting the Reason, and Searching for Truth in the Sciences (Discours de la méthode pour bien conduire sa raison, et chercher la verité dans les sciences)* which he published in 1637. It included the memorable assertion, "Je pense, donc je suis" which translates into "I think, therefore I am." According to his biographer Desmond Clarke (Clarke 2006), Descartes emphasized but cautioned about the value and importance of experiments to advance knowledge. Lamenting the complexity of nature as overwhelming he stressed the need to break complex investigations into their simplest components and to commence experimental discoveries with the simplest steps. Most vigorously, however, he demands that every researcher document design and conduct results of experiments without leaving out any detail. The latter requirement is based on his frustration caused by not being able to verify the work of other "discoverers" of the time as they commonly provided very little in way of descriptions beyond their claims. Descartes found many followers who published a steady stream of methods that improved the validity of experimental findings mostly through proper design and evaluation of the measured data. The basic premise, of course, was to make the experiment reproducible so that the results could be reproduced if the descriptions were followed.

The late 1920s and 1930s brought serious of behavioral experiments that revitalized Descartes' call to skepticism concerning the meaning of data. What is referred to as Hawthorne effect or observer/experimenter bias began with a set of illumination experiments at the Hawthorne works of the Western Electric Company in Chicago. The goal was to find the lighting condition that allowed the workers involved in winding coils and assembling relays to be more productive. After a long study that involved the workers themselves, the light bulbs for a test group were made brighter while a comparison group had to continue with the old ones. Strangely, the production of both the test as well as the control group improved. Equally unexpected was the drop of the productivity after researchers had left. In fact, the productivity before and after the change of the lighting system was exactly the same. It was subsequently discovered that the researchers' actions

[1]Professor, Dept. of Civil and Environmental Systems Engineering, Hanyang Univ., 426-791 Korea. E-mail: Leonhard.bernold@gmail.com

[2]Professor, Dept. of Civil and Environmental Systems Engineering, Hanyang Univ., 426-791, Korea. E-mail: cmtsl@hanyang.ac.kr

Note. This manuscript was submitted on August 6, 2008; approved on April 30, 2009; published online on May 4, 2009. Discussion period open until June 1, 2010; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Construction Engineering and Management***, Vol. 136, No. 1, January 1, 2010. ©ASCE, ISSN 0733-9364/2010/1-26–35/$25.00.

Table 1. Procedure to Conduct Experimental Research

	Main activity
1.	Definition of a vital question for which no answer exists
2.	Establishment of the state-of-the-art (data, information, expertise, etc.)
3.	Formulation of subquestions each accompanied by a hypothesis (a priori)
4.	Design and development of experimental procedures and tools
5.	Design and calibration/validation of methods needed to measure and observe
6.	Execution of experiments accompanied by data collection
7.	Statistically sound analysis of data collected during experiments
8.	Interpretation of data to establish cause-effect relationships that either prove or disprove the hypothesis
9.	Conclusions founded on the interpreted data and publication of results
10.	Stipulation of new questions that have arisen

(e.g., the interactions with the workers) impacted the outcome of the experiment. Feeling as being a part of the scientific study the workers changed their original behavior only to fall back when the original factory management team returned.

Finally, a "lethal" problem that every experimental researcher faces is bias. Expecting and whishing for a certain outcome, a biased researcher tends to accept data, even false data, that supports the expected outcome while discounting what does not fit. Only a critical and open examination of the data by a range of experts will provide an adequate barrier to biased conclusions.

Process of Empirical Research in Science

As mentioned earlier, the object of empirical research is to establish "a posteriori" knowledge, meaning scientific knowledge that is backed up by evidence rather than established through intuitive reasoning. In natural as well as in social sciences, the scientific experiment is the main tool of empirical research not only to discover or explain our world but also in proving new theories. The tenet of scientific discovery in engineering is a logic that is based on the causality of nature where a phenomenon can be explained by establishing relationships to its causes. As even early thinkers have pointed out (e.g., Ibn al-Haytham) those cause-and-effect relationships can be extremely "tricky" to discern without offering a critical mind to doubt it. As a result, Ibn al-Haytham's 7 steps have been more clearly defined over time and constitute the accepted procedure to conduct scientific research involving experiments.

Steps 4–6 in Table 1 are uniquely tied to experimental work while other steps are part of other methods of research as well. Consequentially, the remainder of this paper will focus on discussing the fundamental steps of: (1) design of experiment, (2) testing of measuring methods, and (3) implementation of experiments with data collection.

Elegant Case Study: Curing Scurvy

Medicine is a field that has probable the longest history in conducting experiments, unfortunately until recently only on a trial a error basis. Nevertheless, the history of medical research has many excellent examples of experimentations that were not only well designed but led to important discoveries. One such case is the experiment conducted by James Lind in 1747, who, as a British Royal Navy surgeon, was faced with a serious killer disease for which so many tried to find a cure. His approach to this problem was the first controlled scientific experiment that has been thoroughly documented (Simon 2002). He established a hypothesis that at least one of six "talked about" therapies, each involving the intake of some kind of food, would actually restore to health of diseased sailors. For his next long voyage, he loaded the food stuff for the six treatments that he had selected to have available when sailors would get the disease. Fig. 1 depicts his plan for finding a cure.

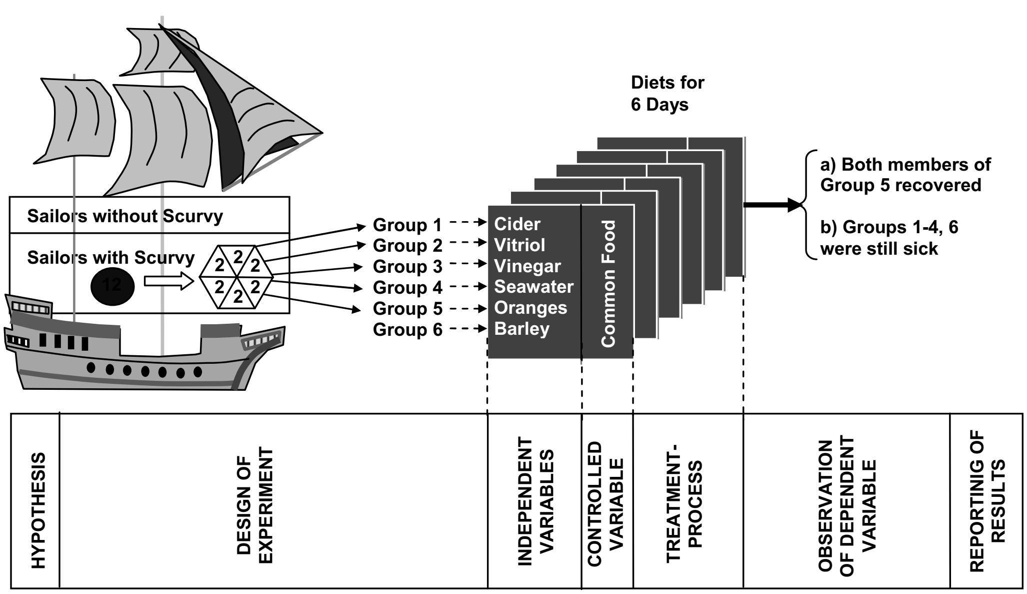

Fig. 1. Design of the first controlled experiment to find cure for scurvy

J. Constr. Eng. Manage., 2010, 136(1): 26-35

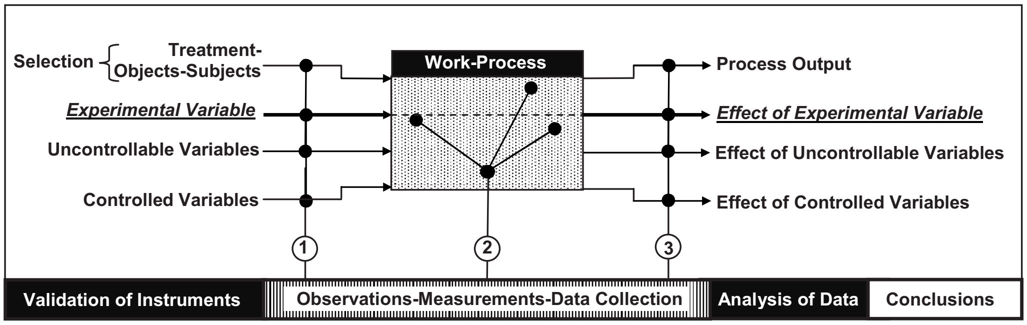

Fig. 2. Architecture of experimental process to establish causality

The life on a navy ship in the 18th century provided for a very controlled environment in that every sailor faced pretty much the same hardship be it weather, food, or physical work. The controlled experiment began when he had the first 12 sailors in the sickbed. Why not start the experiment earlier? The surgeon's objective was not only to improve the reliability of the outcome, by making six groups of two sailors, but also creating a uniform environment for all subjects. All twelve were treated equally except the extra food that was randomly assigned to the sailors. Of course, other sailors who also got sick provided the control group. Should both members of a group recover from scurvy it would be hard to claim that it was a coincidence.

As Fig. 1 indicates, each group was given a daily portion of common food plus a supplement from the assigned "medicine," called the independent variable. In this experiment, observing the effect of the different input variables was straight forward. After only six days, both members of group that was receiving the oranges had recovered. Unknown to the surgeon at that time, vitamin C from the oranges was and still is the key ingredient to prevent and fight scurvy.

While this case comes from outside the construction arena, its simplicity helps one to easily identify the key characteristics of controlled experiments. The next section of the paper will now expand this base model to the more complex world of construction.

Process Model of Scientific Experiments

Modelers of construction processes or supply chains recognize the input-output nature of a scientific experiment. The goal is to measure the relationships between an input and an output characteristic. Fig. 2 presents a model of the core relationships of process oriented experiments.

Again, the goal of a scientific experiment is to establish empirical evidence of a relationship between an independent or experimental variable and a dependent variable that is being effected by it. As Fig. 2 reemphasizes, the key to this process are observations and measurements with calibrated instruments at Stages 1–3 to establish causalities that are cleared of extraneous effects caused by uncontrollable variables. Also highlighted in Fig. 2 is the fact that measurements are not restricted to a before-and-after scenario but that may involve the process itself. What guides the final design of an experiment is the need for data that either proves or disproves the hypothesis "a posteriori."

The process model in Fig. 2 indicates three different data-collection points for: (1) pre, (2) In, and (3) postprocess measurements. Applied to the scurvy healing experiment one can see that the James Lind used sick sailors as subjects of his treatment process while the six different food supplements served as experimental variables. The controlled variables consisted of the daily food rations and the un-controlled variables were the temperature, humidity, wind, etc. The surgeon made is daily rounds and recorded what changes he observed in the 12 sailors. It is common practice to record as much as possible during the experimental process from independent "sensory" points. The outcome of the treatment were 10 sick sailors and two who had recovered. The effect of the experimental variable was clear in that the treatment with oranges led to the recovery of both members of group 5. Since all 12 sailors had experienced the same uncontrollable and the one controlled variable, their effect on the recovery was nil.

Measuring Construction Work

Measuring the effectiveness of a new device, method, or material in the construction arena is not as straight forward a task as investigating the effect of a food on curing scurvy. The most important differences are the many extraneous factors that interfere with the independent variable and the difficulties in measuring meaningful values with sufficient reliability. Pioneers in this arena were Frederick W. Taylor as well as Frank and Lillian Gilbreth. While Taylor focused on timing work elements while the Gilbreths categorized basic work motions as ineffective or effective leading to what is now called motion studies or predetermined time systems that calculate work time based on needed motions. L. H. C. Tippett later addressed one of the key weaknesses in both namely the large amount of time needed for the studies to be done by highly trained personnel. Instead of measuring the continuously he recommended to randomly pick time instances to take "snapshots" of what is going on at that particular time. Today, we refer to it as activity, work, or productivity sampling. Despite the emergence of digital tools that alleviate the burden of a continuous time study, work sampling is still the main workhorse in industries where the pace of the work is not controlled by robots or a conveyor. Ergonomic studies, on the other hand, have developed standard procedures to quantify the amount of "stress" work has on humans while quality sampling can be used to measure output quality.

Work Sampling

A good description of this method can be found in a paper by Salim and Bernold (1994). The purpose of this method is to quantify the activities during the work process and is indicated in Fig. 2 as Measurement 3. A critical step in the process is to define all work-elements that are relevant to prove the hypothesis and organize them into productive, contributory, personal or wasteful

J. Constr. Eng. Manage., 2010, 136(1): 26-35

Table 2. Main Experimental Approaches in Construction

Method	Object of study	Procedure	Results
1. Pilot test of devices and methods	Device A Method B	Operation (Lab-Field) → Measure Implement (Lab-Field) → Measure	-Performance -Performance
2. Passive observation (case study)	Subject-group or operation A Subject-group or operation B	(Events) → Measure (Events) → Measure	-Functioning of A/B -Performances -Variance
3. Controlled experiment (direct comparison)	Subject (group) A Subject (group) A/B	Pre-M → T_{Trad} → Post-M Pre-M → T_{Mod} → Post-M	-Improvement by A -Variance -Covariances
4. Randomized experiment (conventional)	Random subject (group) X, Strata 1 Random subject (group) Y, Strata 2 Random subject (group) X, Strata 1 Random subject (group) Y, Strata 2	Pre-M → T_{Trad} → Post-M Pre-M → T_{Trad} → Post-M Pre-M → T_{Mod} → Post-M Pre-M → T_{Mod} → Post-M	-Correlations -Improvement within strata -Variances -Covariances
5. Four group experiments (minimized Hawthorne)	Random subject (group) X Random subject (group) Y Random subject (group) Z Random subject (group) G	Pre-M → T_{Trad} → Post-M Pre-M → T_{Trad} → Post-M T_{Mod} → Post-M → Observ.	-Improvements of Strata 1 members -Variances -Covariances

groups. The number of random observations can be calculated and are based mainly on the confidence level (e.g., 95%) and the size of the allowed errors (e.g., 5%). This method is easy to learn and use, and it provides more operational detail than historical data. Other work measurement techniques can be used in parallel to ensure that the work sampling results are reliable.

Taxonomy of Experimental Methods for Construction

Collecting a lot of evidence is necessary but not sufficient to discovery and the establishment of new knowledge. As the Hawthorne tests that the experiments with placebo medicines show, what looks like clear relationships between independent and dependent variables is but a "mirage." The productivity of the coil winders did not improve because the light was changed. Executing a series of repetitive experiments within a stable environment is only possible inside strictly controlled laboratories. Experiments involving test on a construction site or observing office work require a different approach. One serious difficulty is the many uncontrollable variables that create the "unwelcome" extraneous side effects. Safety and scheduling concerns makes it unlikely that a contractor is willing to allow an experimental test of a new device or method on a real construction project. Thus, construction researchers have been challenged to develop experimental procedures within the given constraints while finding ways to collect data that address their hypothesis. Table 2 arranges the various procedures that have been used in the recent past into five clusters each with some unique features.

It is important to recognize that only the first method focuses solely on the experimental performance testing of a new tool or method. Methods 2–5 on the other hand focus on individual, groups of individuals or the output of entire operations. Before exploring in more detail, the main experimental research methods that reader must take notice that various statistical methods are available to quantify the variances of data, to assess the variability of data sets or the goodness-of-fit. Many books and documentations are available on the internet while the Statistics Department at any university generally provides assistance to faculty who are facing challenging problems in designing valid experimental tests [National Institute of Standards and Technology (NIST) 2009].

Method 1. Pilot Test of Devices and Methods

Assuring the quality of materials or the precision of building structures (e.g., the slope of pyramids) has always involved extensive measurements and standardized tests. The rapid growth of the well known ASTM, established over 100 years ago, was based on its scientific approach and the involvement of all the constituents in the standardization process. New or improved construction devices or methods, however, are not covered by such standards. Manufacturers are generally interested to pilot new technology not only to assess its performance but also verify its ease of use, etc. Innovating new technologies in construction also depends on field testing of prototypes to demonstrate technical functionality and opportunities for further improvements. Such pilot tests require a creative researcher to develop solid test procedures. The most common approach is to deploy and collect data related to the performance under laboratory or field conditions.

Laboratory Tests

Tests in a protected environment with access to standardized large testing apparatuses provide many advantages but limit in most cases the applicability to the real world of construction. Peng (2002) published an example where observed conditions on site were recreated in the lab to test the load capacity of shoring systems: "*The experimental test results indicate that the base stiffness to the ground of shores is … the critical load of the shoring system directly measured from the tests varies from 76 to 80% of the tested critical load of an individual shore …*" The test results were compared to theoretical calculations in order to provide recommendation for safe designs. A different type of performance investigation of a new tool in construction involved assessing the survivability of various bar code labels in the construction environment. For this purpose, an assortment of label types was attached to a random selection of hand-tools and gas-

J. Constr. Eng. Manage., 2010, 136(1): 26-35

bottles. After incremental periods, the "health" of the labels was inspected.

Scaling from Laboratory to the Field

The common path for new devices, after they pass laboratory tests, is to be brought into the "real world" to assess performance under field conditions. This nontrivial step requires not only an upgrading of the human-machine interfaces, in order to facilitate operation by field personnel, but also the redesign of power supplies, hardening of the protective structure (e.g., protection against potential rain) as well as mountings for safe transportation and interface with other systems on site. One such example is the equipment mounted utility detection system discussed by Kolera and Bernold (2006). They reported about extensive laboratory work to develop control and validate the performance of an innovative system that integrated mechanics, geophysics, electronics, and computing. In a second development, the hardened system was attached to a backhoe loader for field tests that demonstrated that it: "*… not only detected the presence of the pipe but aimed correctly at the midpoint of the pipe within the space of the trench*" (Kolera and Bernold 2006). A second paper that falls in this category presents the work in testing a real-time asphalt density measuring device. The writers of the paper explain that: "*Several field studies have been conducted … Additionally, laboratory experiments were conducted … Core samples were also taken to relate the variance of the signals to the densities*" (Jaselskis et al. 2001). A drastic upscaling of a laboratory to integrate with a commercial device was necessary to test an intelligent controller on a shop rebar bender. The writers describe how a manufacturer was willing to loan a bender but requiring them to develop a unique interface that allowed the transfer of the control model developed in the lab: "*Bending tests were conducted with both a laboratory prototype and an actual shop table bender … Springback model evaluations revealed that empirical statistical models, neural networks, and in-process relaxation performed equally well*" (Dunston and Bernold 2000). Two more recent example are the development and field testing of a machine vision-assisted, teleoperated pavement crack sealer (Lee et al. 2006) and the instrumentation and field testing of a roller compactor to monitor vibration behavior during earthwork compaction (Rinehart and Mooney 2008).

Many Demands of Field Tests

While still only having to deal with nonhuman test objects, orchestrating field experiments requires many human skills from the director or manager including coordination, communication, and flexibility besides a thorough understanding of the technology to be tested and the construction process itself. How much may be involved in testing a new device in the field was verified by Jaselskis et al. (2005) who explained the procedure and the outcome of testing the performance of a laser scanning device the Cyra 2500 Laser Scanning Unit. The system was deployed in six different field situations to collect raw data while traditional surveying approaches were used to create benchmarks for accuracy. The preparatory work not only included hands-on training of the study team in the use of the hardware and software before it was sent to perform the entire process on its own: "*… there were six test areas involved in this pilot study: (1) An intersection including a railroad bridge, (2) a section of highway including a pair of bridges,…*" (Jaselskis et al. 2005). The objective of the study not only focused on the accuracy of the laser scanner in the six different setting but also on establishing the time and to cost for creating those outputs. Not surprisingly, the learning curve in developing efficiency in operating such complex systems was considered very costly. An excellent example of a project that focuses on a method rather than a device tested the effectiveness of ASTM C1074 (standard practice for estimating concrete strength by the maturity method) to predict the concrete strength in a cold environment recently presented by Bagheri-Zadeh et al. (2007). Eight foundations of an actual project were equipped with 29 data loggers measuring the development of concrete temperature while casting of standardized concrete cylinders provided validations in relation to established testing methods (breaking of concrete cylinders). After a statistical treatment of the different data sets the researchers were able to conclude that: "*… the concrete maturity method could produce a reliable and accurate prediction of in situ concrete strength on a continuous real-time basis during curing*" (Bagheri-Zadeh et al. 2007).

Method 2. Passive Observations

The descriptor, case study, used in Table 2 indicates a one-shot nature of this quasi-experiment where the researcher is does not administer a test but remains a passive observer of what is happening. This method actually dates back to Aristotle who stipulated the need for observations as the basis for knowing. Galileo's study and descriptions of the Lunar surface using one of the first telescopes of the time, but invented in Holland, is a good example. The JCEM contains several examples including a paper by Lee et al. (2007) expounding on an effort to monitor and compare the production rates of freeway rehabilitation projects in California: "*A comparison of production rates was based on the as-built progress data … recorded during one 55 h weekend closure … Examination of the as-built progress data indicated that … A higher production rate and a noticeable 'learning-curve effect' was observed …*" (Lee et al. 2007) The writers also point to the limitation of the small sample sizes to create generalizable knowledge. Nevertheless, as another example demonstrates the collection of useful data may require an extensive effort. Here, the quality of asphalt overlay application during day and night cycles on a 8.85 km (5.5 miles) stretch of urban highway was assessed. Surface smoothness, in-place density, gradation, and asphalt content were measured during two weekends of work by one contractor for comparison with published averages. "*Production rates were compared to those from a comparably sized nighttime project. The investigation revealed that.… High paving production rates resulted from … with respect to these parameters was decidedly better than average. Average shift production rates for the weekend closure were higher than those documented for a similar nighttime paving project*." (Dunston et al. 2000)

While both cases present extensive amount of sampled data, mainly to measure work results, even employing standard instruments (e.g., California Profilograph), it is impossible to reliably draw cause-effect conclusions that are transferable. In cases where the observed company is aware of the study arranged by the owner the surveyed laborers, operators, crew foremen, superintendents, as well off-site personnel most probably changed their normal behavior much like the factory workers at Hawthorne Works in the 1930s.

J. Constr. Eng. Manage., 2010, 136(1): 26-35

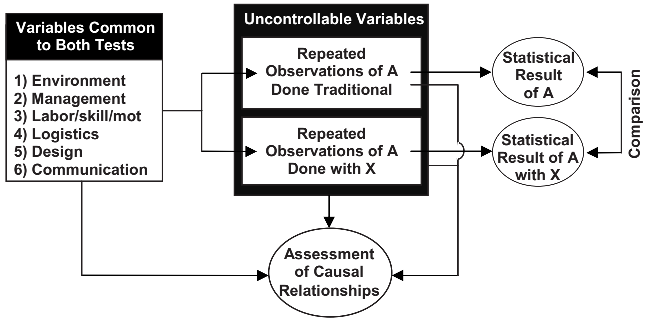

Fig. 3. Design for comparative field experiment with minimal impact of extraneous variables

Method 3. Controlled Experiment

As we learned earlier, this method was first used by the British Royal Navy surgeon James Lind in 1747 (see Fig. 1). The difference to Method 2 is the purposeful setup of a comparative test including the collection of relevant data before, during and after the ongoing experiment. Reviewing the JCEM one quickly realizes that this approach is heavily used albeit with several variations.

Comparative Evaluation

A hypothesis may state: "*Procedure X will increase the productivity of construction process A by 30%*." The most common approach to prove or disprove such a hypothesis is a controlled experiment to compare the productivity of process A with and without with X. While this might sound simple, the difficulty lies in the design of experiments that produce data that create clear and "clean" causal relationships between procedure X and the productivity of A. Since most construction takes place within an uncontrollable environment many extraneous influences may "sneak in through backdoors" to create misleading correlations. A recent paper by Shin and Dunston (2009) shows how researchers reduced the complexity of the construction site by designing experiments that could be executed inside a controlled space. They correctly state this limitation in the conclusions and point to the need for further work on site.

In some situations, this risk may ask for a preexperiment to assert that all influential factor variables are being either controlled or their effect captured. The model in Fig. 3 is based on the assumption that all cause-effect causing factors are known, captured and assessed.

Fig. 3 highlights a strategy that does not eliminate uncontrollable variables but minimizes their conflicting impacts. As shown, the comparative experiments are executed within the same environment, management, labor, etc. The underpinning for this design is the expectation that hidden factors impact both experiments in the same manner, thus cancel each other out. It is apparent, that this cannot necessary be true in all cases. For example learning skills, motivation or changes in health of persons involved in comparative experiments can have drastic impacts on the measurements.

Four examples where this approach has been used involved the comparison of automated with human control of a dredging operation (Tang et al. 2009), the pollutant level and energy efficiency of an intelligent controller for a tunnel ventilation system as compared to a conventional controller (Chu et al. 2008), effectiveness of different approaches to rebar placement for a multistory building (Salim and Bernold 1994) and the laying of concrete pipes with traditional and advanced technologies (Lee et al. 2003). In the latter case, a utility contractor was willing to dedicate one and the same crew to lay nine pipe-sections traditionally on one day and the same pipes using the new technology during the next. Furthermore, these two operations took place in the same field and under the same weather conditions. During both operations, each repetitive task was timed allowing the calculation of mean task as well as cycle times. One cycle consisted of (1) excavation for one section; (2) lowering of one pipe; (3) jointing of new piece; and (4) initial backfill. Because of the nine repetitions, the learning effect could be easily detected. Following the design in Fig. 3 the operation was continuously observed in order to measure the effect of uncontrollable variables on the productivity. For example, during the experiment, an electric winch broke and had to be replaced with a chain and come-along system drastically lengthen the time of rigging. Safety and quality were constantly monitored as well. Due to the inherent simplicity of the system, the crew showed an extremely steep learning curve. Still, the Hawthorne effect could not be measured but assumed equally strong during both tests.

Multifactor Comparative Evaluation

Human factor studies have shown that the performance of humans may be effected by a range of causes some counteracting and some synergistically supporting each other. Such "hidden" factor combinations may result in such large variances when correlating cause-and-effect variables, that no sound conclusions can be drawn. Two experiments where this phenomenon surfaced involved the productivity assessment of a sensor-equipped nailgun (Miller and Bernold 1991) and the speed and safety of different manipulator control concepts for bridge paint-removal (Moon and Bernold 1998). Fig. 4 depicts three graphs with hypothetical data to demonstrate in a simplified manner the underlying principles and the merits of considering a multi-factor approach when analyzing experimental data.

Fig. 4(a) plots the average productivity of seven volunteers testing a new device. A point represents the mean value for 10 observations, once working with the traditional and then with the new device. As shown, the surfacing improvements show large changes and range from 1.5 to 16 units per Person-hr. No linear relationships are apparent rather some small clusters. Is there a hidden factor that has extremely strong relationships with productivity? If data for all potential impact variables has been collected an ANOVA will find the strongest relationships and even the effect of covariances. The critical nature of having "all" the necessary data after all the experiments are completed reemphasizes the importance of preexperiments.

Preexperiment

Fig. 4(b) depicts graphically the result of one preexperiment that had the objective to test if and how the years of experience effect productivity. Twenty-two laborers with different years of experience were separately observed nailing the plywood for the roofs of single-family houses with the traditional nailgun. The only experimental variable was the years of experience working with nailguns. The hypothetical data is shown in Fig. 4(b) shows that there exist three definite clusters or strata demarcated by 3.5 and 6.5 years of experience. The importance of this data are not only

J. Constr. Eng. Manage., 2010, 136(1): 26-35

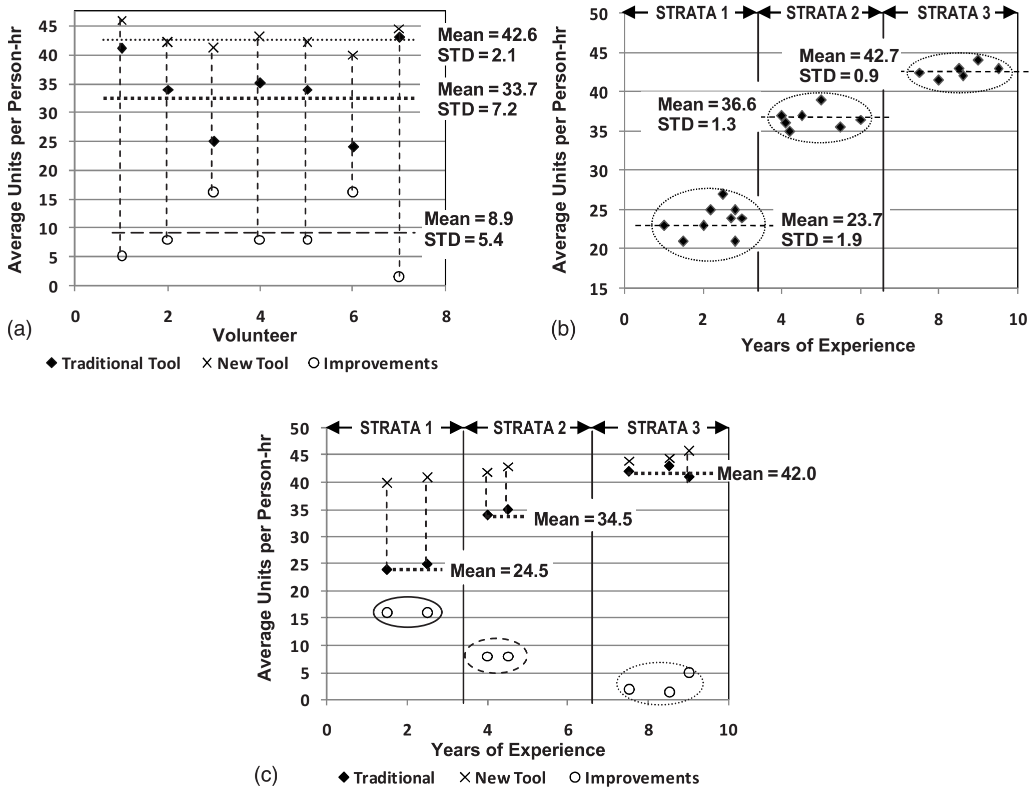

Fig. 4. Experimental results with and without stratification (a) results of initial experiments; (b) identification of strata within labor-pool; (c) initial experimental results stratified

in proving a correlation between experience and productivity but more importantly, this information will be instrumental in designing the actual experiment.

Experiment with Stratified Samples

In order to create generalizable results, conducting experiments with a representative sample of laborers is important. The preexperiment showed the existence of three experience groups showing different productivities. It is apparent that the samples of laborers to be invited to the comparative experiment should include representatives from each strata. In fact, the number of laborers in each strata should reflect the proportion of the strata in the overall population. Fig. 4(c) plots the data from Fig. 4(a) as a function of experience with the oldest strata being represented by three subjects. It can be observed that the average performances of the small samples are all close to the strata mean values in Fig. 4(b) although no subject participated in both experiments. It is now immediately apparent that those with least experience benefit most from the new device. This important conclusion gives immediately raise to a new question: How quickly will the inexperienced learn to handle the new device most effectively?

Learning Effect

As mentioned earlier, in experiments that involve human expertise to coordinate manual operations one has to be aware of the learning effect. Learning effects the cycle time of an operation not dictated by the uniform pace of a machine because repeated execution of that same motion leads to better hand-eye coordination, less experimentation, and time consuming mistakes. Based on empirical data, the general form of the learning follows a polynomial function with scale invariance of the form $T(x)=T_1 * x^{-\alpha}$. T_1 represents the duration of the first cycle and α the learning curve exponent. Task duration $T(x)$ reduces for each repetitive cycle x along the curve defined by $x^{-\alpha}$ until it reaches a level that is limited by extraneous factors. This curve is further bound by the maximum number of cycles it will take to reach the minimum time. Experience shows that learning curves reach from 90% for repetitive welding or 75% for more complex tasks, meaning that a cycle time will drop by 75%.

Method 4. Randomized Experiment

Proving the hypothesis that: "*Total Stations improve the productivity of U.S. motor grader operators by 28%*" will require experiments with a large number of operators, using many different grader models, working under different conditions and regions of the United States. As it is impossible to test every single operator a sufficiently representative sample has to be selected. How big that representative sample has to be depends on the required precision and the variability in the entire population. Various formals have been proposed that allow us to calculate. For the sake of this demonstration, expecting a confidence level of 95%, we shall use the following simplified formula: Sample size $n=N/[1+N(e)^2]$ where N=total population and e=acceptable error. If we assume that there are 3,500 active grader operators in the United States and we consider the acceptable sampling error as 5% the calculation of sample size looks as follows: $n=3,500/[1+3,500(0.05)^2]=359$ operators. Reducing the confidence level to 10% would still require 98 operators, a large number considering the fact that each would have to be tested in various environments. Stratification of the total population might help but would

J. Constr. Eng. Manage., 2010, 136(1): 26-35

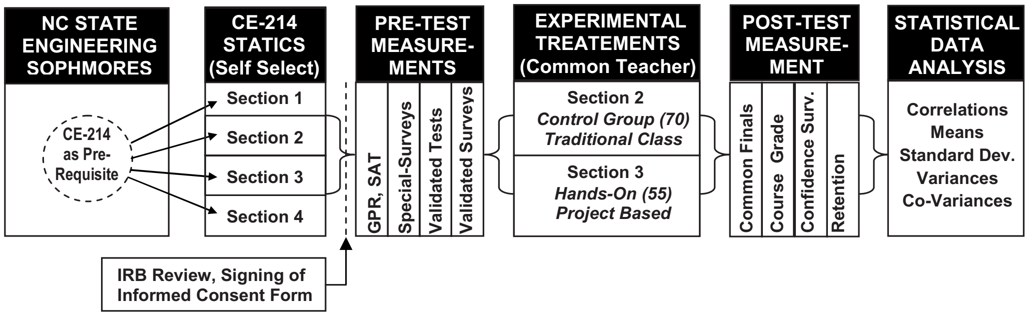

Fig. 5. Design of an educational experiment with control group and self-selection

require the availability of statistical data about the different strata (see previous section). This demonstration of sample size calculation should be sufficient in demonstrating that such projects would require funding much higher than is common in construction engineering and management. On the other hand, selecting operators working for one company only might skew the results that they may show similar and unique traits having been hired based on specific company criteria. The same holds true for limiting the sampling to one town, state or region of the United States. The only solution to this dilemma is to limit the hypothesis to cover a population that can be sampled in a sufficient manner. It goes without saying that the conclusion will only apply to the targeted population.

Self-Select within a Limited Population

A more feasible alternative is to sample a smaller population or even a strata such as students attending a given university or a program. For example North Carolina State University (NCSU) has approximately 1,000 engineering freshmen every year. Applying the same formula as above would result in a still large sample size $n=1,000/[1+1,000(0.05)^2]=285$. Instead of sampling from the entire population, considered homogenous, we might want to stratify by matriculation and pick randomly from each cluster so that each cluster is equally well represented in the final sample. While such a procedure might be desirable, the reality of a university prohibits that students are being "coerced" to participate in experiments that might influence their grades and with it their future. One way to circumvent this problem is through self-selection and student consent. Bernold et al. (2000) presented the result of such a study in engineering education where students self-selected into two sections of statics taught by the same professor at NCSU.

Fig. 5 reveals that only a subset of all engineering sophomore needed to take CE-214 (Engineering Mechanics-Statics) as a prerequisite for many other classes. While the students' self-selection from several parallel sections cannot claim to represent a perfect randomization process one may be able to verify that it is sufficient for a particular experiment. It goes without saying that engineering students who become sophomores at NCSU do also not represent a randomized sample of all engineering students in the United States. Still, by comparing the representation of critical factors within the population (e.g., female versus male) with the engineering at other universities may give sufficient indications about the soundness of the sample. Fig. 5 involves a very important topic that effects every research that involves people, namely protecting the safety of human subjects.

Informed Consent

According to the Code of Federal Regulations Title 45 Part 46, Protection of Human Subjects, every institution that is involved in funded research is required to operate an office that reviews, approves, and documents that every project is following approved procedure regulations. At NCSU, this office was called Institutional Review Board or IRB for short. In cases where research is "conducted in established or commonly accepted educational settings, involving normal educational practices … (Code of Federal Regulations 2009)" Researchers will be exempt from continued oversight by the IRB but will have to design an informed consent form that explains the goals, risks, and the context of the involvement to be signed by both the student and the research director (see Appendix I) The main hypothesis of the educational project underlying Fig. 5 claimed that: "Hands-on project based teaching of statics will improve the grades of students who depend on concrete experiences to understand and learn." Thus, the objective of the experiment was to establish strong correlations between the learning preferences of students, 55 in the experimental and 70 in the control group, the method of teaching and the course grade. Using a validated survey, the students were stratified into four clusters of students who had similar learning preferences. It was interesting to realize that the self-selection process produced distributions in each section that were close to the overall population at NCSU (10, 22, 55, and 13%), taught by the same professor the students faced similar tests but different methods of teaching and learning. The stratified comparison of the grades from both student cohorts confirmed the hypothesis only partially. While one of the two subgroups needing concrete experiences did in average gain significantly higher course grades, did the second subgroup still equally did poorly in both sections? For the results of analyzing additional data, the reader is referred to subsequent papers in JCEM (Bernold 2005, 2007).

Method 5. Four Group Experiments—Minimized Hawthorne

This is considered the most powerful method to study the effect of any intervention on human performance because it considers the previously discussed Hawthorne effect. As shown in Table 2, four subjects or groups are randomly picked but only two of them are subjected to pretest observations. The other two groups are left without knowledge that they are being observed one using the traditional and the modified method. This design provides the advantage that the consequence of knowing that one is part of the experiment, due to the pretest, can be detected. However, this design requires large samples while the federal requirement to get

J. Constr. Eng. Manage., 2010, 136(1): 26-35

the consent from anybody participating in the research project will add an additional burden for the research team. Hence, it is not surprising that the JCEM does not contain examples of four group experimental designs.

Summary

Finding economic means to improving the quality, safety, and productivity of construction or reducing waste and pollution requires the development of a posteriori knowledge about the effect of various factors. However, coming up with measurable evidence needed to support a hypothesis or theory is extremely hard to come by in the construction industry. Furthermore, the large size of the industry puts extreme requirements on establishing research findings that are applicable to the entire or even a section of the industry. The inherent problem is the cost involved in observing sample sizes big enough to produce data points that are statistically significant.

As the paper shows, published research results demonstrate that it is still possible to perform useful experimental projects in construction. The key to a successful project is the establishment of a hypothesis that fits the available data and the data acquisition capabilities. Likewise, a careful hypothesis will lead to a robust design of the experiment and consequential conclusions.

The majority of experimental projects discussed in the JCEM are pilot tests of devices or methods. It was discussed how controlled experiments that compare the performance of one device or method against another require extensive planning and data collection procedures to ensure that extraneous effects are held at a minimal level or are at least being documented and considered in the final analysis. The existence and the size of multiple impact factors is best established by performing preexperiments. Such preliminary results not only provide evidence if multiple factors exist they help to improve the design of the final experiment using stratified sampling. Examples were used to demonstrate how such an approach could ensure the representative presence of every cluster existing within a larger population.

Experimental research in construction in most cases involves human subjects. Many unexpected things can and do happen despite the best preparations. A researcher working with humans within a laboratory or in the field should, in his own best interest and in the interest of each participant, acquire the guidance and the approval of the IRB. This step will not only prevent accidents but protect the researchers from the unconscious misuse of personal information.

Appendix I. Informed Consent Form

Research Study: Study Skills and Success of Engineering Students

Principal Investigator: Dr. Leonhard E. Bernold

You are invited to participate in a research project to better understand the interrelationships between learning strategies and academic success of students in the College of Engineering. This information will be used to develop new educational material that will designed to enable students to reach their academic and professional goals….

BACKGROUND INFORMATION

1. As part of your classwork for E101: Intro to Engineering and Problem Solving, you will be asked to complete three standard survey forms and keep an electronic journal answering short questions about your progress in class. The electronic journaling will take approximately 15 min per week. Some of you will be invited to group interviews at the end of each semester.

2. ..

BENEFITS

Considered a first in the nation, this study will provide important information about students' learning preferences, how engineering students study, and how it impacts their success and confidence.

CONFIDENTIALITY

The information in the study records will be kept strictly confidential. For example, your journal entries will be coded electronically. When data are given to us by the course instructors, it will be stripped of identifying information. Data will be stored securely and made available only to persons conducting the study unless you specifically give permission in writing to do otherwise. No reference will be made in oral or written reports that will identify you as an individual.

PARTICIPATION

Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed. If you decline to participate or withdraw from this study you are still expected to submit the assigned journals to your instructor for you to earn your final grade, but your data will not be forwarded to us for analysis.

CONTACT

If you have questions at any time about the study or the procedures, you may contact Dr. Leonhard E. Bernold, at .. If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact Dr. … Chair of the NCSU IRB for the Use of Human Subjects in Research Committee, Box.., NCSU Campus … or Mr. …, Ass. Vice Chancellor, Research Admin., Box .., NCSU Campus…

CONSENT

I have read and understand the above information. I have received a copy of this form.

I AGREE TO PARTICIPATE	I CHOOSE NOT TO PARTICIPATE
Subject's Signature:	Subject's Signature:

References

Bagheri-Zadeh, S. H., Kim, H. K., Hounsell, S., Wood, C. R., Soleymani, H., and King, M. (2007). "Field study of concrete maturity methodology in cold weather." *J. Constr. Eng. Manage.*, 133(11), 827–835.

Bernold, L. E. (2005). "A paradigm shift in education is vital for the future of our profession." *J. Constr. Eng. Manage.*, 131(5), 535–542.

Bernold, L. E. (2007). "Teaching evaluations for construction engineering and management: Opportunity to move us forward." *J. Constr. Eng. Manage.*, 133(2), 146–156.

Bernold, L. E., Bingham, W.L., McDonald, P.H. and Attia, T.M. (2000). "Influence of learning type oriented teaching on academic success of engineering students." *J. Eng. Educ.*, ASEE, April, 191–199.

Chu, B., et al. (2008). "GA-based fuzzy controller design for tunnel ventilation systems." *Autom. Constr.*, 17(2), 130–136.

Clarke, D. (2006). *Descartes: A biography*, Cambridge University, Cambridge, U.K.

Code of Federal Regulations. (2009). "Title 45-Public Welfare." *Dept.*

J. Constr. Eng. Manage., 2010, 136(1): 26-35

of Health and Human Services, Part 46, Protection of Human Subjects, ⟨http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm#46.101⟩(Jan 12, 2009).

Dunston, P. S., and Bernold, L. E. (2000). "Adaptive control for safe and quality rebar fabrication." *J. Constr. Eng. Manage.*, 126(2), 122–129.

Dunston, P. S., Savage, B. M., and Mannering, F. L. (2000). "Weekend closure for construction of asphalt overlay on urban highway." *J. Constr. Eng. Manage.*, 126(4), 313–319.

Jaselskis, E. J., Gao, A., and Walters, R. C. (2005). "Improving transportation projects using laser scanning." *J. Constr. Eng. Manage.*, 131(3), 377–384.

Jaselskis, E. J., Han, H. C., Grigas, J., Tan, L., and Fahrion, D. (2001). "Status of roller mountable microwave asphalt pavement." *J. Constr. Eng. Manage.*, 127(1), 1–46.

Kolera, B. T., and Bernold, L. E. (2006). "Intelligent utility locating tool for excavators." *J. Constr. Eng. Manage.*, 132(9), 919–927.

Lee, E. B., Lee, H., and Ibbs, C. W. (2007). "Productivity aspects of urban freeway rehabilitation with accelerated construction." *J. Constr. Eng. Manage.*, 133(10), 798–806.

Lee, J., Lorenc, S. J., and Bernold, L. E. (2003). "A comparative performance evaluation of tele-operated pipe laying." *J. Constr. Eng. Manage.*, 129(1), 32–40.

Lee, J. H., Yoo, H. S., Kim, Y. S., Lee, J. B., and Cho, M. Y. (2006). "The development of a machine vision-assisted, teleoperated pavement crack sealer." *Autom. Constr.*, 15(5), 616–626.

Miller, M. L., and Bernold, L. E. (1991). "Sensor-integrated nailing for building construction." *J. Constr. Eng. Manage.*, 117(2), 213–225.

Moon, S., and Bernold, L. E. (1998). "Graphic-based human-machine interface for construction manipulator control." *J. Constr. Eng. Manage.*, 124(4), 305–311.

National Institute of Standards and Technology (NIST). (2009). "e-Handbook of statistical methods." ⟨http://www.itl.nist.gov/div898/handbook/index.htm⟩ (Jan 12, 2009).

Peng, J. L. (2002). "Stability analyses and design recommendations for practical shoring systems during construction." *J. Constr. Eng. Manage.*, 128(6), 536–544.

Rinehart, R.V., and Mooney, M. A. (2008). "Instrumentation of a roller compactor to monitor vibration behavior during earthwork compaction." *Autom. Constr.*, 17(2), 144–150.

Salim, Md., and Bernold, L. E. (1994). "Effects of design-integrated process planning on productivity in rebar placement." *J. Constr. Eng. Manage.*, 120(4), 720–737.

Shin, D. H., and Dunston, P. S. (2009). "Evaluation of augmented reality in steel column inspection." *Autom. Constr.*, 18(1), 118–129.

Simon, H. B. (2002). *The Harvard medical school guide to men's health*, Free Press, New York, 31.

Tang, J., Wang, Q., and Zhong, T. (2009). "Automatic monitoring and control of cutter suction dredger." *Autom. Constr.*, 18(1), 194–203.

J. Constr. Eng. Manage., 2010, 136(1): 26-35