Implementing Observational Research Methods to Study Team Performance in Construction Management

Robert M. Leicht[1]; Samuel T. Hunter[2]; Chitwan Saluja[3]; and John I. Messner, M.ASCE[4]

Abstract: The evolution of research in the area of construction engineering and management has moved toward the use of quantitative metrics and research methods for identifying and evaluating the impact of events or procedures on the construction process. While quantitative research methods can be very important for answering certain research questions, there are important reasons for adopting qualitative or mixed methodology studies to gain a better understanding of complex phenomena. This paper explores the use of observational studies and their potential use within the field of construction research. Observational studies can provide answers to "what" phenomena occurred, particularly when people are involved in a process, along with gaining insights into "why" the phenomena occurred. This paper discusses two types of observational studies, structured and unstructured, and provides a procedure for their implementation within construction research. To clearly demonstrate the methodology, a case study focused on the investigation of two different versions of an educational simulation application, the virtual construction simulator (VCS), is used to illustrate the benefits and challenges of implementing mixed methodology observational studies. The case study involved the video recording and analysis of interactions between student team members when using the VCS application for a construction sequencing task. The video recordings were analyzed, and important insights were identified, both qualitative and quantitative. Through content analysis, it was determined that the improvements made in a new version of the VCS application were beneficial, and the detailed observational studies identified insights into why the revisions in the application yielded improved results. This case study details the steps and considerations involved in planning an observational study, as well as the benefits and challenges that researchers may encounter when using observational research methodologies.

DOI: 10.1061/(ASCE)CO.1943-7862.0000080

CE Database subject headings: Research; Methodology; Construction management; Productivity; Teamwork.

Author keywords: Research methods; Observational studies; Content analysis.

Introduction

The use of observation as a means of scientific research is in many ways one of the oldest methods recognized. Darwin used sketching and personal observation of Galapagos tortoises as the backbone for his Origin of Species (Darwin 1859). Newton developed the concepts for the laws of universal gravitation from his own "observation" when hit by a falling apple (Conduitt 2006). Archimedes used simple observation for determining concepts of material density through his "Eureka" moment in a bath tub (Biello 2006). The use of observation is indelibly linked with most major advances in science.

Since these historical uses, the methodology and application of observation as a form of research has grown and developed into a regular practice in many areas of research ranging from documenting physical symptoms of diseases to behavioral studies, to studies of communication. Developments within computing and audio-visual technology have led to the utilization of better cameras, microphones, and software for capturing activities for observation as well as more thorough and accurate analysis capabilities. The value of the use of observational studies, using these new tools, lies in the inherent mixing of the methods. The use of observation is inherently qualitative, but by integrating content analysis of the video it provides the ability to generate quantitative data to support the inferences from the qualitative observation. While the use of observational methods is not unknown in construction and dates back to work on themes such as worker efficiency (Taylor 1903), the use of these methods is fairly rare in construction research. Also, the use within construction is focused on simple tasks like that of Taylor, which focuses on fairly simple repetitive tasks rather than more complex and dynamic issues such as means of effective collaboration, interdisciplinary teaming, or the use of new technology in construction. With recent trends in the delivery of projects focusing more on integrated delivery of facilities, means of closer collaboration, and the use of technology for virtual teaming, this methodology could potentially move to the forefront in research related to teams and technology as these tools become more widely used in the industry.

This paper demonstrates an overview of the different manners

[1]Graduate Research Assistant, Dept. of Architectural Engineering, Pennsylvania State Univ., University Park, PA 16801 (corresponding author). E-mail: rml167@psu.edu

[2]Assistant Professor, Dept. of Industrial Psychology, Pennsylvania State Univ., University Park, PA 16801.

[3]Graduate Research Assistant, Dept. of Architectural Engineering, Pennsylvania State Univ., University Park, PA 16801.

[4]Associate Professor, Dept. of Architectural Engineering and Engineering Design, The Pennsylvania State Univ., University Park, PA 16801.

Note. This manuscript was submitted on August 8, 2008; approved on April 2, 2009; published online on April 30, 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 0733-9364/2010/1-76–86/$25.00.

J. Constr. Eng. Manage., 2010, 136(1): 76-86

in which observational studies may be used, citing examples to clarify the unique implementation and variety of uses. Following this background and explanation, a case study of implementation is presented to clearly demonstrate the steps and the potential for use in construction research contexts.

Use of Observational Methods in Social Science Research

It is difficult to determine precisely when observational research methods were first used (Wax 1971; Bauer et al. 2000). These techniques are, in many ways, natural steps in the research process. Simply stated, if we are curious about why something occurs we are likely to seek observations of that phenomenon to more fully understand it. Thus it would appear more fruitful to consider when the observational method gained formal recognition as a research tool rather than when it was initially employed. In doing so, we find some early use in studying the London working class during the 1880s. The technique became more formalized, however, in the United States—particularly in Chicago during 1920s and 1930s (Gill and Johnson 2002; Wax 1971). It was during this period that a group of sociologists formed the Chicago School and trained some of the most influential and important qualitative researchers. Termed ethnographers, these researchers sought to understand the Chicago working class by examining them in their own environment—that is, observing them in their natural context. Ethnography has gone through a number of changes since its formal inception. In fact, Denzin and Lincoln (1994) suggest that there are five historical stages the approach has gone through: (1) traditional (1900–1950); (2) modernist (1950–1970); (3) blurred genres (1970–1986); (4) crisis of representation; and (5) the post modern challenge/the present. Notably, Denzin and Lincoln suggest that recent stages were brought about by technology such as video cameras and tape recorders. It should be noted that the distinctions and current status of these stages has been challenged by some (e.g., Atkinson et al. 1999). What is clear, however, is that observational methods have had both high and low points with regard to their approach and favor with researchers. At present, it is safe to say the method is being welcomed by an emerging group of researchers eager to understand their phenomena in a more complete fashion.

Qualitative methods such as observation have been adopted by a number of groups within the social sciences including, most notably, the sociologists out of the Chicago School mentioned above. Other groups include anthropologists, who have often embraced the "naturalistic" component of ethnography with some choosing to wholly embed themselves in the culture they are studying in an attempt to gain a comprehensive understanding of that culture as well as limit their own influence, or bias, in their observations. Ethnography is used to some degree in psychology, most notably in the study of children as well as the workplace (Mintzberg 1973; Gill and Johnson 2002). Historically, however, many areas of psychology have been skeptical of the method viewing it as unscientific and lacking empirical strength. Instead, experimental and survey-based techniques have been emphasized in an attempt to obtain perceptions of scientific rigor. This negative perception has begun to wane, however, as researchers begin to dissolve the distinction between "softer" qualitative research and "harder," quantitative research. In fact, most researchers today are open to and understand the benefits, as well as costs of, combining the two methodological approaches when studying a phenomenon.

Quantitatively Coding Qualitative Data

Termed mixed-methods research, current ethnographic studies are often characterized by collecting qualitative data and quantifying it in such a way as to allow for the application of traditional statistical techniques (Plano Clark et al. 2008). Better characterized as a quantitative analysis of qualitative data, mixed-method approaches have a number of unique advantages over the use of qualitative and quantitative approaches alone. For example, although a quantitative researcher may observe a correlation between variables X and Y, supplementary qualitative information may provide the researcher with the context to better understand the processes involved in that relationship. On the other hand, applying traditional quantitative methods to qualitative data allows researchers to more strongly establish generalizability to other, similar contexts. It is not difficult to see how gaps between theory and practice are bridged by providing both numbers and context to more completely understand a given phenomenon.

The use of observational methods within these contexts is divided into two main categorizations, usually identified as structured and unstructured observation (Yin 2003). Structured observations rely on frameworks of predefined actions, discussion content, or even body language to fit the activity within the variables and scope of the research question. By using an existing framework or developing a specific framework to track observed activities, the researcher has chosen the architecture around which they are fitting the hypothesis and they utilize the data from the observations to create the correlations between the pieces. Buchholz and colleagues observed 17 ironworkers for a period of 13 days during construction of a tunnel construction project in Boston (Buchholz et al. 2003). They were focusing on ergonomic concerns of specific tasks performed while installing the reinforcing in preparation for the concrete placement. The coding was performed live with a very detailed coding framework focused on posture, activities, tools, and handling of materials (PATH). The detailed structured framework created for the study used exact calculations about the loads, material weights, and tool use techniques to incorporate how the workers physical forms were distributing the stresses of the construction tasks to their backs, legs, and arms. The detail and structure of the framework enabled the researchers to relate the data collected about times and frequencies of activities to the safety measures and longer term physical impacts on the workers of performing specific tasks. In this case the researchers had very specific concerns regarding body position and movement. By coding the field durations and frequencies of the various construction activities, the researchers could draw conclusions about these variables in the changing setting typical to construction, which previously had been considered too dynamic a setting for this type of study as compared to the typical factory floor studies.

In contrast to structured observations, an unstructured observation utilizes no preset framework. The concept driving unstructured observation is to have the researcher enter into the observations with no preconceptions regarding the expected outcomes. The researcher simply records and notes the behaviors in question and the trends in the behaviors will emerge with time. Along with these, nonfrequency techniques and contingency analysis, coding whether or not something is used, rather than how often, or the absence or presence of an attribute may be more significant than the relative frequency of other characteristics. In a study of lab processes in science classes, Dov and colleagues used unstructured observation of classrooms to identify the different processes used in conducting lab activities in science classes (Dov

J. Constr. Eng. Manage., 2010, 136(1): 76-86

2005). The identification of seemingly unique activities offered insight into areas that set apart those teachers and their methods from the typical process.

Conducting Impactful Observational Research

One of the primary reasons qualitative methods have not been embraced by all researchers is the types of bias and subjectivity associated with such approaches. That is, many have argued that a researcher's personal bias plays too great of a role in gathering, synthesizing, and analyzing qualitative data (Hammersly and Gromm 1997). It should also be noted that, in contrast, some researchers have countered this argument, suggesting that this bias is evident in all types of research and that qualitative studies are unfairly criticized. We concede that this may in fact be true, but contend that studies are significantly improved to the extent that bias is reduced and some degree of rigor is employed. Fortunately, the reduction of bias is conversely related to two critical components of research: reliability and validity. We continue then, with a discussion on how to maximize both and, in turn, minimize investigator bias.

Reliability. With regard to the social sciences, reliability refers to the consistency of a given phenomenon or observation. For example, measures of intelligence are created in such a way that if we give an individual an IQ test at Month 1 and then again at Month 6 we would expect roughly the same score each time. Similarly, if a scale is intended to measure job satisfaction and each item on the scale is scored the same way by each employee (with reasonable variation among participants) that scale is said to be reliable—each item produces a score similar to the other items. Central to our discussion, reliability is often viewed as inversely related to random error. That is, scales and measures that are reliable are said to have less error than those which are unreliable. Thus, in understanding how to analyze data we seek to minimize error variance or use approaches that result in reliable data. It is important to keep in mind, however, that reliability is not equivalent to validity but rather is a necessary precursor to it.

Reliability and its assessment come in many forms. One of the simplest is test-retest reliability where scales are given at multiple times and assessed for consistency between time points. Others include internal consistency and split-half reliability where scale items are compared to one another with the goal of determining consistency among items purported to measure the same construct. The type of reliability most applicable to observational research is inter-rater reliability. That is, assessing the extent to which judges make the same assessment about a given phenomenon or event. For example, if we ask three judges to rate the quality of a given product we would hope to find consistency among their rating and if we did, this would suggest adequate inter-rater reliability. Further, by finding adequate reliability, we would be more confident that our ratings minimize the potential for error. On the other hand, if the judges' ratings were largely divergent, it would be unclear as to what we were, exactly, measuring. In short, we would have less confidence in our data and, ultimately, our results.

How then, do we maximize reliability and therefore reduce error in observational research? The simplest and most effective answer is training. Raters must be exposed to the phenomenon they were chosen to assess and gain experience in making judgments on that phenomenon. The end result is the formation of shared mental models—or similar perspectives on rating the behaviors or events at hand (Mathieu et al. 2000). This may be accomplished several ways; the first is providing them with clear examples (good and bad) of what they will be rating. The second is a series of rating exercises where raters make ratings independently, share their results and rationale for why ratings were made followed by another round of ratings, and discussion until ratings are consistent among judges. This requires some guidance by researchers to ensure that one judge is not dominating the discussion but also to ensure that judges are sharing their rationale for their ratings. The latter is essential for creating shared mental models among judges—each judge must understand what the other is thinking during the rating process. Nearing the end of training and prior to conducting data analysis, it is still essential to assess reliability among judges via the application of statistical techniques. Approaches vary, but typical data analyses include r_{wg} (James et al. 1984) $ICC1$ (Bliese 2000), and more basic forms such as percentage of agreement. The reader is invited to view Bliese (2000) for a thoughtful discussion on inter-rater assessment as well as Bernardin and Buckley (1981) and Woehr (1994) for a more specific discussion on types of rater training.

Validity. If reliability is viewed as consistency of a given phenomenon, validity occurs when a phenomenon operates, in actuality, as we propose it does. It is important to note that validity is not an all or nothing distinction—rather, it must be viewed as a matter of degree (Messick 1995). Validity has been conceptualized in a variety of ways, although for the present discussion it is useful to break it into two basic forms: test validity and what might broadly be termed design validity (Cook and Campbell 1979). Test validity refers to scales used in measuring psychological phenomena. For example, a measure of intelligence is said to be, at least partially, valid if it is related to other established tests of intelligence. More central to the present discussion, observational research is valid if it is free from design confounds or biases introduced by the researchers—that is, it may be viewed as having design validity. Referred to as internal validity, Campbell and Stanley (1963) note that there are at least eight threats to internal validity—a list expanded by four in later works by Cook and Campbell (1979). The reader is invited to view these manuscripts to gain a more complete understanding of bias in research.

Of these threats, the most impactful is confounding or systematic covariation among events or manipulations that precludes causal interpretation on the part of the researcher. Consider, for example, a researcher that is observing a group of students designing a new cell phone. Let us imagine that the researcher observes the group every Friday. Let us also imagine that the design team uses a new technology to help them in their design efforts on Fridays. In this scenario, it is impossible to determine if a team's success is due to the use of new technologies or due to being observed, as both occur simultaneously on Fridays. Thus it is essential that observations are conducted in such a way as to not systematically covary with events they are observing.

In addition to internal validity it is also important to briefly discuss external validity (Cook and Campbell 1979) as it applies directly to the use of observational methods. More precisely, observational research has been criticized for lacking generalizabilty (i.e., external validity) or the capacity to extend findings beyond those events and participants observed in the study. Although this criticism may be true for isolated case studies, by applying a mixed-method approach to observational research a greater degree of external validity may be obtained. For generalizability to be maximized, it also is important to choose a sample of participants that are representative of the population to which we hope to generalize. In doing so, we can be more confident that our findings are applicable to the broader population and not just our select sample.

J. Constr. Eng. Manage., 2010, 136(1): 76-86

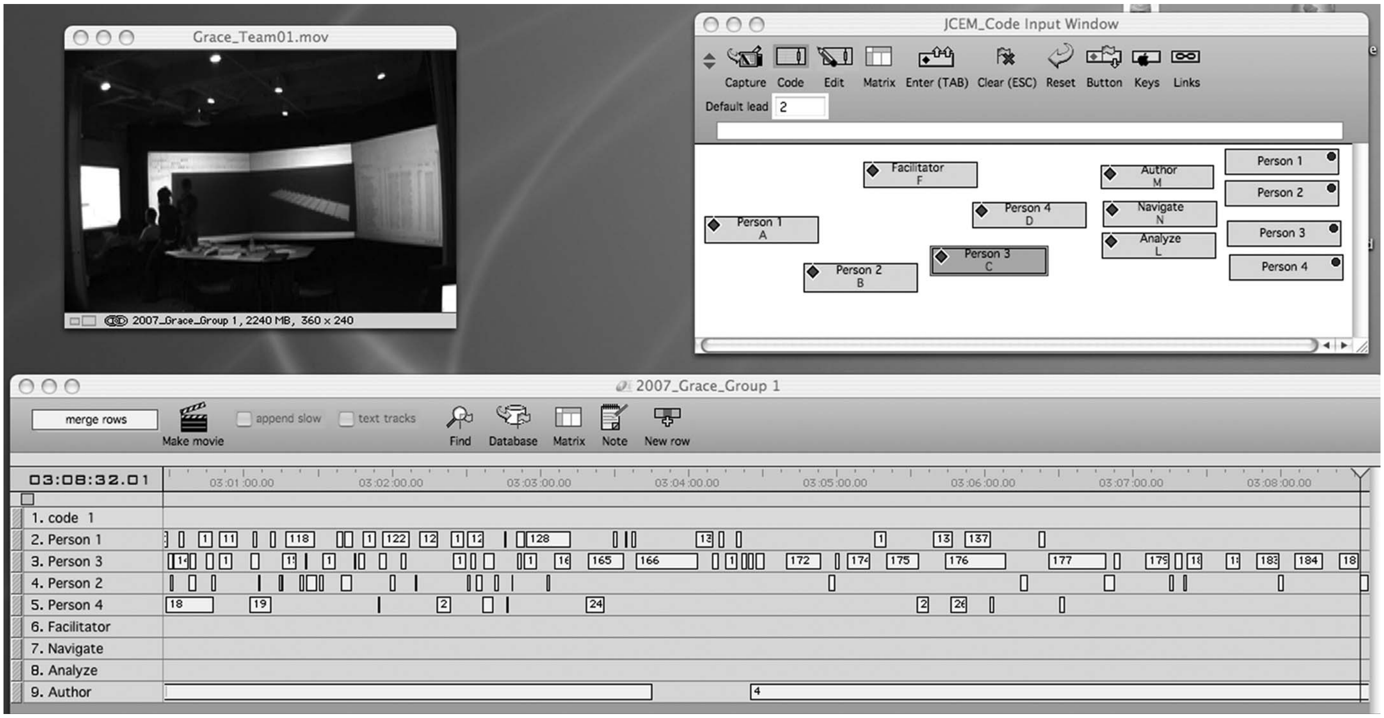

Fig. 1. Image of the Studiocode software interface, showing the video in the upper left, the framework coding window in the upper right, and the content analysis timeline along the bottom

Analyses and Results

The type of analyses employed in observational research varies as widely as the research questions driving them. For example, researchers interested in comparing the mean performance of teams using technology versus those who do not will likely employ an ANOVA based analysis model (Tabachnick and Fidell 2001). On the other hand, researchers interested in assessing the relationship between frequency of technology use and performance are likely to use correlation or regression based models (Aiken et al. 2003). Going further, as research questions become more complex researchers may employ the use of structural equation models or a analytic tool gaining popularity due to its ability to appropriately handle nested data structures, hierarchical linear modeling (Raudenbush and Bryk 2002).

The driving force behind analytic decisions must be the research question at hand. Factored into the decision, however, are more pragmatic issues such as sample size and nature of the study hypotheses. For example, if the research is exploratory, the researchers should note their results and conclusions as such (Tukey 1977). On the other hand, if the researchers are testing explicit hypotheses, more rigorous standards of analyses are often employed. Finally, it is important to blend both quantitative and qualitative elements when presenting results from mixed-method observational research. For example, in their study of leadership Mumford and colleagues (Mumford 2006) often conducted discriminant function analyses on their coded results—a wholly quantitative endeavor. In explaining these results, however, the researchers also provided quoted material from the leader biographies to provide concrete examples as well as general context for their findings. Although this approach is time-consuming, the benefits of a more complete research picture are well worth the mixed model effort.

Computer Software for the Analysis of Audio-Video Recordings

The use of audio-visual recording devices allow for the capture of an activity under study. Once the activity is recorded, greater rigor or exactness in the content analysis can be achieved. The use of video creates the ability for the coder to return to periods of activity. Absent the use of audio-video recordings for coding, observers typically identify activities at set intervals during direct observation. A recent example of this type of observational coding was performed by Gopinath who used 30 s intervals to identify types of conversation used in regular construction job progress meetings to demonstrate inefficiencies in communication (Gopinath and Messner 2004). This remains necessary at times due to limitations in recording activities. The use of audio-video recordings offers the ability to code entire durations of a discussion topic or activity. Also, by using recordings there is the opportunity to clarify or refine the coding framework by direct comparison of the coded outcomes between multiple observers. In the setting of a live coding, the results would only contain the quantitative data. Using audio-video recordings for coding allows for the exact time of each occurrence to be recorded and when discrepancies occur between coders, the occurrence can be reviewed, allowing for the removal of errors, clearer communication by using challenging examples to train coders, and also the ability to revise the coding framework in cases of ambiguity.

The introduction of specialized computer software has allowed for greater efficiency in coding and analyzing audio-video data. Several of these software applications allow users to create custom frameworks for coding categories, as well as labeling functions to correlate data directly, along with shortcuts to make the coding process more user-friendly (Gyorke 2006) (Fig. 1). Along

J. Constr. Eng. Manage., 2010, 136(1): 76-86

with the efficiencies in coding the content, the applications typically have added functionality for searching the results, creating matrices to organize the results, exporting selected video for identified categories, and in some cases even incorporate statistical packages for analysis of the resulting quantitative data. The major challenge of performing content analysis is removing bias from the results inherent in the process from the coders seeing what is expected. The use of recorded sessions allows for verification of the coding (Scott and DeSanctis 1992) and the newer software applications improve the efficiencies and exactness of this process, resulting in minimization of the bias through a more thorough and rigorous process.

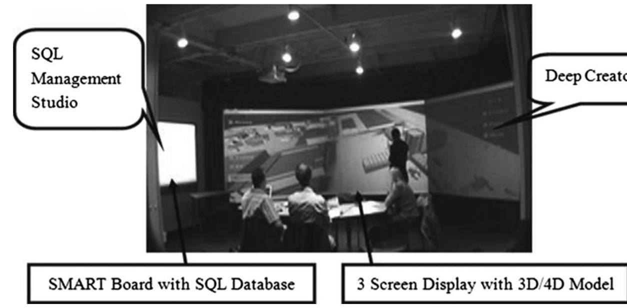

Fig. 2. Photo of a student group in the Immersive Construction (ICon) Lab using the VCS

Illustrative Example of Using Observational Time Studies in Construction Research

Time studies have been used in construction research, though only to a small extent of the potential. Several research studies investigating worker productivity has used time studies to track the time spent performing particular tasks in an effort to identify field challenges, differing conditions, or to identify means for improving worker efficiency (Gillbreth 1912; Taylor 1903). Though the nature of capturing quantitative data are seemingly in line with the proposed methodology, the use of time studies often does not rely on the qualitative data available but instead uses standard time and efficiency measures as a basis of comparison. Thus, these studies are only partially benefiting from the available opportunities. More recently, researchers studying the use of three-dimensional (3D) and four-dimensional (4D) software have used observational studies to demonstrate improved efficiencies of communication. Liston used observational studies to quantify the time spent in "less productive" conversations when changing from standard two-dimensional project documentation to a 3D computer-aided drafting (CAD) model and 4D CAD simulation of the construction process (Liston et al. 2001). The study was one of the first within construction to use the qualitative observations about the seemingly greater efficiency of communication when using such technology and support it with the quantitative data developed through coding communication using time studies of project meetings. In similar efforts, Wang and colleagues developed an educational module for improving sequencing knowledge through the use of 3D CAD and 4D software (Wang et al. 2007). Wang performed an experiment and then coded student groups using a new 4D simulation module to demonstrate shifts in time spent during different stages sequencing design during the development process.

To clearly present the use of observational studies with the use of content analysis for evaluation, a case study was performed. The use of this case study allows exploration of issues on how to implement this methodology, from defining what to measure, to developing the coding framework and reliability measures, by addressing the methodology steps point by point throughout the case study example. This study is an outgrowth of the work performed by Wang at Penn State evaluating the use of a virtual construction simulator (VCS) using an education module for construction sequencing. In an ongoing research project, a software change was made in the development of successive versions of the VCS interface. In the 2006 course offering, the students used an initial version of the VCS created in the Deep Creator game engine, while in 2007 the student used a new interface developed in the Irrlicht open source rendering engine. In both offerings the course had 10 groups of students randomly assigned to three or four member teams. When comparing the outcomes of the use of the VCS application for the education exercise, there was an observation that the shift to the new user interface with slightly different functionality including a slightly improved user interface for entering activity and sequence data, led to student groups spending a much shorter time period working on the project, and also yielded a decrease in the average project grade by 3%, which was not statistically significant. This led to the fundamental question regarding "why" the amount of time varied significantly for the groups along with quality differences in the final grade. Each group was video-taped during their exercise and this data provided an opportunity to investigate through observation whether the interface and functionality were mitigating factors in the shift.

In the first offering of the educational module in 2006, two different software interfaces were compared to identify changes in the process used to develop their sequence of construction activities. The students were asked to develop a short interval production schedule for the structural sequence of the MGM Grand renovation project in Las Vegas. For the study there were 10 participating groups in the course, which was an upper level construction project controls class in the Architectural Engineering curriculum at Penn State. All student groups were developed through random assignment, and they consisted of three or four members. The study was designed where one set of 5 groups used the newly developed VCS, and another set of 5 groups used a commercial scheduling and 4D CAD application for developing their construction sequence. In the analysis of the groups using the two different software packages, videos were taken of each of the groups working for a continuous period within the Immersive Construction Lab at Penn State. Each group was given up to 4 h, with all groups finishing in less time. In Fig. 2 a group is shown using the first version of the VCS in Penn State's Immersive Construction Lab. In the initial results the study showed a change in the time spent in different stages of the sequencing process, with less time spent trying to understand the problem and more time performing analysis and creating the sequence.

For the 10 videos recorded during the group meetings, the average duration of the videos in the first offering was 3 h, while in the second offering it was 50 min The set up of this particular study is simpler than most because the data, in this case videos for the groups using the different software packages, was already collected. Since the collection of the data was set up as an exploratory study for the use of the software, there is little room to control for the variables but they still need to be clarified for the purposes of identifying limitations of the analysis and use of the outcomes. All of the class members were either in the final year of undergraduate work in Architectural Engineering or graduate students. This gave the population as homogeneous a range of par-

J. Constr. Eng. Manage., 2010, 136(1): 76-86

ticipants as possible given the experience and knowledge about the scheduling exercise they were undertaking. The participants were all 22–24 years of age and in their final year of the Architectural Engineering program at Penn State, over 90% of each group was Caucasian, and 70% of each group was male.

Research Question

Having identified the dynamics of the exploratory study, a question to investigate using observational methods is necessary to move forward with identifying a framework for analyzing the data. Potential differences were identified along two main areas of interest. First, since the change between the two groups in the study was believed to be mainly a more user-friendly interface, the interaction between the student groups and the software interface was a suitable starting place for trying to identify differences in how they were used. The intention in studying this was to see if there are clear differences in the time or manner of use of the software.

The other potential area of difference between the two groups is in the dynamics of the groups themselves. This area of research could potentially be a very challenging area to study, and the intention with this exploratory study was to see if there is some evidence to show that the interface is the main reason for the change in the results. The interest in studying this area is to identify if there are noticeable differences in the group dynamics that could be a greater factor than the software interface.

We employed the guidelines suggested by Holsti (1969) for using content analysis. Although these guidelines do not fully eliminate potential biases, they do minimize the possibility that the findings will reflect the analyst's subjective bias rather than the actual content.

Step 1. Identify the Categories

Within the use of the VCS interface, two main categories were chosen to compare the two groups of participants, reflected in the core research questions: (1) interaction with the software and (2) discussion between the team members. For software interaction, the subcategories were broken into: authoring content, navigating the model, and analyzing outcomes. For team member discussion, the subcategories were tracking the time each team member spends discussing the task. Along with tracking the team member discussion, the lab facilitator was tracked as well. Since the study used software which was new to the student groups, there was a facilitator available to answer questions or help if there was difficulty using the interface. This was coded like the team discussion, by tracking any time the facilitator spoke with the team. Although these categories may appear simple, they were discussed in detail to identify which actions by the students could at the end of the process indicate the differences in value of the two interfaces.

Another approach was considered for the case study analysis wherein the coding would track the number of specific references to the model. As such, it would be necessary to code references to the model, as well as determine how to code physical gestures, verbal references, and when both are used together. The question which ruled out the use of this approach was what information will we have when complete, and will we be able to use that information to draw conclusions? In this case, the results of the content analysis would produce frequency information about the discussion references to the model, but the information would only convey how often they used the visual information in the model and not how they actually used the software interface. Since the visual information in the model was identical between the two study groups, the information in the outcomes would not be related to the hypotheses or allow for conclusions of value about the interfaces. The reason behind this seemingly tangential discussion about the case study is not the details of the framework used, but the designing of the framework and why was one approach chosen over another. Similar to design intent explanations architects and engineers use to explain facility systems at early design stages, the intent of a coding framework should be clear and should consider the type of results the analysis will return.

Step 2. Distinguishing between Categories

With the main and subcategories for the framework identified, the definitions for the subcategories were developed to clearly communicate the differences in interaction. The definitions should be simple, concise, and clear to minimize ambiguity. For the software interaction, the three subcategories were developed and defined as follows:

1. Authoring content: using the software interface to create or document information or content related to the task at hand, e.g., typing in activity durations or selecting geometry in the model to create groups;
2. Navigating the model: using the software interface to move through the model or information contained therein either for general exploration of the information or to locate specific information, e.g., rotating the model to view it from the top or side; and
3. Analyzing outcomes: using the software interface to manipulate the data to synthesize new information/results or to create new ways of viewing the data. Examples include performing calculations to get results or developing an animation, e.g., reviewing a 4D simulation.

For the discussion among the team members, the category was the actual team members, rather than separate categories of discussion. The members were generically labeled 1–4, starting from the left side of the video and continuing to the right and clockwise, depending on how the team members orient themselves. The breakdown of the discussion in this case is more a question of what discussion constitutes discussion worth coding for analysis.

Step 3. Identifying the Content Unit

The content unit can vary with the intent of the coding, and the manner of breaking it down impacts the type of analyses which can be performed at the completion. For the discussion content of the groups, the focus was not on the number of contributions an individual made but was on the amount of time each team member spoke. Under this scenario the definition only needs to cover what meets the minimum standard for the speaker to be coded as part of the discussion. To consider a verbal statement as a contribution to the discussion by one of the group members it needed to meet the following two criteria:

1. The contribution, either a question or statement, needs to constitute a whole statement and complete thought. Started but not completed statements and simple one word comments, such as yes or why, were not coded as contributing time.
2. The contribution needs to be related to the topic at hand. The participants in the study are working for an extended period

J. Constr. Eng. Manage., 2010, 136(1): 76-86

of time and it is expected that tangents to the conversation will take place.

Step 4. Clear Reasoning behind the Inferences

The reasoning in the background of the coding process constitutes the underlying backbone of the research outcomes. It was necessary in the development of the definitions and content unit to pilot the coding process using the videos from the study. The piloting process allowed for clarification and examples to be used to help train multiple coders and also for iterations and clearer reasoning behind the framework definitions to be developed. The reasoning and clarifications developed during this process was, and usually should be, incorporated into the category explanations to demonstrate the intent of the categories.

The interaction with the software interface was coded to identify if the design or the user friendliness of the interfaces affected the educational outcomes. By tracking the different manners of its use and their times, we could determine if the two software interfaces were used in a similar manner or in a different manner, and also were able to identify if that was the factor which affected the project outcomes.

Ensuring Reliability

There were two issues identified as central to ensuring impactful research. The first was validity and the elimination of covariation in the collected research data, which is ensured through the design of the research task. In the case study, the research is using data from previous studies and though covariation is minimized through the use of similar pools of participants, there is the still some small amount of error created from the use of different offerings of the course and dynamics of the participant pool which cannot be controlled. The second issue raised was the reliability of the content analysis for ensuring reproducibility of the results, often referred to as inter-rater reliability. The method used identifies the level of coding errors among the various coders as defined in Eq. (1)

$$\text{Level of coding error} = \frac{\text{Time in agreement}}{(\text{Time in agreement} + \text{Time in disagreement})} \quad (1)$$

The act of coding a video for this level of framework took an individual coder approximately three to four times the duration of the video being coded which made testing the reliability of the framework for the entire data set unrealistic for an exploratory study until there is some evidence to support the hypotheses. All of the chosen coders had experience with the sequencing activity, with 4D modeling techniques, and were trained in the coding and use of the framework. Determining an appropriate level of inter-rater reliability should be very thoughtful since "defining an acceptable level of reliability is one of the many problems in content analysis for which there is no single solution" (Holsti 1969). The reliability standards need to be answered within the context of a given research problem (Krippendorff 1980). Since the data set was relatively small and the framework simple, two coders were able to perform all of the video coding, and these coders were then used to determine the level of reliability. For more challenging frameworks and larger data sets the reliability would need to account for all of the coders and may utilize one of the other reliability tests suggested previously. The agreement level between the coders exceeded the standard of reliability using a 95% standard which had been predetermined as the level expected for the framework, as shown in Table 1.

Rigorous Coding versus Data Sampling

One of the challenges which was identified and reasons why the recent developments of technology has help invigorated observational studies is the use of data sampling to represent the true events which take place in an observational setting. In live observational studies, typically, the observers will note activities as regular intervals, similar to the way a time lapse video shows steps in activity without showing all of the activity between frames. However, as with any study there is efficiency in the process if a representative sample can be used to represent a larger population. Using video enabled software to code allows for very exact time to be measured for an activity. However, there is an intense labor effort involved in generating the quantitative data from the observational video. The challenge then is to identify when a representative sample can be used, and when coding of an entire study is worthwhile.

For the undertaken study, there were two areas of interest raised and each had a correlating category within the coding framework. It was decided for the study, to undertake a thorough and complete coding of the software interaction because the software is the known difference between the two participant populations. When considering the use of the software, if there is a difference in how the groups utilize the software then a sampling approach would be significantly more likely to miss significant data about how the groups used the software. From the preliminary studies performed by Wang, it was shown that using the software shifted the process for undertaking sequencing. If a similar shift occurs from the first to the second version of the software, it would be challenging to choose truly representative samples of the groups' activities.

On the other hand a sampling approach was taken for the group discussions. The interest in the group discussions was more exploratory and rigorous coding of the issue would be time intensive with the possibility of no valuable data. For this reason random sampling of the discussion was coded to see if the data suggested substantial differences and value in further analysis or more detailed coding. For the discussion, the interest is not the process of use as it is with the software interface, but purely in the dynamics of the group. The group dynamics are unlikely to change to such an extent that random samples of the discussion would be largely different than considering the whole duration of activity. To sample the discussions it was decided to code five

Table 1. Coder Reliability Levels for a Sample Portion of Video Observation

Category	Coder 1 Total time (min:s)	Coder 2 Total time (min:s)	Reliability level
Navigating the model	2:11	2:42	96.6%
Authoring content	1:00	1:31	95.7%
Analyzing outcome	0:00	0:00	100.0%
Person 1	0:47	0:50	99.7%
Person 2	2:04	1:44	97.8%
Person 3	2:59	3:15	98.3%
Person 4	6:05	6:38	96.4%
Facilitator	0:00	0:00	100.0%

J. Constr. Eng. Manage., 2010, 136(1): 76-86

intervals of 10 min each. This provides a significant amount of time for groups to hold discussion and allows for lulls in conversation relating to tasks such as calculation or use of the software. It is important to note that there is additional error incorporated into the results for the discussion coding that is not present in the software coding because of this choice.

Results

When reviewing the videos taken during a study, content analysis is a valuable tool for deriving the quantitative measures, but the first step is to review the video and observe the content and group activity. The quantitative measures follow to demonstrate how often an action takes place or how much time is spent performing a given activity. For the case study there were two areas of interest noted and these will be addressed first.

It was clearly evident from observing the videos that the second version of the software was easier to use, more intuitive, and user friendly. For the second version of the software, none of the groups spent more than 20 min creating their sequence, whereas the groups using the original version of the software all spent over 50 min creating their sequences. Also, the facilitator spent more time assisting the teams using the original version of the software, while the facilitator helping with the second version rarely had to assist with the software, if at all.

However, when watching the manner of use of the software, the overall process of interacting with the software was very similar. The software was oriented around creating a 4D model of the construction sequence. All of the groups generated their sequence, created the schedule using the software, and then reviewed the 4D sequence. All of the groups went through a similar process in this regard, and after reviewing the 4D simulation, none of the groups went back to their schedule and made significant changes. In all of the cases there were adjustments made to the schedule, typically to perform minor adjustments when the team missed a component in the sequence or an element was out of order. The average time spent analyzing the simulation for the first version of the software was 4.5 min while for the newer version it was 3.6 min. When reviewing the timing of the 4D analysis by the groups, all of the groups performed the review within the last 15 min of their time in the lab (Table 2).

There was a clear difference in the amount of discussion held among the groups depending on which version of the VCS interface was used. For the groups using the first interface, each team member spoke during the sampled video for an average of more than 11 min, while for the second interface it was only just over 6 min. These results indicate that groups using the first version of the software spent approximately twice as much time discussing the problem than with the newer version of the software. Also, as shown in Table 3, the time any individual spent speaking varied from as little as 5 to over 60% of the time. This indicates that there was no typical time one team member spent speaking, and the variation was consistent between the two versions of the software interface. Considering that the range of time speaking was consistent, but there was a clear difference in the overall time spent in discussion, there is some different dynamic between the two groups regarding the approach to the problem.

Looking at the same consideration from a different angle, the first time the teams chose to author content for the simulation was identified within each video (Fig. 3). The groups using the first version typically spent a greater amount of time in preparation before moving to the use of the software and took more time to

Table 2. Time Spent Using the Software by Each Group, Shown in Minutes, and as Percentage of the Total Interaction Time with the Software

		VCS 1				VCS 2			
		Group 1	Group 2	Group 3	Average	Group 4	Group 5	Group 6	Average
Navigating the model	Time in minutes (time as %)	7.87 (6.6%)	13.02 (17.2%)	14.71 (19.7%)	11.87 (13.2%)	41.02 (60.0%)	25.3 (77.3%)	20.15 (63.8%)	28.82 (65.2%)
Authoring content	Time in minutes (time as %)	108.93 (90.9%)	59.57 (78.7%)	52.61 (70.5%)	73.70 (81.8%)	18.58 (27.2%)	7.21 (22.0%)	9.67 (30.6%)	11.82 (26.7%)
Analyzing outcome	Time in minutes (time as %)	3.06 (2.6%)	3.14 (4.1%)	7.33 (9.8%)	4.51 (5.0%)	8.81 (12.9%)	0.21 (0.6%)	1.75 (5.5%)	3.59 (8.1%)
Total interaction time with the software	Time in minutes	119.86	75.73	74.65	90.08	68.41	32.72	31.57	44.23

J. Constr. Eng. Manage., 2010, 136(1): 76-86

Table 3. Percentage of Total Conversation Time That Each Team Member Spoke during the Meeting

	VCS 1[a]			VCS 2[b]		
	Group 1 (%)	Group 2 (%)	Group 3 (%)	Group 4 (%)	Group 5 (%)	Group 6 (%)
Person 1	26.8	61.9	34.6	9.8	37.2	12.2
Person 2	12.2	22.2	50.3	14.8	37.2	30.6
Person 3	55.8	15.9	15.1	14.8	24.8	57.2
Person 4	5.2	[c]	[c]	60.6	0.7	[c]

[a]VCS 1 calculated from a sample of the meeting.

[b]VCS 2 calculated from entire meeting.

[c]The group only contained three members.

prepare before using the software. The simulation using the software was one task within the assignment, and the teams using the original version seem to have planned to perform the whole class assignment during the time they were in the lab, and spent time performing calculations and holding discussions for how the construction progresses before considering how to detail the erection sequence. The groups using the second version of the interface moved directly into creating the content for the simulation. The groups using the second version seemed to have had their preparatory work done in advance of their time in the lab, and used the time to focus only on the sequencing aspect.

Along with the results which allowed the hypothesis to be addressed, there was another potential indicator noticed during the observation. As noted above, the facilitator spent a noticeably greater amount of time assisting with the first interface. When watching these discussions, it was noted that in order to utilize the first version of the interface, there were certain assumptions made in the creation of the software which affected the manner of data input. The software was created to help students find the "best" way to sequence the construction, so the interface was designed to complement better strategies. When the facilitator was interacting with the student groups, there was often dialog about how to group geometry or create schedule activities to best work with the software interface. The discussions, however, often induced some changes regarding the breakdown of the construction sequencing. These discussions may have helped the teams using the first interface to better schedule their sequences, which the teams using the second interface were not able to receive. The results in this regard were not conclusive, because the content of the discussions between the facilitator and the teams were not the focus of the analysis, but this dialog could be reviewed in more detail to evaluate the potential impact.

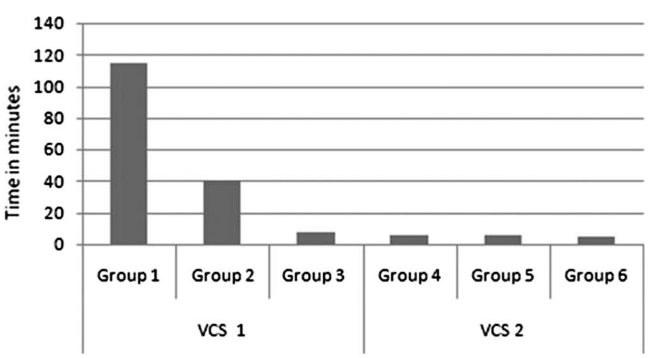

Fig. 3. Figure shows the time elapsed in minutes when the groups first authored content for the simulation

Discussion

Though the case study example is a simple study using an exploratory question, the potential implications of this mixed methodology study within construction research is quite powerful. Examples of current mixed-method studies are readily available for a multitude of varying research questions. For example, in a series of studies designed to examine leadership, Mumford and colleagues (Mumford 2006; Mumford and Van Doorn 2001; Strange and Mumford 2005) content analyzed biographies of over 200 historic leaders. By applying strict coding methods and using well-trained coders, the researchers were able to apply traditional statistical techniques to largely qualitative data. The results of their research stood in sharp contrast to current ways of thinking about leadership. What is underscored by these studies is that by applying these mixed methods, researchers may observe phenomena and trends that would not be available using either qualitative (e.g., single case study) or quantitative (e.g., survey) methods alone. Turning to more traditional observational approaches, Amabile and colleagues (Ruscio and Amabile 1999; Ruscio et al. 1998) have examined the creative process in several studies. Using videotaped accounts and think-aloud protocols, the researchers once again applied traditional coding techniques to the study of innovation in a natural setting. By using well-trained raters, the researchers were again able to apply traditional statistical techniques in their study. What is particularly notable about the work by Amabile and colleagues is how well they were able to study creativity—a complex phenomenon that is, at a minimum, a challenge to investigate. More precisely, by applying both qualitative and quantitative methods the researchers were able to shed light on a topic that is in desperate need of greater understanding. Although countless other examples exist, these should suffice to make our basic point: observational research—particularly mixed-method observational research—has a number of unique advantages that make the challenges associated with the method worthwhile. The implications for the unique interactions involved in the design and construction process offer an opportunity to utilize these advantages. The emergent question, then, is how to conduct this type of research in such a way that maximizes the researchers' capacity to both understand and convey their findings to others. The integration of the qualitative and quantitative data sets allows for the researcher to draw inferences from the qualitative video, and review the quantitative data for the level of impact, or correlation between variables.

While the case study was only exploratory in nature, the intention was to demonstrate the process and how the qualitative and quantitative results can be used to complement each other. The proper design of the study is essential for the researcher to be

J. Constr. Eng. Manage., 2010, 136(1): 76-86

able to draw valuable data from the video and to be able to correlate the coded content with the research questions or hypotheses. It is strongly suggested that any researchers interested in pursuing this type of work spend time testing and piloting the method, both in terms of the framework used and how to draw relationships between the qualitative and quantitative aspects of the methodology.

Conclusions

The construction industry is dynamic in nature with constantly changing teams, team members, settings, and methods. The use of observational studies within construction can create quantitative data about the processes and interactions used, while being able to fit that data within a qualitative context can prove very valuable in a range of research topics.

The example observational study presented in the paper illustrates a technique for developing rigorous data collection and analysis techniques for observational studies using content analysis and time study data from the coding of audio-video recordings. The case study demonstrated the process for developing a structured coding framework, incorporated rigor and reliability into the coding process, and presented some examples of how the observed content could be represented in the quantitative data developed during the content analysis. While the study was exploratory in nature, the ability of the tool to demonstrate quantitatively what is visually observed creates great potential for employing more qualitative research techniques in a more rigorous manner.

Challenges and Limitations

> "Hence efforts to analyze motives, values, attitudes, and other interesting but elusive variables necessarily require imaginative and often painstaking efforts to validate inferences drawn from component data. Quantification of documentary materials may yield important and interesting data about many aspects of human experience and behavior, but the temptation to count things for the sake of counting, unless resisted, is almost certain to yield precise findings which are meaningless, trivial, or both" (Holsti 1969).

The use of content analysis for observational studies enables the creation of vast amounts of quantitative data which can support research efforts. The process of generating the data are slow and painstaking, even using the available software for such efforts. Thus it is essential to focus the coding only on the content which will contribute to the research question at hand. Should a subcategory need to be refined or a new category added, the content can easily be recoded, but the time spent generating data that will never be used is lost forever. Piloting the overall process and coding framework on a single video or small set of videos is strongly suggested to quickly refine the framework from the content which does not contribute to the overall research efforts.

Observing and coding human interaction and communication is complex and even a trained analyst with keen insight may find it difficult to make maximum use of their data unless they use systematic methods. As with all research, proper design of the study and thorough attention to detail is essential for the collection of data. The methodology presented offers new means for combining qualitative and quantitative data, but cannot overcome the bias and covariation introduced in the original design of any study.

Along with the challenges and limitations of the coding, the challenges inherent in observation as a methodology are always difficult to foresee. The example from the case study regarding the facilitator assisting with the use of the interface shows that the observer can inadvertently alter the interactions and outcomes of the research participants. This limitation and challenge of the methodology can never be wholly removed, but using technology to take the person observing out of the setting or finding means to make the study more rigorous through larger samples and rotating facilitators could minimize the error created by these variables. One benefit of using observational research methods is that there is a better opportunity for identifying these type of experimental design issues since the researcher can review the specific activities that occurred during the experiment. In other more quantitatively focused studies, this type of subtle interaction which can influence the data collected, may easily go undetected.

Potential Uses for Observational Studies

There are many potential uses for observational studies, and this section is by no means exhaustive. When we wish to gain a better understanding from human interaction or behaviors, it is important to observe and study behaviors. Construction has many such interactions, both in industry and academia. Studying the team dynamics of group interactions can provide valuable insight into how people lead and make decisions. Further studies into leadership, team dynamics, and decision making within construction are all important areas that could benefit from more observational studies. The use of observational studies will create opportunities to track field construction using observation to identify trends or techniques to be further studied, and allow more quantitative analyses of the trends.

Another important area where many observational studies are performed is within human computer interaction. The ability to study and refine technology so that it is more easily used can improve the productivity of those who use the computer applications. Computers and software are rapidly growing in popularity in construction professions and on job sites. Finding ways to maximize the value through improved efficiency of these processes is ongoing, but tracking how it can be used in various forms on construction projects, similar to the case study presented, could identify methods of better using computers for communication and construction process planning.

In addition to group dynamics, observational studies can also be beneficial for studying jobsite activities, with a particular focus on productivity and safety. It is critical that we continue to gain a better understanding of the actual performance, priorities and decision models of our workforce, and best methods for investigating these factors through observational research methods is important. The site of construction activity is ever changing and through observational studies it is possible to quantitatively document these activities, while still considering them within the context of their projects and sites. The potential for this methodology to be applied within construction research is extensive and thus far has been underutilized.

Acknowledgments

We would like to thank the student participants for agreeing to be observed and recorded during the study presented in this paper. We also thank the National Science Foundation for their support

J. Constr. Eng. Manage., 2010, 136(1): 76-86

of this research through Grant Nos. 0348457 and 0342861. Any opinions, findings, conclusions, or recommendations are those of the writers and do not reflect those of the National Science Foundation or the project participants.

References

Aiken, L. S., West, S. G., and Pitts, S. C. (2003). "Multiple regression analysis." *Comprehensive handbook of psychology. Volume 2. Research methods in psychology*, J. A. Schinka and W. F. Velicer, eds., Wiley, New York.

Atkinson, P., Coffey, A., and Delamont, S. (1999). "Ethnography: Post, past, and present." *J. Contemp. Ethnogr.*, 28, 460–471.

Bauer, M. W., Atkinson, P., and Gaskell, G. (2000). *Handbook of ethnography*, Sage, Thousand Oaks, Calif.

Bernardin, H. J., and Buckley, M. R. (1981). "Strategies in rater training." *Acad. Manage. Rev.*, 6, 205–212.

Biello, D. (2006). "Fact or fiction?: Archimedes coined the term "Eureka!" in the bath." *Scientific American*, ⟨http://www.sciam.com/article.cfm?id=fact-or-fiction-archimede⟩ (July 24, 2008).

Bliese, P. D. (2000). "Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis." *Multilevel theory, research, and methods in organizations*, K. J. Klein and S. W. J. Kozlowski, eds., Jossey-Bass, San Francisco, 439–381.

Buchholz, B., Paquet, V., Wellman, H., and Forde, M. (2003). "Quantification of ergonomic hazards for ironworkers performing concrete reinforcement tasks during heavy highway construction." *Am. Ind. Hyg. Assoc. J.*, 64(2), 243–250.

Campbell, D. T., and Stanley, J. C. (1963). "Experimental and quasi-experimental designs for research on teaching." *Handbook of research on teaching*, N. L. Gauge, ed., Rand McNally, Chicago.

Conduitt, J. (2006). "Conduitt's account of Newton's life at Cambridge." *Newtonproject*, King's College, Cambridge, Keynes Ms. 130.4, ⟨http://www.newtonproject.sussex.ac.uk/texts/viewtext.php?id=THEM00167&mode=normalized⟩ (Apr. 27, 2008).

Cook, T. D., and Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis for field settings*, Rand McNally, Chicago.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, 1st Ed., John Murray, London, ⟨http://darwin-online.org.uk/content/frameset?itemID=F373&viewtype=text&pageseq=1⟩ (May 27, 2008).

Denzin, N. K., and Lincoln, Y. S. (1994). *Handbook of qualitative research*, Sage, Thousand Oaks, Calif.

Dov, N. (2005). "Developing an observation model in the conventional and simulation lab lessons: Observations on learning and instruction processes." *Proc., 2005 ASEE/IEEE Frontiers in Education Conf.*, IEEE, Indianapolis.

Gill, J., and Johnson, P. (2002). *Research methods for managers*, Sage, Thousand Oaks, Calif.

Gillbreth, F. B. (1912). *Primer of scientific management*, Van Nostrand Reinhold, New York.

Gopinath, R., and Messner, J. I. (2004). "Applying immersive virtual facility prototyping in AECO." *Proc., CONVR 2004: 4th Conf. of Construction Applications of Virtual Reality*, Dept. of Information Sciences and Technologies of ISCTE, Lisbon, Portugal, 79–86.

Gyorke, A. (2006). *Studiocode white paper: A video coding and analysis tool with applications for education and research*, Pennsylvania State Univ., University Park, Pa., ⟨http://ets.tlt.psu.edu/whitepapers/Studiocode_Whitepaper.pdf⟩ (Aug. 5, 2008).

Hammersly, M., and Gromm, R. (1997). "Bias in social research." *Sociological Research Online*, 2(1), ⟨http://www.socresonline.org.uk/socresonline/2/1/2.html⟩ (July 29, 2008).

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*, Addison-Wesley, Reading, Mass.

James, L. R., Demaree, R. G., and Wolf, G. (1984). "Estimating within-group interater reliability with and without response bias." *J. Appl. Psychol.*, 69, 85–98.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*, Sage, Thousand Oaks, Calif.

Liston, K., Fischer, M., and Winograd, T. (2001). "Focused sharing of information for multi-disciplinary decision making by project teams." *J. Inf. Tech. Constr.*, 6, 69–82.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., and Cannon-Bowers, J. A. (2000). "The influence of shared mental models on team process and performance." *J. Appl. Psychol.*, 85, 273–283.

Messick, S. (1995). "Validity of psychological assessment." *Am. Psychol.*, 50, 741–749.

Mintzberg, H. (1973). *The nature of managerial work*, Harper & Row, New York.

Mumford, M. D. (2006). *Pathways to outstanding leadership: A comparative analysis of charismatic, ideological and pragmatic leaders*, Erlbaum, Mahwah, N.J.

Mumford, M. D., and Van Doorn, J. R. (2001). "The leadership of pragmatism: Reconsidering Franklin in the age of charisma." *Leadership Q.*, 12, 279–309.

Plano Clark, V. L., Creswell, J. W., O'Neil Green, D., and Shope, R. J. (2008). "Mixing quantitative and qualitative approaches." *Handbook of emergent methods*, S. N. Hesse-Biber and P. Leavy, eds., Guilford, New York.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, Sage, New York.

Ruscio, A. M., and Amabile, T. M. (1999). "Effects of instructional style on problem-solving creativity." *Creat. Res. J.*, 12, 251–266.

Ruscio, J., Whitney, D. M., and Amabile, T. M. (1998). "Looking inside the fishbowl of creativity: Verbal and behavioral predictors of creative performance." *Creat. Res. J.*, 11, 243–263.

Scott, M., and DeSanctis, P. G. (1992). "Microlevel structuration in computer-supported group decision making." *Hum. Commun. Res.*, 19(1), 5–49.

Strange, J. M., and Mumford, M. D. (2005). "The origins of vision: Effects of reflection, models and analysis." *Leadership Q.*, 16, 121–148.

Tabachnick, B. G., and Fidell, L. (2001). *Using multivariate statistics*, Allyn and Bacon, Needham Heights, Mass.

Taylor, F. W. (1903). *Shop management*, Harper & Row, New York.

Tukey, J. W. (1977). *Exploratory data analysis*, Addison-Wesly, Reading, Mass.

Wang, L., Messner, J. I., and Leicht, R. (2007). "Assessment of 4D modeling for schedule visualization in construction engineering education." *Proc., 24th Int. Conf. on Information Technology in Construction*, Univ. of Maribor, Maribor, Slovenia.

Wax, R. H. (1971). *Doing field work: Warnings and advice*, Univ. of Chicago, Chicago.

Woehr, D. (1994). "Understanding frame-of-reference training: The impact of training on the recall of performance information." *J. Appl. Psychol.*, 79, 525–534.

Yin, R. K. (2003). *Case study research: Design and methods*, Sage, Thousand Oaks, Calif.

J. Constr. Eng. Manage., 2010, 136(1): 76-86