Professional Development

Assessing SH&E Research

Key principles and practical strategies improve understanding

By E. Scott Geller

SH&E PROFESSIONALS ARE BOMBARDED with information that they need to consider as they strive to improve their efforts to address the human dynamics of injury prevention. This includes marketing brochures touting research evidence that "proves" a given approach is best. The term "research" is periodically used in presentations and journal articles, and at professional development conferences to support one safety perspective or intervention approach versus another.

Thus, SH&E professionals need to understand some basic concepts about research in order to assess the value of research evidence presented to them and become more evaluative consumers (Metzgar). They need to know how to discriminate between good and bad research; sometimes, inferior research is not clearly evident. Increasing SH&E professionals' knowledge and appreciation of applied research could also lead to more objective evaluation studies in occupational safety. and health. Not only should SH&E professionals propose important questions for empirical study, they should also recommend certain situations to evaluate, methodologies to implement, interventions to compare and interpretations to consider. They might also design, conduct and analyze their own field experiments and, thus, contribute to the limited research literature relevant to improving safety-related behaviors and attitudes.

This article reviews basic principles to help SH&E professionals better evaluate research results and formulate research questions and procedures. For example, to become better consumers of information, practitioners need to understand critical differences between case studies, surveys, laboratory experiments and field studies. With this understanding, they can also design and conduct their own research to evaluate various intervention techniques. Such studies can be convenient to conduct and require no financial support. In other words, SH&E professionals can actually collect objective data throughout their daily routines that not only improve decision making, but also motivate application of improved safety management strategies.

E. Scott Geller, Ph.D., *is a senior partner and cofounder of Safety Performance Solutions Inc., a consulting firm specializing in the human dynamics of safety, and director of the Center for Applied Behavior Systems in the psychology department of Virginia Tech, both located in Blacksburg, VA. A professional member of ASSE's Colonial Virginia Chapter, Geller is a Fellow of the American Psychological Assn., American Psychological Society and World Academy of Productivity and Quality.*

Empirical research is founded on the scientific method, which has three basic objectives: 1) to describe (e.g., case studies); 2) to predict (e.g., surveys); and 3) to control (e.g., manipulations of independent variables to study cause-and-effect relationships). Unfortunately, safety presentations often use a certain kind of research (e.g., a case study) for an inappropriate purpose (e.g., to claim understanding of a cause-and-effect relationship). Or, sometimes the manipulations of independent variables in a field study are flawed, and a purported cause-and-effect relationship is unfounded.

For these three types of research (descriptive, correlational and experimental), it is important to discriminate between reliable and unreliable measures, and between valid and invalid results. This increases a person's ability to distinguish between subjective opinions and objective evidence. Bottom line: When selecting or developing an intervention approach for occupational safety, more applications of the scientific method—and less reliance on common sense—are needed.

The Fallacy of Relying on Common Sense

Aubrey Daniels, a renowned educator and consultant in the field of organizational performance management, asserts he is "on a crusade to stamp out the use of common sense in business. Contrary to popular belief there isn't too little common sense in business, there is too much" (Daniels 10). He offers the following distinctions between common sense and scientific knowledge, reflecting the need to be cautious when relying on only common sense to deal with the human dynamics of occupational health and safety:

•Common-sense knowledge is acquired in ordinary business and living, while scientific knowledge must be pursued deliberately and systematically.

•Common-sense knowledge is individual; scientific knowledge is universal.

•Common-sense knowledge accepts the obvious; scientific knowledge questions the obvious.

•Common-sense knowledge is vague; scientific knowledge is precise.

•Common sense cannot produce consistent results; applications of scientific knowledge yield the same results every time.

•Common sense is gained through uncontrolled experience; scientific knowledge is gained through controlled experimentation.

Profound knowledge relevant to managing and improving the human dynamics of safety can only be gained through objective, systematic observation and evaluation (i.e., research). Research enables the verification or refutation of common sense and, therefore, enables people to improve their common sense. Let's begin with a brief review of the scientific method and its objectives. Its three basic objectives define three different research approaches that SH&E professionals need to appreciate.

The Scientific Method

Every reader has some knowledge of the scientific method, including the basic steps:

1) Define the problem.
2) Form a hypothesis.
3) Collect data.
4) Draw conclusions.

Years ago, the author and colleagues incorporated these basic steps into a comprehensive employee training process for behavior-based safety (BBS) (Geller, et al). Figure 1 depicts this DO IT process:

•D = Define critical target behaviors to increase or decrease (pinpointing the problem in BBS).

•O = Observe target behaviors and environmental conditions that could be influencing the target behaviors (collecting data and forming a hypothesis).

•I = Intervene to change the target behavior(s) in desired directions (this is an added stage to the scientific method when the research is applied and the primary objective is to improve performance).

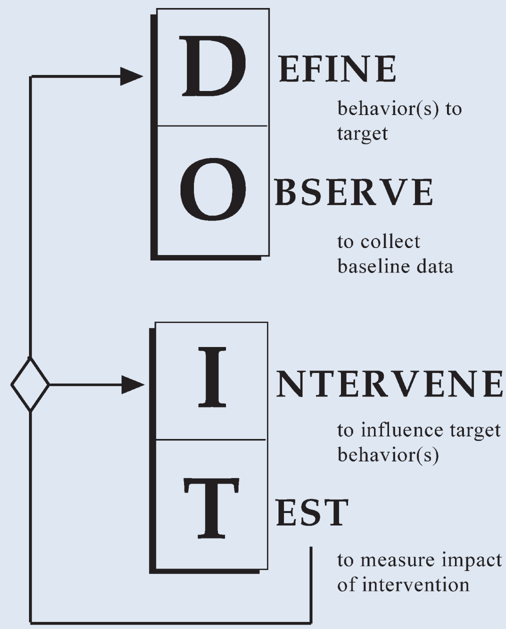
•T = Test the impact of the intervention procedure by continuing to observe and record target behaviors (this leads to drawing conclusions, in this case about whether the intervention has beneficial impact).

Note that the arrows in Figure 1 reflect the scientific method as a continuous process. That is, after drawing conclusions about the impact of an intervention (i.e., after the test phase), the applied researcher makes one of two choices. If the intervention worked sufficiently, the DO IT process starts over, with one or more additional behaviors targeted for improvement. If it is concluded that the intervention did not work as desired, the process returns to the intervention stage where the prior approach to behavior change is refined or an entirely new treatment is implemented and tested.

This process is a special application of the scientific method—when an intervention (or independent variable) is implemented to affect change in designated behavior (the dependent variable). Much scientific research is not of this form. But the scientific method has specific characteristics that distinguish it as a special approach for obtaining profound knowledge. Consider the following five qualities.

Figure 1

DO IT Process



Quality 1: An Empirical Approach

Researchers rely on direct experience rather than hearsay or other people's opinions, unless the others base their information on direct experience (or empirical evidence). Thus, researchers are skeptical of a person's opinion unless it is based on objective observation. It is critical to evaluate how empirical evidence is obtained, which leads to the next characteristic of the scientific method.

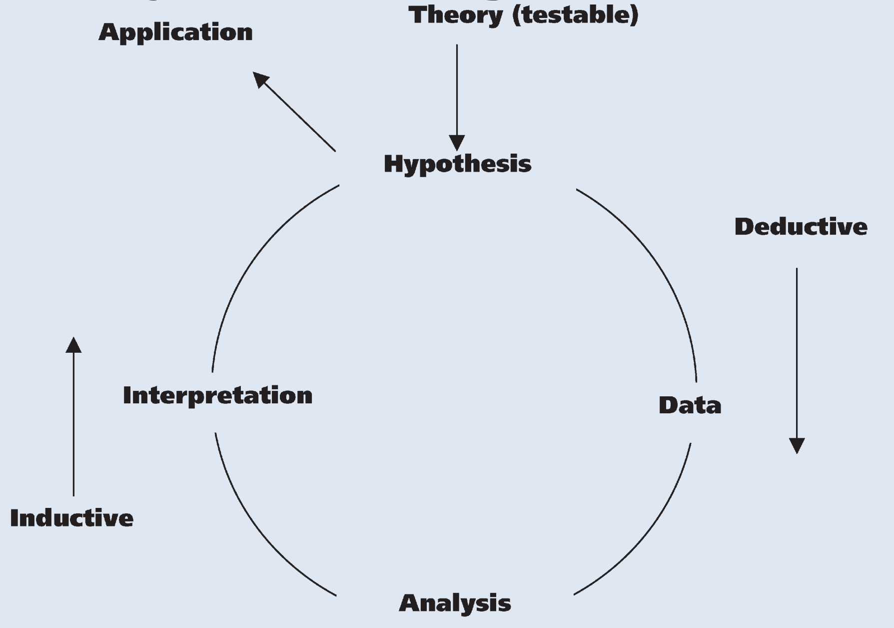
Quality 2: An Objective Approach

Reporting observations from experience must be objective, meaning that other people would have made the same observations had they been there. For this to occur, the procedures involved in the experience must be documented so other individuals could observe the same event. Thus, the scientific method involves careful and unbiased documentation of exactly what happened and what was observed. Objectivity means the experience observed by one person is available to any other person if s/he repeats the same conditions, including the observation process.

A primary process in BBS is objective observations of workers' behavior and the environmental conditions in which such behavior occurs. By making these observations with the aid of a checklist of target behaviors and environmental conditions, the

Figure 2

Theory: Guides & Organizes Research



results are more objective and likely more comprehensive. In other words, a behavioral checklist makes replication more likely because it is possible for another observer to follow the same procedures and obtain the same results.

Objectivity can be assessed by comparing two sets of observations. In BBS, this is accomplished by counting the number of times two independent observers agree and disagree with regard to a set of behavioral observations, then calculating the percentage of agreement with the formula: percent agreement = total number of agreements divided by the total number of behavioral observations (i.e., number of agreements + disagreements) multiplied by 100. If the observers do not reach at least 85-percent agreement, scientists would usually consider the observation data unreliable or not objective enough for dissemination as research-based information (Kazdin).

Quality 3: A Self-Correcting Process

The DO IT process reflects a continuous improvement process whereby interventions are progressively improved and new behaviors are successively defined and observed for beneficial treatment. For example, a work team with which the author consulted was not satisfied with an observed 10-percent increase in the use of safety gloves following the display of new safety signs. The team designed and implemented a new intervention. The group simply posted the daily percentage of employees in a certain work area who wore safety gloves. Then, a steady increase in glove use was observed, until usage reached nearly 95 percent. Subsequently, the work team defined another behavior to observe, intervene upon and test.

This process of never-ending improvement epitomizes science. Empirical evidence is constantly being discovered that contradicts prior knowledge and opinion. Thus, scientists are open to change and modify their opinions and common sense as a result of research results. For example, years ago, psychologists emphasized the impact of the environment on people's behavior and personality, and downplayed biological or hereditary influence. Today, however, as a result of research evidence, psychologists give more credit to biological and hereditary determinants of both behavior and personality (Bouchard; Roberts). In this case, common sense is founded on applications of the scientific method and changes only as a function of objective observation.

Quality 4: A Progressive Process

Unlike human activity in the arts and humanities, science is progressive. In other words, the knowledge obtained from the scientific method is tentative and continuously advances in terms of its accuracy or its ability to improve quality of life. In contrast, while paintings, music, literature and fashion certainly change, such change is usually not considered progress. Quality of life does not advance with periodic changes in popular clothes, music or art (McBurney).

Quality 5: Theory Testing & Development

Theory has two basic functions: 1) to integrate and organize facts from research into a parsimonious concept or principle; and 2) to guide the design and interpretation of research. Most contemporary research is theory-driven, and the focus of statistical analysis is to accept or reject a specific hypothesis deduced from a particular theory. Figure 2 depicts the role of theory in guiding and integrating research. As shown, research is designed to test hypotheses derived from theory. Development of an observation and data-collection process to evaluate a theory is deductive, meaning the research design defines a specific set of circumstances inferred from a certain theory or hypothesis. Data analysis supports the theory, rejects the theory or suggests that it be refined in particular ways.

The process of interpreting research data as supporting or not supporting a theory-derived hypothesis is inductive because general principles are generated from a particular research observation. The reasoning is from the more specific to the more general. The inductive interpretation of research observations often leads to theory modification and the derivation of another hypothesis to test. Research procedures are deduced from the new hypothesis, and the improvement cycle depicted in Figure 2 continues.

Sometimes, after successive theory testing, data interpretation and theory refinement, application is warranted. In other words, the scientific method enables objective refinement of a theory or set of principles that can be applied in real-world circumstances to benefit people and/or the environment in which they exist. In occupational safety, this means research has uncovered a procedure, process or intervention that can reduce the probability of unintentional injury.

However, clear, straightforward guidelines are not available for determining when the research and development of an intervention technique is sufficient for a transition to real-world application. Sometimes, individual, organizational or social needs influence premature use of an experimental procedure. Furthermore, marketing strategies can give an innovative treatment approach more credibility than it deserves, leading to risky applications. Thus, it is important to collect objective data on the individual and organizational impact of a program or process purchased from a consulting firm. The research methodology and validity principles discussed in this article are relevant to understanding the credibility of a procedure designed to evaluate an intervention technique, whether in the laboratory or field.

All research results should not be judged the same, even though all adhered to the scientific method. SH&E professionals need to know how to determine the relative credibility of a research methodology and outcome, especially when findings have apparent implications for the management of occupational safety. The three fundamental types of research reflect the three basic objectives of research: to describe, to predict and to control.

Descriptive Research

The objective of descriptive research is to obtain an accurate description of a particular situation or phenomenon. This approach attempts to neither predict future situations nor understand the cause of an observed event. Rather, it applies the scientific method to identify the various characteristics of a particular circumstance, sometimes to describe apparent relationships between certain aspects of the situation. Polls or opinion surveys are descriptive. The value or usefulness of the description depends on the source of the data. If data are nonempirical and subjective, based on intuition rather than unbiased observation, the description is not useful from a research perspective.

On the other hand, an objective description of an event or phenomenon can provide information that is relevant to developing a theory for research testing. For example, clinical psychology advanced substantially following objective description and classification of various unusual behaviors observed in clients. Often, such descriptions are referred to as case studies, describing individual events or circumstances that present themselves as opportunities for systematic research.

SH&E professionals conduct descriptive research when completing an injury report. Careful analysis of an injury-causing event or situation reflects descriptive research and can be invaluable in preventing future injuries. However, such research should not be considered a "root-cause analysis." While the analysis may be useful, it is not possible to discover the root cause of a mishap from descriptive research. Defining a cause-and-effect relationship requires much more experimental control and observational rigor than is possible from even the most careful and comprehensive description of an injury-producing incident.

Survey Research

Perception surveys are popular tools in the SH&E profession. In fact, some professionals call them "one of the more recent and best measurement tools" (Petersen 169). Surveys can indeed be used to assess human factors such as attitudes, beliefs, personal opinions and personality characteristics that contribute to workplace safety. They can even be used to assess corporate culture. However, perception surveys have limitations. Therefore, understanding basic principles that underlie their development, administration and interpretation can help the SH&E professional make appropriate decisions regarding their use.

First, one must realize that most surveys provide only descriptive information. They do not imply cause-and-effect relationships and, therefore, should not be used to evaluate the direct influence of an intervention procedure. Some surveys can be used to predict performance, but few surveys do this well. When is a survey a useful research tool? Answering this question requires some basic information about reliability and validity.

Reliability & Validity

What is the practical value of a questionnaire or survey? This can be assessed with various research methods and statistical tools. Although many are beyond the scope of this article, a few basic concepts are pertinent. First, questionnaires to measure perceptions, attitudes or person factors can be reliable, although not valid. However, to be valid, they must be reliable. A reliable survey gives consistent results. This is assessed by comparing answers across different survey items that supposedly measure the same factor, or by comparing two different administrations of the same survey. For example, if a scale indicated that an individual weighed 250 pounds on Monday, 249 pounds on Wednesday and 251 pounds on Saturday of a given week, the scale would get a high reliability rating, even though the person really weighs only 150 pounds. This scale gave consistent results, so it is reliable, but the numbers are invalid.

Types of Validity

Validity refers to whether the survey instrument measures what it claims to measure. For example, does the perception survey really assess attitudes toward occupational safety and health? Does the personality scale measure a person's "injury proneness" (i.e., likelihood of receiving a personal injury) or "injury preventiveness" (i.e., willingness to participate in a safety-improvement effort)? There are three basic types of validity for a measurement scale, each with particular experimental and statistical methodologies for evaluation: 1) content validity (do relevant experts agree that the survey appears to measure what it is supposed to measure?); 2) criterion validity (can scores from the survey be used to predict individual behavior or performance?); and 3) construct validity (are the relationships found with the survey consistent with relevant theory and research?).

In the weight example, the scale appears to meas-

Profound knowledge relevant to managing and improving the human dynamics of safety can only be gained through objective, systematic observation and evaluation.

Common sense might be the basis for an original theory, but it must be modified according to the results of real scientific inquiry. Common sense can benefit from research.

ure the correct weight of a person standing on it (content validity), but if results were compared with readings from another scale, or with results of another estimate of weight, the numbers would not correspond. Construct validity would be questionable. Plus, this scale could not predict other variations in individual performance, such as running speed or calories consumed per day, presumed to be influenced by weight. Thus, criterion validity could not be demonstrated.

Types of Criterion Validity

Criterion validity can be evaluated with two different techniques: concurrent and predictive validity. Concurrent validity is most frequently used and refers to the relationship between the scale results (e.g., weight in pounds) and another simultaneous assessment of the factor that the scale is supposed to measure. This assessment could result from measuring one's weight with another scale or visually estimating that person's weight.

A valid measure of "injury proneness" should correlate significantly with the number of injuries employees experience, whereas an assessment of "injury preventiveness" should vary directly with a measure of safety participation (e.g., number of safety audits conducted, safety hazards removed, safety meetings attended).

Predictive validity is more difficult to assess, but it is the ultimate objective of most research applications of surveys. It refers to the ability of an evaluation tool to predict future behavior. In the weight example, testing predictive validity requires that the scale information (i.e., how much someone weighs) is compared with a future outcome the scale is purported to predict, such as a person's quickness, general health or diet. Determining whether measures of "injury proneness" and "injury preventiveness" obtained at employee orientation predict the number of workplace injuries and amount of employee participation in a safety-improvement effort, respectively, also exemplify tests for predictive validity.

Types of Construct Validity

The construct validity of a scale is usually evaluated with tests of convergent and divergent validity. Convergent validity refers to the extent to which other measures of the same construct (e.g., a visual estimate of one's weight) relate to each other. Divergent validity indicates the extent scores from surveys unrelated to the construct do not correlate with survey scores related to the construct. In other words, divergent validity implies the extent to which a particular questionnaire measures special characteristics not measured by other scales. Theoretically, injury proneness should not be related to injury preventiveness, so measures of these constructs should not correlate. And, if such measures do not correlate, divergent validity would be demonstrated.

When Validity Is Critical

Regardless of how a person scale is used—whether it is for teaching, pinpointing problems, or measuring trends or change—it is important to use measurement tools with acceptable levels of reliability and validity. However, if results of a person scale are only used to teach diversity or to measure group perceptions, statistically unacceptable levels of reliability or validity will not cause harm or injustice to anyone. However, if a questionnaire is used to identify and select individuals for a particular job, prior research with the scale must have demonstrated acceptable levels of criterion and construct validity.

What Is Acceptable?

The basic statistic used to measure reliability and validity is a correlation coefficient, which describes the relationship between two sets of survey scores with a number ranging from -1.0 to 1.0. The greater a positive number (between 0 and 1), the greater the direct relationship between measures (a high score on the predictor scale indicates a high score on the criterion). Negative correlations (between 0 and -1.0) indicate an indirect (or inverse) relationship (a high score on the predictor scale indicates a low score on the criterion); the closer the correlation to -1.0, the greater the inverse relationship.

The closer the correlation to 1.0 or -1.0 and the larger the sample, the more confident one can be that the relationship is true. However, one must recognize the difference between statistical significance and practical significance. For example, a correlation of 0.30 would be statistically significant for most statistical tests and sample sizes. In some cases, however, this number may not represent practical significance, given that the square of the correlation coefficient indicates the degree of variance overlap between the two measures.

For example, a correlation of 0.30 between the results of a safety perception survey and other measures of safety, such as employees' frequencies of completing coaching sessions or percentages of at-risk behavior per observation period, sounds good until one realizes that only nine percent of the variance in one measurement device could be accounted for by the other (0.302 = 0.09 or nine percent). In this case, 91 percent of the variance in people's safety perception scores could not be explained by the other estimate of a person's safety.

Importance of Construct Validity

It is possible that a direct relationship can be found (predictive validity) between a predictor (such as a measure of accident proneness) and a criterion (such as number of at-risk behaviors or recordable injuries) without supporting the underlying principle(s) or theory. This would indicate the absence of construct validity. For example, suppose an individual could determine how to answer survey questions in order to receive a favorable score. Construct validity would then be questionable, even if criterion validity were high.

Every perception survey the author has seen has included items that are transparent and enable a respondent to "fake good." In research literature, this is called impression management (Schlenker;

Umstot), and it leads to significant bias in many survey administrations. For example, if a scale is used to select individuals for a job, and if the respondents know this, impression management could easily bias test results.

Determining an acceptable level of validity is not straightforward. It requires a distinction between statistical and practical significance, and a careful evaluation of the experimental methodology used to assess validity. In some cases, this kind of evaluation requires the special training and experience of a statistical consultant who has nothing to gain if the target survey is used. A more cost-effective approach is to study the research literature associated with the survey. If the research evaluating it is published in a scientific journal, then the survey likely has passed at least one rigorous test of validity.

Table 1

Measuring an Intervention

What Is Measured?	How Is It Measured?	What Is the Score?
Injury-related incidents	•Near-hit reports •Injury reports •Workers' compensation costs	•Frequency and type of near-hits •Number and type of injury-producing incidents •Monetary expenditures
Environment	•Observation of worksite •Housekeeping audit	•Percentage of safe conditions per opportunity •Percentage of items in proper location
Behavior	•Direct observation •Corrective action survey	•Percentage of safe behaviors per opportunity •Number of items corrected for safety
Attitudes/perceptions/ person states	•Questionnaire •Interview	•"Safety score" reflecting overall safety attitude, perception or person state •Statements of specific and general attitudes about safety
Knowledge	•Questionnaire •Interview	•Percentage correct •Statements indicating awareness of a hazard or a safety procedure
Opinions	•Questionnaire	•Opinion score
Participation	•Direct observation •Attendance records	•Number of participants per opportunity

Social Validity

One final type of validity must also be assessed. It is used to evaluate the utility of an intervention program, such as a particular process to improve safety-related behavior. It was developed by researchers and practitioners in behavior-based psychology (Baer, et al), and essentially refers to the practical significance of an intervention, product or change process. To evaluate social validity, researchers use rating scales, interviews and focus-group discussions to assess: "1) The social significance of goals . . . 2) the social appropriateness of the procedures . . . [and] 3) the social importance of the effects" (Wolf 207). It is often useful to obtain social validity evaluations from actual recipients of the program or intervention.

Comprehensive evaluation of an intervention's social validity is more complex than it seems [Geller(a)]. Many perspectives are available on what makes intervention goals socially significant, procedures socially appropriate and results socially important. Furthermore, various approaches can be used to assess an intervention's social validity—from objective behavioral observations to surveys of reactions from those involved in the process.

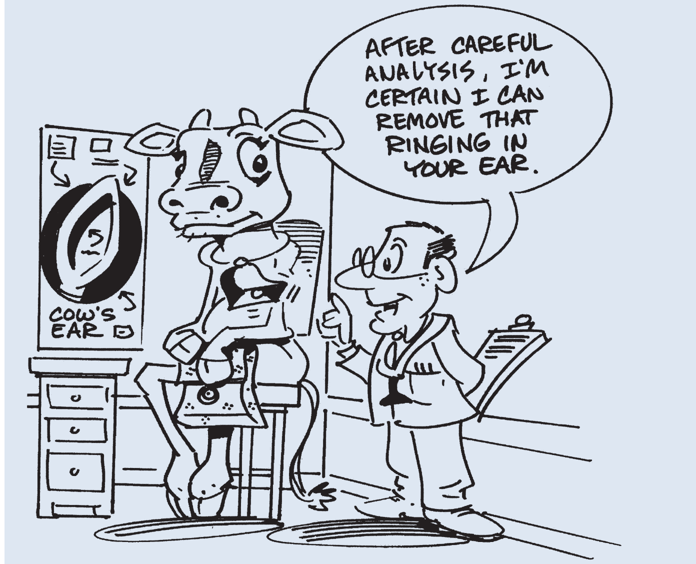
To understand various perspectives regarding social validity, the author has found it useful to consider the four basic components of an intervention process: selection, implementation, evaluation and dissemination [Geller(a)]. Selection refers to the importance or priority of the target problem and population addressed. The social validity of selecting workplace and community safety as an intervention target is obvious, given that unintentional injury is responsible for the greatest percentage of years of potential life lost before age 65 (Sleet).

Assessing the social validity of the implementation stage includes evaluating the goals and procedures of the program plan. How acceptable is the intervention to potential participants and other parties, even those with tangential connections (Schwartz and Baer)? In the case of a corporate safety program, this means obtaining acceptability ratings not only from target employees, but also from employees' family members and company customers. It is easy to obtain opinions about an intervention technique from those directly affected (participating employees) is convenient, but it is difficult (but usually not impossible) to solicit reactions from those indirectly influenced by a corporate program. Ask the question: Are the intervention procedures used to promote safety consistent with an organization's values, and do they reach the appropriate audience?

The social validity of the evaluation stage refers to the impact of the intervention process. This includes estimating an intervention's costs and benefits as well as measures of participant or consumer satisfaction. Table 1 depicts the various ways to evaluate program impact. The first column lists aspects of a work setting that can be measured before, during and after implementation of a safety intervention. The top items are directly measurable and relate most immediately to the ultimate purpose of a safety intervention—injury prevention. Therefore, improvements in injury-related incidents, behaviors and environmental conditions would indicate more social validity for the evaluation phase than would beneficial changes in attitudes, perceptions, knowledge, opinions or program participation.

Figure 3

Cause-and-Effect Relationship



The second column of Table 1 includes examples of the type of measurement tool or index that can assess the dimensions in the left column. Each measurement device can be classified according to three basic sources of data: 1) direct observation; 2) archival data (obtained from examining plant documents, memos and government reports); and 3) self-report (such as verbal answers to interview questions or written reactions on questionnaires). If the direct observations and archival data are reliable, these measures have greater social validity than self-report measures (Hawkins).

The third column reflects the scores or numbers obtained from various measurement tools. What kinds of numbers are meaningful and useful? Of course, the numbers (or assessment scores) need to be reliable and valid, and they also must be understood by those who use them. If they are not, the evaluation scheme cannot hold people accountable and lead to continuous improvement. Meaningless numbers also limit the dissemination potential and large-scale applicability of an intervention.

This review of basic concepts about surveys is offered to help SH&E professionals realize that even the best surveys only assess correlations between factors. Survey results can only estimate how one perception, attitude or opinion relates to another; they say nothing about cause-and-effect. In other words, survey results offer little about how one factor influences (or causes) another. Thus, answers on a questionnaire or from an interpersonal interview cannot define cause-and-effect relationships. This is why the "root cause" of an incident cannot be determined from survey data. Let's now consider what applications of the scientific method can show something valid about cause-and-effect relationships.

Cause-&-Effect Relationships

Researchers assume that events have causes. Events, including behaviors, do not happen independently or for no reason. In fact, the notion that every event has a specific cause is a basic assumption of science. Researchers apply the scientific method to search for the causes of events or behaviors—which is easier said than done.

Do They Vary Together?

To infer causation between a factor and an event, three criteria must be met. First, covariation must be shown; that is, the factor and the event (or behavior) must vary together. When the event (or behavior) occurs, the factor must be present. This criterion for defining cause-and-effect can be measured with a survey or by behavioral sampling. Surveys can identify correlations between two or more factors, and observations of a behavior within specific surroundings can show that a behavior and an environmental factor covary—that they occur together in time and space.

Which Came First?

Cause-and-effect relationships, however, require more than a demonstration of covariation. One must also determine the time-order relationship between two variables. In other words, to define a cause-and-effect relationship, one must know which variable, factor or event occurred first. Relationships derived from questionnaire data are rarely given a time-order attribute. Furthermore, an observation of behaviors and environmental factors does not necessarily indicate which came first. When a chain of behaviors is observed, as when a person gets in a vehicle and buckles up, a precise sequence of behaviors can be specified. In this case, two criterion for defining causal relationships are satisfied.

Is There an Alternative Explanation?

However, covariation and a time-order relationship are not sufficient criteria for defining a cause-and-effect relationship. A third, and difficult-to-meet, criterion is necessary. Specifically, no alternative explanation can be provided for the observed cause-and-effect relationship. Researchers manipulate an independent variable and look for predicted change in a dependent variable. However, even when they observe expected change in behavior or attitude following prior introduction of a particular environmental factor, a cause-and-effect relationship cannot be assumed. A cause-and-effect statement is only legitimate when all possible alternative explanations are eliminated.

Researchers eliminate alternative explanations for a cause-and-effect relationship through the design of an experiment. They might use a control group or observe the behavior of the same individuals before and after introducing an independent variable. A research design that eliminates alternative explanations for a cause-and-effect observation is considered internally valid. (Defining the variety of research designs and their concomitant internal validity is beyond the scope of this article. Interested readers may consult texts on the principles of psychological research, such as Goodwin; McBurney; and Shaughnessy and Zechmeister.)

To appreciate the challenge in satisfying this third criterion for causal inference, consider Figure 3. Does this illustration depict a cause-and-effect relationship? Actually, the humor in the illustration is based on a self-evident root cause of the ringing in the cow's ear. The bell causes the ringing, correct? But is another explanation possible? Could other noises have caused the ringing problem? Was the cow born with the problem or did it develop over time, independent of the bell around its neck? How can one test whether the bell is the cause? Suppose the bell is removed. Might the ringing continue? Perhaps the bell has caused permanent damage to the ear, and the ringing continues without the bell.

Bottom line: It is not easy to show a cause-and-effect relationship. Researchers rarely state a causal relationship between a manipulated independent variable and an observed change in a certain dependent variable. For example, researchers will not say environmental factor "A" caused the observed change in behavior or attitude "B." The most they might say is, "Factor B changed after the manipulated change in Factor A, suggesting a causal link between the two variables."

Researchers in the behavioral and social sciences are skeptical and conservative, and rarely claim to observe a cause-and-effect relationship. The author suggests that SH&E professionals assume a similar stance and search for contributing factors of an incident or injury. The analysis might be the same, but the conclusions different. Through interviews, surveys and perhaps behavioral observations, SH&E professionals should assess the variety of possible determinants of a workplace mishap, and the elimination of any one of these factors could prevent a recurrence of the undesirable incident. This change in language and perspective could help transition the incident analysis from fault-finding to fact-finding, and from asking the question, "What was the root cause that needs to be eliminated?" to "What were the contributing factors I could help remove from the workplace?"

Conclusion

W. Edwards Deming said many times, "There's no substitute for knowledge" (Deming). In this author's opinion, profound knowledge for SH&E management should come exclusively from scientific research, not from intuition or common sense. Common sense might be the basis for an original theory, but it must be modified according to the results of real scientific inquiry. Common sense can benefit from research, in that as the theory continuously improves from the deductive-inductive process of collecting, analyzing and interpreting objective data relevant for supporting, rejecting or refining a particular presumption or thesis. People with a research perspective continuously change their common sense according to empirical, objective and progressive information.

Research has shown some current common sense or theory applied to safety management is flawed. However, research can be used to improve these approaches to safety management. This requires a paradigm shift for some SH&E professionals. This needed change in perspective is contingent on embracing the basic research principles and procedures reviewed here. ■

References

Baer, D.M., et al. "Some Current Dimensions of Applied Behavior Analysis." *Journal of Applied Behavior Analysis.* 1(1968): 91-97.

Bouchard, T.S. "Genetic and Rearing Environment Influences on Adult Personality: An Analysis of Adopted Twins Reared Apart." *Journal of Personality.* 58(1996): 263-292.

Daniels, A.C. *Bringing Out the Best in People.* 2nd ed. New York: McGraw-Hill Inc., 2000.

Deming, W.E. "Quality, Productivity and Competitive Position." Workshop. Los Angeles: Quality Enhancement Seminars Inc., 1991.

Geller, E.S.(a). "Applied Behavior Analysis and Social Marketing: An Integration for Environmental Preservation." *Journal of Social Issues.* 45(1989): 17-36.

Geller, E.S.(b). *The Psychology of Safety Handbook.* Boca Raton, FL: CRC Press, 2001.

Geller, E.S., ed.(c). *Social Validity: Multiple Perspectives.* Monograph No. 5. Lawrence, KS: Society for the Experimental Analysis of Behavior Inc., 1991.

Geller, E.S., et al. *Behavior Analysis Training for Occupational Safety.* Newport, VA: Make-A-Difference Inc., 1989.

Goodwin, C.J. *Research in Psychology.* 2nd ed. New York: John Wiley & Sons Inc., 1998.

Hawkins, R.P. "Is Social Validity What We Are Interested In? Argument for a Functional Approach." *Journal of Applied Behavior Analysis.* 24(1991): 205-213.

Kazdin, A.E. *Behavior Modification in Applied Settings.* 5th ed. Pacific Grove, CA: Brooks/Cole, 1994.

McBurney, D.H. *Research Methods.* Belmont, CA: Wadsworth/Thompson Learning, 2001.

Metzgar, C.R. "Evidence-Based Loss Control." *Professional Safety.* Aug. 2002: 39-43.

Orne, M.T. "Demand Characteristics and the Concept of Quasi-Controls." In *Artifact in Behavior Research,* R. Rosenthal and R. Rosnow, eds. New York: Academic Press, 1969.

Petersen, D. *Authentic Involvement.* Itasca, IL: National Safety Council, 2001.

Roberts, B.W. "Organizational Behavior Management and Personality Psychology: Reunited and It Feels So Good?" *Journal of Organizational Behavior Management.* 22(2002): 59-70.

Schlenker, B.R. *Impression Management.* Monterey, CA: Brooks/Cole, 1989.

Schwartz, I.S. and D.M. Baer. "Social Validity Assessments: Is Current Practice State of the Art?" *Journal of Applied Behavior Analysis.* 24(1991): 189-204.

Shaughnessy, J.J. and E.B. Zechmeister. *Research Methods in Psychology.* 4th ed. New York: McGraw-Hill Inc., 1997.

Skinner, B.F. *The Behavior of Organisms.* Acton, MA: Copley Publishing Group, 1938.

Skinner, B.F. "Are Learning Theories Necessary?" *Psychological Review.* 37(1950): 193-216.

Sleet, D.A. "Motor Vehicle Trauma and Safety Belt Use in the Context of Public Health Priorities." *The Journal of Trauma.* 27(1987): 695-702.

Umstot, D.D. *Understanding Organizational Behavior.* St. Paul, MN: West Publishing Co., 1984.

Wolf, M.M. "Social Validity: The Case for Subjective Measurement or How Behavior Analysis Is Finding Its Heart." *Journal of Applied Behavior Analysis.* 11(1978): 203-214.

Your Feedback

Did you find this article interesting and useful? Circle the corresponding number on the reader service card.

RSC#	Feedback
25	Yes
26	Somewhat
27	No