

# PERCEPTION SURVEYS

## Development & Analysis to Study Safety Issues

By Christopher Janicak and Majed Zreiqat

**SAFETY PROFESSIONALS PERIODICALLY RELY ON** the use of employee perception surveys to monitor and gauge safety performance in the workplace. When appropriately developed and assessed, these tools can provide invaluable information. Because of the proliferation in use of perception surveys over the years in the safety profession, this article will address the accepted practices of perception survey development, analysis and interpretation.

### Use of Surveys in the Workplace by Safety Professionals

Surveys can play a vital role in safety program management. Perception surveys have been used to assess employee perceptions of the safety culture, safety climate, perceptions of the leading indicator effectiveness, incident risk perceptions and measurement of safety management system components. Surveys can also be a useful component of safety management systems. “Continuous improvement process as part of a safety management system relies on data collection” (Herrera, 2018). Surveys can play an integral part in this data collection. They can be used to determine employee needs and services that should be included as part of a wellness program (Rosen & Spaulding, 2009). Employee surveys are also a useful tool as part of a VPP program for obtaining opinion data pertaining to an organization’s safety culture (OSHA, 2008).

**KEY TAKEAWAYS**  
• This article provides guidelines for safety professionals to follow when developing surveys used to gather important information concerning safety issues.

• It also describes methods for analyzing and interpreting survey data.

A literature review using the Science Direct search engine identified an increase in published research studies utilizing perception surveys from three published articles in 2000 to more than 30 in 2018. Most notably, perception surveys assessing safety culture and safety climate have accounted for a large number of the published research articles during this period.

### Define What Will Be Measured

The first step in perception survey development is to clearly define what will be measured. Common uses of percep-

tion surveys in safety include safety climate, safety culture, risk perceptions and psychological constructs.

### Safety Culture Surveys

According to International Nuclear Safety Advisory Group (INSAG, 1996):

While there is no single definition of *safety culture*, in light of the Chernobyl disaster, safety culture was defined by INSAG as an organizational atmosphere where safety and health is understood to be, and is accepted as, the number one priority.

The Advisory Committee on the Safety of Nuclear Installations (ACSNI, 1993) defines the safety culture of an organization as:

The product of individual and group values, attitudes, perceptions, competencies and patterns of behavior that determine the commitment to, and the style and proficiency of, an organization’s health and safety management. Organizations with a positive safety culture are characterized by communications founded on mutual trust, by shared perceptions of the importance of safety and by confidence in the efficacy of preventive measures.

Additional example definitions of *safety culture* include:

- Safety culture incorporates the values and norms and beliefs of a particular company.
- Safety culture is a group’s initiatives, actions, exercises, processes, habits, training and education, and relationships, etc., that pool to establish the core principles and values of the group.
- Safety culture is the overall mind-set of what folks think about safety on the jobsite, that yes, we want to be a safe company.
- Safety culture is how people act when nobody’s watching.
- Safety culture is a subset of the culture of the organization. It represents not necessarily well-articulated expressions of how and why things are done within the organization.

•Safety cultures reflect the attitudes, beliefs, perceptions and values that employees share in relation to safety. (Gillen, Goldenhar, Hecker, et al., 2014)

To measure the perceptions of a safety culture, the observable components that comprise the safety culture must be identified. Singer, Gaba, Geppert, et al. (2003), defined the components of a safety culture as:

- 1) Commitment to safety articulated at the highest levels of the organization and translated into shared values, beliefs and behavioral norms at all levels.
- 2) Necessary resources, incentives and rewards provided by the organization to allow this commitment to occur.
- 3) Safety is valued as the primary priority, even at the expense of "production" or "efficiency"; personnel are rewarded for erring on the side of safety even if they turn out to be wrong.
- 4) Communication between workers and across organizational levels is frequent and candid.
- 5) Unsafe acts are rare despite high levels of production.
- 6) There is an openness about errors and problems; they are reported when they do occur.
- 7) Organizational learning is valued; the response to a problem focuses on improving system performance rather than on individual blame.

An example of a perception survey used to assess the safety culture is NIOSH's (2010) safety culture survey, which uses six questions to measure safety culture (Figure 1).

### Safety Climate Surveys

Zohar (1980) first coined the term *safety climate* as his methods to measure workers' perceptions of on-the-job safety. Zohar defined the safety climate as "the summary of molar perceptions that employees share about their work environment and found safety climate was related to safety audit scores" (Schwatka, Hecker & Goldenhar, 2016). Safety climate is a measure of perceptions about safety that are reflective of the immediate circumstances, whereas "*safety culture* tends to be a term that is generally used to describe an overall sense of shared beliefs, values and traditions around workplace safety that is viewed within the larger framework of organizational systems" (Goulart, 2013).

Additional definitions of the term *safety climate* include:

- 1) Safety climate is what happens on a day-to-day basis, sort of a snapshot of what's actually happening and how employees perceive how the company is actually implementing safety on the ground.
- 2) Safety climate is how things are being done, you know how it really is right now, and is it really being practiced? Is safety a major concern for the company, do they really care about safety or are they just talking about it?
- 3) Safety climate is more of an encouragement, enabling and giving people the tools and education. It is very much about support for the ability for people to perform their work safely.
- 4) Safety climate is the shared perceptions of the workforce at a given point in time as to the extent hazard identification and injury performance are important to the organization as perceived by em-



SDI PRODUCTIONS/E+/GETTY IMAGES

## FIGURE 1 NIOSH SAFETY CULTURE SURVEY

Please indicate how much you agree or disagree with each of the following statements about safety behavior in the organization where you work. Use this scale to answer the questions:

Strongly disagree	Disagree	Agree	Strongly agree
1	2	3	4
1	2	3	4
1	2	3	4
1	2	3	4
1	2	3	4
1	2	3	4
1	2	3	4

### Scoring

Add up your score. The range of possible scores is 6-24. If your score is:

- between 9 and 15, this indicates a poor safety culture at work;
- between 16 and 20, this indicates a fair safety culture;
- between 21 and 24, this indicates a good safety culture.

Analyses of data from 700 nurses in a large Northeastern U.S. hospital showed that groups with a good safety climate were more than four times more likely to follow safe work practices than those with fair or poor climates scores.

**FIGURE 2**  
**EXAMPLE LIKERT SCALE**  
**RESPONSE ANCHORS**

<b>Agreement</b> 1 = Strongly agree 2 = Agree 3 = Undecided 4 = Disagree 5 = Strongly disagree	<b>Level of difficulty</b> 1 = Very difficult 2 = Difficult 3 = Neutral 4 = Easy 5 = Very easy
<b>Importance</b> 1 = Very important 2 = Important 3 = Moderately important 4 = Slightly important 5 = Not important	<b>Likelihood</b> 1 = Extremely unlikely 2 = Unlikely 3 = Neutral 4 = Likely 5 = Extremely likely

**Note.** Adapted from "Likert Scale Examples for Surveys," by S. Brown, 2010, ANR Program Evaluation. Ames, IA: Iowa State University Extension.

employees' interactions with their direct supervisors. (Gillen, et al., 2014)

Many safety climate surveys have been developed.

In a review of the methods and results from a sample of industrial climate surveys, the basis of 18 scales used to assess safety climate were examined and the results suggest the most typically assessed dimensions relate to management (72% of studies), the safety system (67%) and risk (67%). In addition, themes relating to work pressure and competence appear in a third of these studies. (Flin, Mearns, O'Connor, et al., 2000)

**Risk Perception Surveys**

Another common approach to the use of surveys in the safety field involves the assessment of risk perceptions. Their use can be found in environmental exposures, ergonomic hazards and public health issues, to name a few.

Judgments about risk, otherwise known as risk perceptions, are viewed as a fundamental element of most theoretical models of health behavior and behavioral decision making, including social cognitive theory (Bandura, 2001), the health belief model (Rosenstock, 1974), the theory of reasoned action (Fishbein & Ajzen, 1975), the theory of planned behavior (Ajzen, 1985), self-regulation theory (Kanfer, 1970), and subjective culture and interpersonal relations theory (Triandis, 1977). (Institute of Medicine, 2012)

The risk assessment methods in these perception surveys can take on many forms. One example of a risk scale assessing risk perceptions related to environmental hazards asks respondents, using a rating scale, to assess the seriousness of a risk for an event and compare the risks of the topic at hand to the risks of another event such as death or injury (Harclerode, Lal, Vedwan, et al., 2016).

**Attitudinal Surveys**

A construct is defined in psychology as a label for a cluster or domain of covarying behaviors (Binning, 2018).

A construct derives its name from the fact that it is a mental construction, derived from the general scientific process: observing natural phenomena, inferring the common features of those observations, and constructing a label for the observed commonality or the underlying cause of the commonality. (Binning, 2018)

Examples of construct classes include personality traits, attitudes and abilities. An attitude is defined as a person's settled way of thinking or feeling about something. Attitudinal surveys (also known as affective surveys) can provide information on employees' perceptions (emotions, feelings, attitudes) of their workplace experiences. In these cases, the perception survey is measuring characteristics of the subjects that are not directly observable. In this case, the perception survey may be attempting to measure a construct (University of Minnesota, 2019). A review of previous research related to the safety attitudes has identified the following common dimensions to comprise the safety attitude (Wu, Yin, Wu, et al., 2017):

- 1) management safety commitment;
- 2) team safety climate;
- 3) fatalism;
- 4) work pressure;
- 5) risk awareness;
- 6) personal safety responsibility.

**Data Formats**

Data collected from survey research can fall into one of four different data format categories: nominal, interval, ratio or ordinal. Classifying the data collected for each variable into one of these categories is crucial because the statistical tests used to analyze the data are dependent on the data format. Applying a statistical test that assumes the data are ratio when they are in fact ordinal will result in invalid results and conclusions.

**Nominal Data**

Nominal data refers to count or frequency data used in tests. This is also referred to as categorical data. They provide no more information than naming the category to which a response belongs (Levin, 1999), for example, hair color (black, gray, blond) and gender (male, female).

**Interval Data**

Interval scales are a form of continuous data that have an arbitrary zero point (Hair, Black, Babin, et al., 2010). The zero has no meaning and does not mean the absence of the thing. The two most familiar interval scales are the Fahrenheit and Celsius scales.

**Ratio Scales**

Ratio scales are continuous data that have a true zero value and can thus tell us about the ratio as well as the difference between scores. In the physical world, height and weight are examples of ratio scales (Levin, 1999) where the zero means no height or no weight.

**Ordinal Data**

The ordinal scale of measurement, which often comes up in the social sciences, provides information pertaining to the rank order of a series of objects or the assignment to categories such as above average, average and below average, which imply an order relation (Levin, 1999).

## Dependent & Independent Variables

When performing statistical tests, it is also important to understand the relationship between the variables in terms of which variable(s) is influencing the other variable(s). Understanding this concept aids the researcher in the data analysis and interpretation of the results.

An independent variable is a variable that is being manipulated in the study to investigate its effect or impact. Examples of independent variables are those that measure characteristics of the study participants, such as gender and age. Independent variables can also include treatments that the participants are being subjected to, such as training type.

A dependent variable is the variable that is being influenced by the independent variable. It is the variable the researcher expects to see differences or changes in as a result of the independent variable. For example, in a survey study, the researcher wanted to see if a relationship exists between the subjects' seniority measured as the number of years on the job and their knowledge of the safe procedures as measured by quiz scores. In this case, the obtained quiz score is dependent upon the person's age, therefore the dependent variable is the quiz score and the independent variable is the person's seniority.

## Developing Likert Items & Scales

The widely used Likert scales were developed by the American sociologist Rensis Likert (1903-1981). A distinction must be made first between a Likert scale and a Likert item.

A Likert scale is the sum of responses on several Likert items, whereas a Likert item is a statement that the respondent is asked to evaluate according to (subjective or objective) criteria—usually, five levels of agreement or disagreement with the statement, where 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree. (Dimitrov, 2011)

## Formatting Likert Items

Wilson (2005) explains:

A general approach to the creation of outcome spaces in areas such as attitude and behavior surveys has been the Likert style of item. . . . The most generic form of this is the provision of a stimulus statement (sometimes called a stem) and a set of standard options among which the respondent must choose.

The scale portion of a Likert scale item can take a number of different formats including four-, five- and seven-point scales. Each scale has its own benefits and drawbacks, and much research has been conducted on the impact of using even- versus odd-numbered scales.

A neutral midpoint on odd-numbered Likert scales provides respondents the opportunity to provide a neutral response. According to Garland (1991):

This research provides some evidence that social desirability bias, arising from respondents' desires to please the interviewer or appear helpful or not be seen to give what they perceive to be a socially unacceptable answer, can be minimized by eliminating the midpoint ("neither . . . nor," uncertain, etc.) category from Likert scales. . . . There is also some evidence that the presence or absence of a

midpoint on an importance scale produces distortions in the results obtained. Arguably, the problem still remains to be resolved: should the market researcher offer respondents the chance to express a truly neutral position?

There has been much debate as to whether Likert scales should be treated as ordinal data or interval data.

Dr. Geoff Norman, one of world's leaders in medical education research methodology, has comprehensively reviewed this controversy. He provides compelling evidence, with actual examples using real and simulated data, that parametric tests not only can be used with ordinal data, such as data from Likert scales, but also parametric tests are generally more robust than nonparametric tests. That is, parametric tests tend to give "the right answer" even when statistical assumptions—such as a normal distribution of data—are violated, even to an extreme degree. (Sullivan & Artino, 2013)

Thus, "parametric tests are sufficiently robust to yield largely unbiased answers that are acceptably close to 'the truth' when analyzing Likert scale responses" (Sullivan & Artino, 2013).

## Typical Likert Scale Response Anchors

The end points of the scale should be the exact opposites. If the scale has an odd number of choices, the center point on the scale should be neutral. The scales makers should correspond to the item question. For example, if the question is asking about the degree to which the respondent agrees to a statement, then the labels on the scale should be worded in terms of agreement. Several examples are available of typical Likert scale stem and response anchors based on a four-, five- or seven-point scale (Brown, 2010). When assembling a questionnaire using a series of scale items, a mixture of positively and negatively worded items should be used (Harvard University, 2007). Some examples of Likert scales appear in Figure 2.

## Constructing a Questionnaire

Designing a questionnaire takes planning. The process of operationalizing a questionnaire is to take a general purpose or set of purposes and turn these into concrete, researchable fields about which actual data can be gathered (Cohen, Manion, & Morrison, 2003). This process can be summarized in three steps: 1) defining the general purpose of the questionnaire; 2) identifying subsidiary topics; and 3) formulating specific information requirements (Cohen, Manion & Morrison).

## Instrument Validity

Instrument validity is a process of determining whether a survey instrument, in this case, is measuring what it is intended to measure. Validity can be measured in several different ways. Examples of validity can include (Ruane, 2015):

- construct validity demonstrating the accuracy of a measure by showing it produces results consistent with theoretically based hypotheses or predictions;
- content validity asserting a measure is accurate because it addresses all dimensions or components of the concept's nominal/theoretical definition;
- face validity claiming a measure is accurate because it "looks" right or appears to be getting the job done, a subjective assessment of validity;

•criterion validity using empirical evidence to establish that a measure is measuring what it claims to measure.

### Instrument Reliability

Reliability of a survey instrument is a measure of its consistency. For example, if a person completes a survey, one would expect, provided there are no interventions, the person to obtain similar or consistent results. Reliability can be assessed in several different ways. Examples include test-retest reliability, split-half reliability and alternate forms reliability. Statistical tests can be used to quantify the reliability of an instrument such as Cronbach’s alpha and Pearson’s correlation coefficient.

### Split-Half Reliability

Split-half reliability is a measure on internal consistency. Ruane (2015) defines split-half reliability check as “assessing the consistency or stability of a composite measure by splitting the measure into two halves and checking to see if the results yielded by each half are highly correlated with each other.” To conduct a split-half reliability analysis, the instrument is administered to a large group of subjects (more than 30). The items are randomly divided into two parts and a Pearson correlation coefficient can be used to correlate the score for one-half of the items to the score for the second half. The following Spearman-Brown formula is tailored to the split-half approach (Salkind, 2010):

$$\text{Split-half estimate} = \frac{2r_{SH}}{1 + r_{SH}}$$

where  $r_{SH}$  is the Spearman-Brown split-half correlation. For example, a split-half correlation of .57 will yield a reliability of .73:

$$\text{Split-half estimate} = \frac{2(.57)}{1 + .57} = \frac{1.14}{1.57} = .73$$

Typically, a correlation coefficient value of .70 or higher is considered as a good indication of reliability (Ruane, 2015).

**TABLE 1**  
**NUMBER OF MISHAP SITUATIONS REPORTED BY EMPLOYEE**

	Attended training	Did not attend training
	3	4
	2	6
	4	4
	5	3
	4	6
	3	5
	2	4
	3	3
	4	5
	5	7
<b>Average</b>	<b>3.5</b>	<b>4.7</b>

### Test-Retest Reliability

Test-retest reliability “[assesses] the consistency of a measure by employing it two times on the subjects and looking for highly correlated results (provided there has been no real change in the subjects between the test and retest)” (Ruane, 2015). As with split-half reliability, a Pearson correlation coefficient is a common statistical test used for assessing test-retest reliability.

### Analyzing Data Using Excel

The data analysis tools in Excel can be used to analyze data to test various statistical hypotheses. A few examples are presented including *t*-tests, chi-square, Pearson correlation, and phi coefficient. For more complex research questions, statistical software such as SPSS, SAS and R language are utilized.

### Basics of Hypothesis Testing

Inferential statistical tests and procedures have hypotheses and assumptions. It is imperative the researcher knows what these are before collecting data and attempting to analyze the results. Failure to meet the assumptions will result in invalid results and conclusions. Improperly worded hypotheses or hypotheses not matched with the correct statistical test will also result in invalid conclusions. An inferential statistical test has a pair of hypotheses associated with it: the null hypothesis and the alternative hypothesis.

An inferential statistical test has an associated set of null and alternative hypotheses. The test could examine differences in average, correlations between two variables, significant differences in proportions, significant differences in the frequency of cases and more. The researchers must formulate their research question and match the question to appropriate statistical test and statistical hypotheses.

### Determining Significance

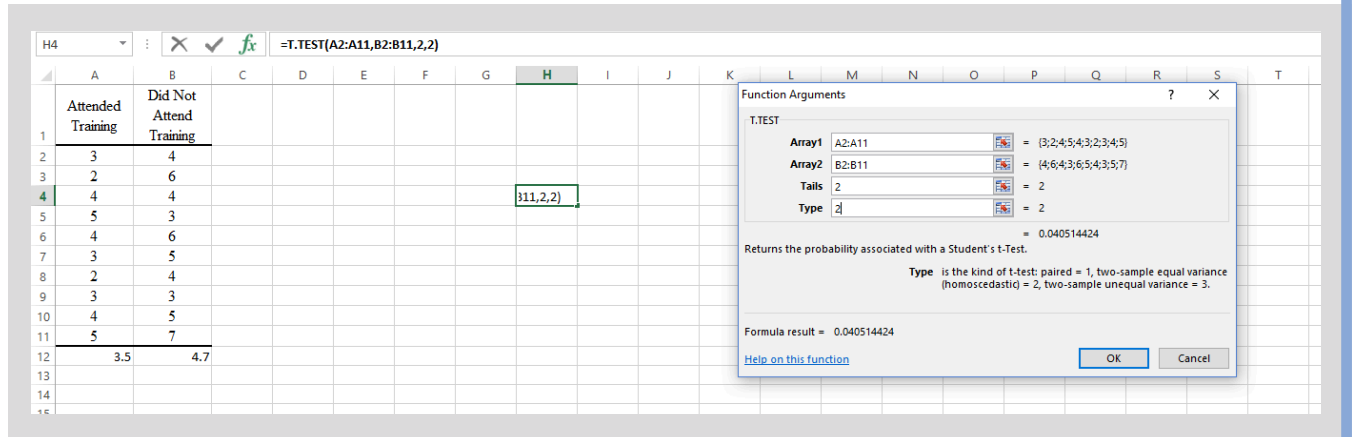
The researcher must first determine what will signify a significant result. Significance is represented by the alpha level ( $\alpha$ ). The alpha level is interpreted as the probability the null hypothesis will be rejected when it is in fact false. An alpha level of .05 is common in social science research. In sensitive research, such as medical application, where an incorrect decision could be harmful as in the case of testing the effectiveness of a drug treatment, then alpha levels smaller than .05, such as .01 and .001, are appropriate.

### Independent Samples *t*-Test

The independent samples *t*-test is used to determine whether a significant difference exists between two averages. The independent samples *t*-test assumes the populations from which the study samples are selected are normally distributed. Another assumption is that the variances are equal in the populations from which the samples are selected (Rosenthal, 2012). When sample sizes are equal or nearly equal, the violation of these assumptions has a minimal impact on the results. However, if the sample sizes differ greatly, the results of the test can be inaccurate (Rosenthal). If the sample sizes are unequal or the researcher suspects the variances of the groups differ significantly, then the appropriate test for unequal variances should be used.

In this example, a safety professional collected survey data and would like to determine whether there are significant differences in the average number of near-hit errors reported on a survey for 10 employees who attended the safety training

**FIGURE 3**  
**EXCEL ENTRIES FOR AN INDEPENDENT SAMPLES T-TEST**



compared to 10 employees who did not attend the training. The sample data collected appears in Table 1.

### Selecting a Hypothesis to Test

The independent samples *t*-test can test the following hypothesis sets:

#### Set 1

Null: There is no significant difference in the averages when comparing Group A to Group B.

Alternative: There is a significant difference in the averages when comparing Group A to Group B.

#### Set 2

Null: Group A average is significantly greater than or equal to Group B average.

Alternative: Group A average is significantly less than Group B average.

#### Set 3

Null: Group A average is significantly less than or equal to Group B average.

Alternative: Group A average is significantly greater than Group B average.

The difference between the three sets is that set 1 is a two-tail test where alpha level is divided equally by two, while set 2 and 3 are one-tail test.

### Steps to Conducting the Analysis in Excel

The following steps will complete this analysis. First, the researcher must select one of the preceding hypothesis sets and input the data into an Excel spreadsheet. Next, in Excel, select "Formulas," "More functions," "Statistical." From the list, select "T.TEST." In the dialog box highlight the 10 numbers in the "Attended training" column for array 1, and the 10 numbers in the "Did not attend training" for array 2. (Do not include the values for the averages in the arrays. These will be used to explain where the averages differ in the next phase of the analysis.) In the "Tails" box, select two tails. Because the alternative hypothesis is testing the means are not equal, then the test is a two-tailed test. If the alternative hypothesis is testing a mean is greater than or less than the other, then it is a one-tailed test. In this example, we will be testing the hypothesis in which the averages are significantly different (equal vs. not equal), which is a two-tailed test.

**TABLE 2**  
**OBSERVED & EXPECTED CASES**

Observed cases				
	High	Medium	Low	Total
Trained	5	9	12	26
Untrained	15	12	5	32
Total	20	21	17	58

Expected cases				
	High	Medium	Low	Total
Trained	9.12	9.58	7.75	26
Untrained	11.23	11.79	9.54	32
Total	20	21	17	58

For the type of test, if it is a paired *t*-test, enter 1, for an independent samples *t*-test with equal variances, enter a 2, and for an independent samples *t*-test with unequal variances, enter a 3. In this example, we are assuming equal variances, so we will enter a 2. A Levene's test can be used to test the equality of the variances and is available in statistical packages that are more advanced than Excel. Online calculators can also be used to run a Levene's test. After you have all of the entries into Excel completed, press Enter and the result is the p value for the *t*-test. In this example, the result is .0405. See Figure 3 for the Excel entries.

### Interpreting the Results

We set our alpha level to .05, so an answer less than .05 means the results are statistically significant. Our result is .0405, which is less than .05. Therefore, we reject our null hypothesis and conclude that the averages are significantly different. Looking at the averages, the group that participated in the training had an average number of errors of 3.5, whereas the group that did not attend the training had an average of 4.7. This can be used to infer that training was effective in reducing the number of errors.

## Chi-Squared Test of Association

The chi-squared test of association is one of several chi-squared tests available. The chi-squared test of association looks for a relationship between two categorical variables in which two or more categories are possible for each variable. The data are presented as frequency data in a table of observed cases that are then compared to a statistically derived number of expected cases. In addition to the requirement of having two categorical variables, each with two or more categories, an additional assumption of the test is that at least 80% of the cells in the table should have a calculated expected frequency greater than 5 (Peacock & Peacock, 2011).

### Selecting the Hypothesis to Test

In this example, frequency data were collected using a survey in which the respondents answered a question as to whether they completed an online safety training course on hazard identification, then how many hazards they correctly identified when presented with a series of workplace hazards videos. The researcher wishes to determine whether a relationship exists between the correct number of hazards identified and attending the online hazard identification training course.

The independent variable in this example is attending or not attending the training program. The dependent variable is the correct number of hazards identified categorized as high (more than 10 hazards identified correctly), medium (5 to 10

hazards identified correctly), and low (fewer than 5 hazards identified correctly). The null hypothesis for the chi-squared test of association is there is no significant relationship between the dependent variable and the independent variable. The alternative hypothesis is there is a significant relationship between the two variables.

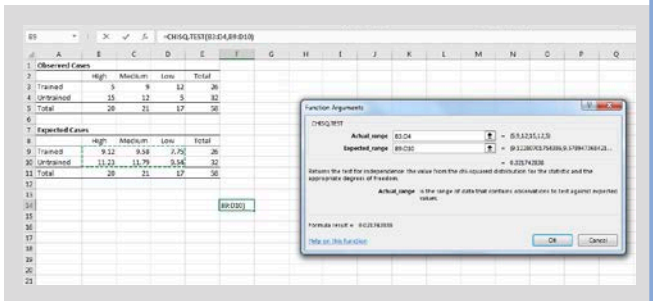
### Steps to Conducting the Analysis in Excel

The first step in conducting a chi-squared test is to set up the data sheet with the observed number of cases and the expected number of cases. The expected number of cases in each cell is calculated by multiplying the row total for that call by its column total then dividing by the grand total.

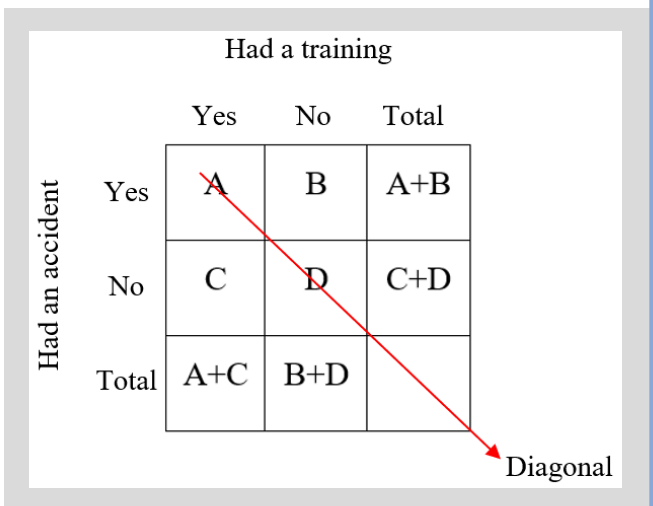
Using the cell for trained and high as an example, the observed number of cases of 5 has a row total of 26 and a column total of 20, as shown in Table 2 (p. 37). Thus, multiplying 26 by 20, then dividing by 58 yields an expected value of 9.12. This is performed for each of the six inside table cells. If the calculations are correct, adding the expected row totals and expected column totals will yield the same results as the observed cases table row, column and grand total (Table 2).

With these tables set up, one can calculate the chi-squared test. The researcher must select "Formulas," "More functions," "Statistical." From the list, select "CHISQ.TEST." In the dialog box, select the six inner table cells from the observed cases table for the "Actual range" in the dialog box. For the "Expected

**FIGURE 4**  
**EXCEL ENTRIES FOR CHI-SQUARED TEST**



**FIGURE 5**  
**EXAMPLE 2 × 2 CONTINGENCY TABLE**



**TABLE 3**  
**STRENGTH OF LINEAR RELATIONSHIP**

Correlation coefficient	Strength of linear relationship
At least .8	Very strong
.6 up to .8	Moderately strong
.3 to .5	Fair
Less than .3	weak

Note. This table also applies to the negative correlation. Adapted from "Biostatistics 104: Correlational Analysis," by Y.H. Chan, 2003, *Singapore Medical Journal*, 44(12), 614-619.

**TABLE 4**  
**SUMMARY DATA**

Subject	Attended training	Had an accident
1	Yes	No
2	Yes	No
3	No	Yes
4	No	Yes
5	No	Yes
6	Yes	No
7	Yes	No
8	Yes	No
9	Yes	No
10	Yes	No

range,” select the six inner table cells from the expected cases table. Press “OK” and result will be the probability of the chi-squared test (Figure 4).

### Interpreting the Results

As with the other tests, we set our alpha level to .05, so an answer less than .05 means the results are statistically significant. Our result is .0217, which is less than .05. Therefore, we reject our null hypothesis and conclude that the frequency of cases is significantly different than what is expected due to chance. Comparing the frequencies in the expected cells to the actual observed cases, the number of cases in the untrained cells is lower than the expected number of cases across all groups, whereas the number of cases in the trained cells is higher than the expected number of cases across all groups.

### Phi Correlation

The phi correlation ( $\phi$ ) is a nonparametric test used to quantify the degree of association (linearity) between two nominal variables (categorical, not continuous; e.g., training type, hair color); each has been measured on a dichotomous scale (binary; e.g., trained/not trained, incident/no incident, cancer/no cancer; Field, 2013). Data are typically presented in the form of two by two (2 x 2) contingency tables (Figure 5). The only assumption for the phi correlation is both variables are dichotomous.

Correlation is quantified by the correlation coefficient, which represents the strength of linear relationship between the investigated variables (Mukaka, 2012). The correlation coefficient, in this case the phi correlation coefficient, has a value that ranges from -1 to +1. A value of  $\pm 1$  means perfect correlation (the two variables are linear), while a value of zero means no correlation (two variables are not linear) (Field, 2013). A positive correlation indicates most of the data are in the diagonal cells in the 2 x 2 contingency table (when one variable increases, the other will also increase; Figure 5). However, negative correlation means that most of the data are off the diagonal cells in the 2 x 2 table (when one variable increases, the other will decrease). In other words, we are fitting a linear model to the data and determining how good the model fitting is (Field).

A guideline on the strength of the linear relationship corresponding to the correlation coefficient value is provided in Table 3 (Chan, 2003).

The phi correlation test is a symmetrical statistic, meaning the independent and dependent variables are interchangeable. To illustrate, consider an example in which a safety professional wants to determine the degree of association, if any, between employees’ training and incident status. The safety professional collected survey data that contained these two questions of many: Did you have a safety training? (Yes/No); and Did you have an accident? (Yes/No). Table 4 shows the summary data.

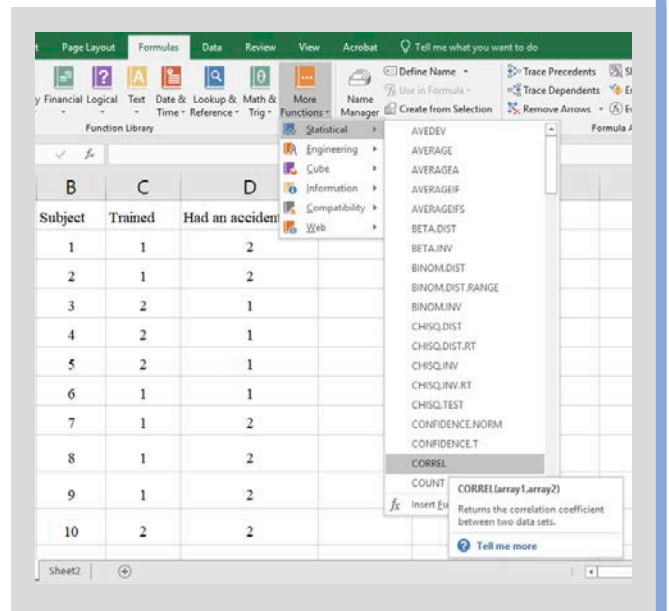
Any one of these variables can be the independent or the dependent variable.

### Steps to Conducting the Analysis in Excel

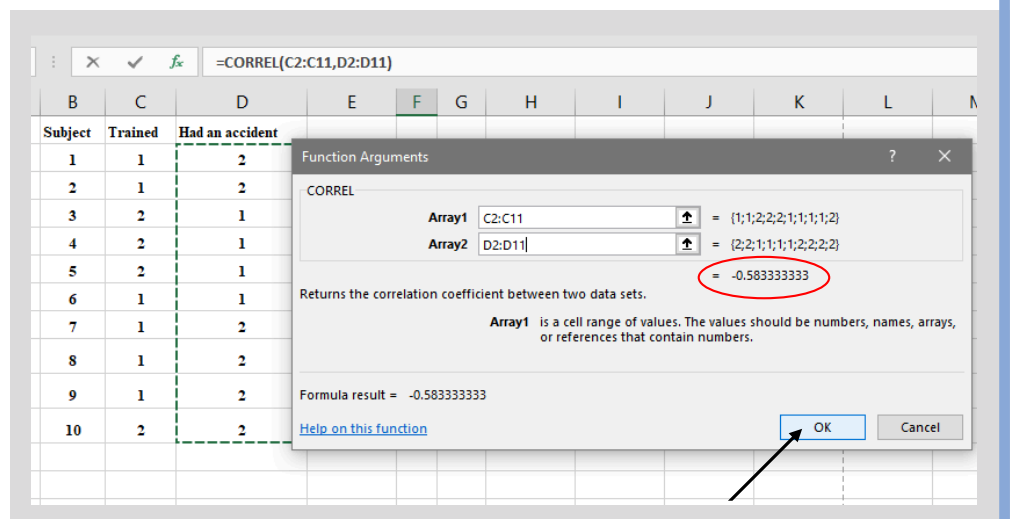
The following steps will complete this analysis. First, in Excel, select “Formulas,” “More functions,” “Statistical.” From the dropdown list, select “CORREL.” Enter the two variables data by highlighting the 10 cells in the Trained column for array 1 and the 10 cells in the “Had an accident” for array 2. You could switch the order of the arrays entries and it will not matter (Figure 6 and 7).

Lastly, click “OK” to get the phi correlation value. In this case,  $\phi = -.583$  (circled in red in Figure 7). This is a moderate

**FIGURE 6**  
**STATISTICAL TEST SELECTION**



**FIGURE 7**  
**DATA ENTRY IN EXCEL**





**TABLE 5**  
**CHI-SQUARED TABLE**

Degrees of Freedom	Probability of a larger value of $\chi^2$								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09

**TABLE 6**  
**SAMPLE CORRELATION DATA**

Employee	Years of experience (X)	No. of lost workdays (Y)
1	7	20
2	12	22
3	15	18
4	25	19
5	33	14
6	34	9
7	36	7
8	39	4
9	41	2
10	44	1

negative correlation between training and accident. This means trained employees are more likely to have fewer accidents than untrained employees.

**Testing the Significance of the Phi Correlation**

The phi correlation can test the following hypothesis:

Null: There is no significant association between the two variables ( $\phi = 0$ ).

Alternative: There is a significant association between the two variables ( $\phi \neq 0$ ).

To test the significance of the phi correlation, the chi-squared ( $\chi^2$ ) test should be used. To calculate the  $\chi^2$  value, use this equation:

$$\chi^2 = n\phi^2$$

where  $n$  = sample size or number of cases

We have 10 cases, so  $n = 10$  and therefore:

$$\chi^2 = 10(-0.583)^2 = 3.40$$

The degree of freedom (df) for the phi correlation = (no. of columns - 1)(no. of rows - 1)

The number of columns and rows are a function of the 2 x 2 contingency table, which has two columns and two rows. Accordingly:

$$df = (2 - 1)(2 - 1) = 1$$

Using the chi-squared Table 5 at an alpha ( $\alpha$ ) level of .05 and  $df = 1$ , the critical  $\chi^2 = 3.84$ .

**Interpreting the Results**

If the calculated  $\chi^2$  value is greater than the  $\chi^2$  critical value, reject the null and accept the alternative. In this case, since the calculated  $\chi^2$  of 3.40 is less than the critical  $\chi^2$  of 3.84, then we

fail to reject the null hypothesis and consequently the correlation is not statistically significant.

**Pearson Correlation**

Pearson’s  $r$  correlation is used to quantify the strength and direction of linear relationships between two continuous variables, such as age and years of experience or incident rate. Like the phi correlation, the Pearson correlation has a value between -1 and +1. A value of zero means no association, while  $\pm 1$  value is a perfect correlation (linearity), which is a rare situation to exist. A guideline on the strength of the linear relationship corresponding to the correlation coefficient value is provided in Table 3 (p. 38).

Assumptions to be met in the Pearson correlation are (Field, 2013):

- 1) Both variables are continuous (ratio or interval).
- 2) Linearity: A straight-line relationship between the two variables is assumed. Use the scatter plot to investigate this.
- 3) Homoscedasticity: The data is assumed to be equally distributed about the regression line. A scatter plot can be used to test this assumption.
- 4) Independent observations: This is usually met by data randomization.
- 5) Normality: Both variables are bivariate normally distributed. This is only an issue for small sample size ( $N < 20$ ). Normality could be evaluated using histogram or P-P plots.
- 6) No outliers: No major outliers should exist. Again, this is a major issue for small sample size. Outliers could be identified by scatter plot or P-P plot.

Let’s consider and analyze an example in which a researcher wants to determine whether an association exists between the annual number of lost workdays and the employees’ years of experience (Table 6). The number of lost workdays is ratio (dependent variable) and the years of experience is also ratio (independent variable). The independent variable will be represented by X and the dependent by Y.

**Steps to Conducting the Analysis in Excel**

The following steps will complete this analysis. First, in Excel, select “Formulas,” “More functions,” “Statistical.” From the dropdown list, select “CORREL.” After that, select the two variables by highlighting the 10 cells in the X column for array 1 and the 10 cells in the Y column for array 2 (Figure 8 and 9).

Accordingly,  $r = -.924$  (circled in red in Figure 9). This is a strong negative correlation between lost workdays and years of experience. This means that employees with more years of experience are more likely to have fewer lost workdays than employees with fewer years of experience.

**Testing the Significance of the Pearson Correlation**

The Pearson correlation can test the following hypothesis:

Null: There is no significant association between the two variables ( $r = 0$ ).

Alternative: There is a significant association between the two variables ( $r \neq 0$ ).

To test the significance of the Pearson correlation, the  $t$ -test should be used. To calculate the  $t$  value, use this equation:

$$t = \frac{r}{\sqrt{(1-r^2)/(n-2)}}$$

$$df = n - 2$$

$n =$  number of cases or data pairs  $= 10$

$$df = 10 - 2 = 8$$

$$t = \frac{.924}{\sqrt{(1 - .924^2)/(10 - 2)}}$$

$t = 6.83$ ; this is called the  $t$ -calculated value

To make a decision, compare the  $t$ -calculated value to the  $t$ -critical value from the  $t$ -test at  $df = 8$  and  $\alpha = .05$  (Table 7; highlighted in red).

### Interpreting the Results

If the calculated  $t$ -value is greater than the  $t$ -critical value, reject the null and accept the alternative. In this case, since the  $t$ -calculated value of 6.83 is greater than the  $t$ -critical value of 2.31 (a two-tail test), then we reject the null hypothesis and accept the alternative hypothesis. We conclude that there is a statistically significant negative correlation between lost workdays and work experience.

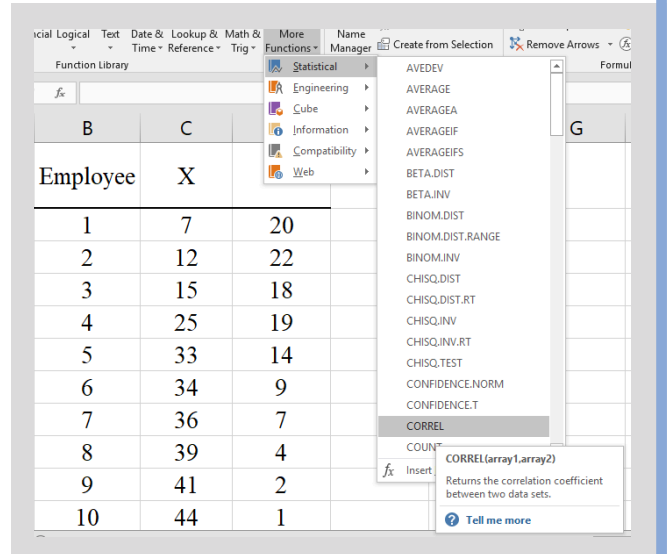
### Conclusion

This article addresses the application of the research process to survey analyses to answer common types of safety-related questions. The research process consists of:

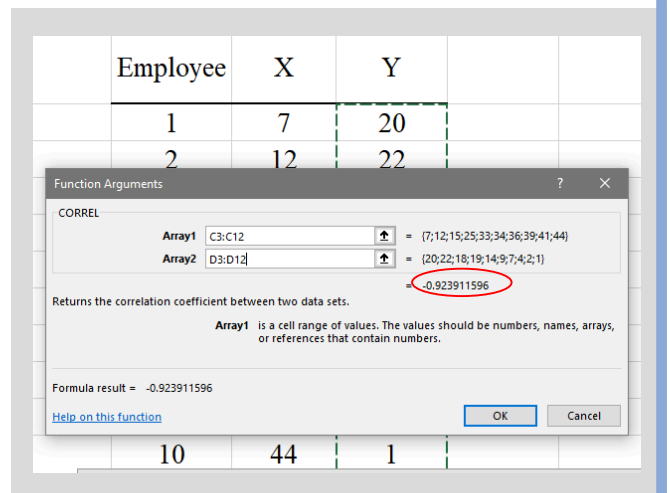
- 1) formulating a series of research questions;
- 2) designing appropriate survey questions used to gather data;
- 3) matching both the research question and the data format to an appropriate statistical test;
- 4) collecting and analyzing the data;
- 5) drawing conclusions and making recommendations for improvement.

The process begins with the development of questions that can be answered through the collection of survey data. Common types of questions include determining employees' perceptions about safe working conditions, the degree of knowledge related to workplace policies and procedures, and measures of the safety climate and safety culture. The compilation of survey items is referred to as the survey instrument. Examples of survey instruments include risk perception surveys, climate surveys and Likert scales. Each instrument example serves a unique purpose and can be tailored to specific workplace needs.

**FIGURE 8**  
**STATISTICAL TEST SELECTION**



**FIGURE 9**  
**DATA ENTRY IN EXCEL**



**TABLE 7**  
**T-TEST TABLE**

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587

The item formats contained in survey instruments and scales play a crucial role in determining the types and questions that can be answered and the appropriate statistical tests for data analysis. For any type of inferential statistical test, such as a correlation procedure, there is a set of statistical hypotheses and assumptions. The hypotheses aid in formulating the research questions to be asked and the assumptions aid in determining the formats of the survey items and characteristics of the data.

This article provides information on a few common basic statistical tests used to analyze safety performance-related data using Excel. More advanced statistical tests require specialized software and a strong foundation in mathematics and statistics.

A purpose of using survey research in the safety field is to identify areas for change and improving safety performance. The survey research process, when used properly, can assist in identifying significant relationships between the safety performance and indicators such as employee perceptions and safety program activities. Inferential statistics are important because when decisions are made about the findings, they are done so with a degree of certainty thereby reducing the chances of implementing changes that may not result in the desired effects. In summary, survey research can be a useful strategy for safety professionals to collect data and use it to make meaningful decisions. **PSJ**

## References

- Advisory Committee on the Safety of Nuclear Installations (ACSNI). (1993). *ACSNI study group on human factors. Third report: Organizing for safety*. London, U.K.: HM Stationery Office.
- Binning, J.F. (2018). Construct. *Encyclopaedia Britannica*. Retrieved from [www.britannica.com/science/construct](http://www.britannica.com/science/construct)
- Brown, S. (2010). *Likert scale examples for surveys*. ANR Program Evaluation. Ames, IA: Iowa State University Extension.
- Chan, Y.H. (2003). Biostatistics 104: Correlational analysis. *Singapore Medical Journal*, 44(12), 614-619.
- Cohen, L., Manion, L. & Morrison, K. (2003). *Research methods in education* (5th ed.). London, U.K.: Routledge.
- Dimitrov, D.M. (2011). *Statistical methods for validation of assessment scale data in counseling and related fields*. Alexandria, VA: American Counseling Association.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: Sage Publications.
- Flin, R., Mearns, K., O'Connor, P., et al. (2000). Measuring safety climate: Identifying the common features. *Safety Science*, 34, 177-192.
- Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, 1991(2), 66-70, Research Note 3.
- Gillen, M., Goldenhar, L.M., Hecker, S., et al. (2014). *Safety culture and climate in construction: Bridging the gap between research and practice* (Workshop report). Silver Spring, MD: CPWR—The Center for Construction Research and Training.
- Goulart, C. (2013, Nov. 8). Resolving the safety culture/safety climate debate. *Occupational Health and Safety*. Retrieved from <https://ohsonline.com/Blogs/The-OHS-Wire/2013/11/Resolving-the-Safety-Culture-Safety-Climate-Debate.aspx>
- Hair, J.F., Black, W.C., Babin, B.J., et al. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Harclerode, M.A., Lal, P., Vedwan, N., et al. (2016). Evaluation of the role of risk perception in stakeholder engagement to prevent lead exposure in an urban setting. *Journal of Environmental Management*, 184(1), 132-142.
- Harvard University. (2007). Harvard University Program on Survey Research: Tip sheet on question wording. Retrieved from [https://psr.iq.harvard.edu/files/psr/files/PSRQuestionnaireTipSheet\\_0.pdf](https://psr.iq.harvard.edu/files/psr/files/PSRQuestionnaireTipSheet_0.pdf)
- Herrera, G. (2018, July 1). Three questions to ensure continuous improvement of your incident reduction strategy. *Occupational Health and Safety*. Retrieved from <https://ohsonline.com/Articles/2018/07/01/Three-Questions-to-Ensure-Continuous-Improvement.aspx>
- Institute of Medicine (2012). *Scientific standards for studies on modified risk tobacco products*. Washington, DC: National Academies Press. doi:10.17226/13294
- International Nuclear Safety Advisory Group (INSAG). (1996). *Defense in depth in nuclear safety* (INSAG-10). Vienna, Austria: International Atomic Energy Agency.
- Levin, I.P. (1999). *Relating statistics and experimental design: An introduction*. Thousand Oaks, CA: SAGE Publications Inc.
- Mukaka, M.M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69-71.
- NIOSH. (2010). Stop Sticks campaign, Safety culture: Evaluation survey. Retrieved from [www.cdc.gov/nora/councils/hcsa/stopsticks/survey.html](http://www.cdc.gov/nora/councils/hcsa/stopsticks/survey.html)
- OSHA. (2008). Participation in VPP helps L'Oreal reduce injuries and illnesses, increase employee involvement and enhance a safety and health culture. Retrieved from [www.osha.gov/dcsps/success\\_stories/vpp/reg5\\_ss\\_loreal.html](http://www.osha.gov/dcsps/success_stories/vpp/reg5_ss_loreal.html)
- Peacock, J.L. & Peacock, P.J. (2011). *Oxford handbook of medical statistics*. Oxford, U.K.: Oxford University Press.
- Rosen, M. & Spaulding, T. (2009, July 1). Best practices for wellness programs. *Occupational Health and Safety*. Retrieved from <https://ohsonline.com/articles/2009/07/01/best-practices-for-wellness-programs.aspx>
- Rosenthal, J.A. (2012). *Statistics and data interpretation for social work*. New York, NY: Springer Publishing.
- Ruane, J.M. (2015). *Introducing social research methods: Essentials for getting the edge*. Hoboken, NJ: Wiley-Blackwell.
- Salkind, N. (2010). Split-half reliability. *Encyclopedia of Research Design*. Thousand Oaks, CA: Sage Publishing. doi:10.4135/9781412961288
- Schwatka, N.V., Hecker, S. & Goldenhar, L.M. (2016). Defining and measuring safety climate: A review of the construction industry literature. *Annals of Occupational Hygiene*, 60(5), 537-550.
- Singer, S.J., Gaba, D.M., Geppert, J.J., et al. (2003). The culture of safety: Results of an organization-wide survey in 15 California hospitals. *Quality and Safety in Health Care*, 12(2), 112-118.
- Sullivan, G.M. & Artino, A.R., Jr. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, 5(4), 541-542. doi:10.4300/JGME-5-4-18
- University of Minnesota. (2019). Understanding psychological measurement. Research methods in psychology. Retrieved from <http://open.lib.umn.edu/psychologyresearchmethods/chapter/5-1-understanding-psychological-measurement>
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Wu, X., Yin, W., Wu, C., et al. (2017). Development and validation of a safety attitude scale for coal miners in China. *Sustainability*, 9(12), 2165. doi:10.3390/su9122165
- Zohar, D. (1980). Safety climate in industrial organizations: Theoretical and applied implications. *Journal of Applied Psychology*, 65(1), 96-102.

**Christopher A. Janicak, Ph.D., CSP, CEA, ARM**, is a professor of safety sciences at Indiana University of Pennsylvania (IUP), Department of Safety Sciences. He earned a Ph.D. in Research Methodology from Loyola University, an M.S. in Industrial Technology/Industrial Safety Concentration from Illinois State University, and a B.S. in Health and Safety Studies from University of Illinois at Urbana-Champaign. Janicak has more than 32 years' professional experience in safety, has presented at numerous conferences and has published books in statistics, safety metrics and fire protection. He has authored more than 30 journal articles on occupational safety and workplace fatalities. Janicak is a professional member of ASSP's Western Pennsylvania Chapter and Chair of the Society's Technical Publications Advisory Committee.

**Majed Zreiqat** is an associate professor of safety sciences at IUP. He earned a Ph.D. in Occupational Safety and Health from West Virginia University. He holds an M.S. in Safety Sciences from IUP and a B.S. in Chemical Engineering from University of Jordan. His areas of expertise are ergonomics and human factors, and his research interest is in work-related musculoskeletal disorders. Prior to the safety sciences field, he was a chemical and process engineer in the manufacturing industry for 10 years. Zreiqat is a professional member of ASSP's Western Pennsylvania Chapter and the Indiana University of Pennsylvania Student Section Faculty Advisor.