

# [The internal and external validity and reliability psychology essay](https://assignbuster.com/the-internal-and-external-validity-and-reliability-psychology-essay/)

This chapter contains a description of the research methodology proposed for this study, including a restatement of the purpose and research questions. The chapter is organized with categories pertaining to the instrumentation, data collection, research methodology, ethical concerns, and explanation of the statistical procedures used for a descriptive quantitative analysis. The purpose of this study is to examine the opinions and attitudes of the Colorado Mental Health Institutes’ clinical staff regarding the perceived value of the Crisis Prevention Institutes’ Nonviolent Crisis Intervention Program used to manage violence in the hospital’s inpatient treatment settings. This will be accomplished through the use of a survey designed to assess the perceptions of those charged with providing direct care to the mentally ill population of the Institute regarding the effectiveness of the Crisis Prevention Institutes’ Nonviolent Crisis Intervention Program to prevent and manage violent behavior. Designing an attitudinal survey in order to quantify results can be a challenging task depending on what is being measured and selecting the appropriate assessment methodology. Most forms of assessment test large sample groups, with divergent sub-groups to contrast and compare in terms of scores (Thorndike & Thorndike-Christ, 2009). The subject areas in the survey that will be used for this assessment are of a non-comparative nature, meaning that each item is scaled independently of the others as opposed to comparing them to each other. The numbers used in the scale will indicate the relative position of items, but not the magnitude of difference between them. With respect to these parameters and limitations, this study will present the survey scores of the participants, along with calculations as to their range and central tendency. In addition, inferences will be made as to factors that may have influenced the scoring choices made by the respondents.

In the fields related to education and social science, research exploring the attitudes, feelings, and lived experiences of the respondents is often conducted through the use of a survey methodology. Research data gathered in this manner can be categorized into four types measuring scales containing different information that determines the method of statistical analysis (Stevens, 1946). Perhaps the simplest though least statistically descriptive measurement scale is termed a nominal scale, in which a name or number is assigned to the value of the data for identifying purposes. Numbers used in a nominal scale have no particular order, magnitude, or significance in value other than for the purpose of identification, and though this scale of measurement yields little in terms of detailed statistical information, there is less potential for error secondary to the simplicity its structure (Stevens, 1946).

In contrast, an ordinal scale is a level of measurement that does provide a differentiation and value in the numerical assignment (Stevens, 1946). Values can be determined by the ranking the assigned numbers in either an increasing or decreasing fashion, dependent on the design and purpose of the research instrument. An example of this would be 5 point ordinal scale used for an assessment of pain utilizing an increasing design. The number “ 1” would indicate the least amount of pain experienced, with the number “ 5” ascribed to the highest level of pain experienced; and the numbers in between reflecting differences in experiential magnitude. Although in this example the level of pain would be indicated by the numerical assignment, the difference in the numbers in the ranking would not necessarily be equal or precise, and the differences between any two rating numbers would not indicate a specific interval. Thus, an ordinal scale may be considered to be a more sophisticated method for measurement than a nominal scale, yet be less precise for in providing specific information in terms of magnitude than with other levels of measurement such as interval or ratio scales.

Researchers seeking this type of information might use an interval scale, as there are equal spaces between any two values which will provide more specific data than with a nominal or ordinal scale. As the distance between any two positions on an interval scale is of a known size, this method allows for the use of arithmetic operations, and the scale might be used for measurements pertaining to temperature, dollars, out to for, nor anything in which there is a direct measurable quantity with equality in units (Stevens, 1946). As with temperature, zero does not represent the absolute lowest value, and therefore inferences regarding the ratios of data cannot be made as an interval scale does not have a true value of zero (Thorndike & Thorndike-Christ, 2009). A level of measurement that does possess a true value of zero and has the property of ratios has some similarities to an interval scale, and is called a ratio scale. The corresponding ratios on the scale have the same meaning and the value of zero absolute, as in a measurement of height and weight (Thorndike & Thorndike-Christ, 2009). The ratio scale can provide more specific information than the other levels of measurement; however, its complexity can be a factor in terms of the potential for research error. All four of these levels of measurement can be used singly and at times conjunctively in the design of a rating scale instrument dependent on the type of research being conducted. Therefore, the selection of a particular level of measurement is a fundamental concern for the researcher, as the success or failure of a research endeavor may lie in making an appropriate choice, as the subsequent types statistical analyses for each are varied and differentiated.

## Likert and Likert-Type Rating Scales

The level of measurement most appropriate for this quantitative statistical study is an ordinal scale. As the focus of this study is on staff members’ attitudes, trying to render subjective and what might be considered qualitative data into quantitative measurements requires a method for categorizing responses. At the same time, care must be given in considering whether a particular scale item can be measured in terms of the distances between any two numbers in the rating scale. In 1932 Rensis Likert created and developed a methodology often used to assess feelings, attitudes, and opinions by providing a rating scale structure that addressed this concern (Likert, 1932). This method has widely utilized and adapted in many venues including education, social science, marketing, and a variety of other applications with empirical research evidence that confirms its reliability and validity (Abdel-Khalek, 1998; Chow &Winzer, 1992; Maurer & Andrews, 2000). This evidence further suggests that rating bias is reduced and the reliability of the rating is highest when a 5 or 7 point or higher rating scale is used (Stennet, 2002). As a result, a typical Likert scale typically has 5 to 7 categories with 3 or 4 point scales often considered too few, particularly when parametric statistical methods are applied (Garson, 2002). The categories in a 5-point Likert scale is often coded as 1- Strongly Disagree, 2 – Disagree, 3 -Neutral, 4 – Agree, and 5 – Strongly Agree, with percentage and nonparametric statistics used for analysis. The simplest form of Likert scale is analyzed in a summative manner; with the mean value of a Likert-item found by adding the number of responses to a Likert-item and dividing it by the number of points used for the rating scale. Whether or not Likert scales can or should be analyzed using parametric versus nonparametric tests is a source of ongoing controversy within the statistical community (Acock & Martin, 1974). Efforts to quantify responses using parametric tests have led to an adaptation in scale design from the classic Likert model, to what is termed a Likert-type scale (Clason & Dormody, 1994).

A Likert-type scale differs from a Likert scale in the statement questions being singular in the response alternatives, allowing for the data to be treated as interval scale data for parametric statistical analysis (Brown, 2000; Clason & Dormody, 1994; Cliff, 1984; Hodgson, 2003). With this approach, the mean score of Likert-type data from each question or statement could be compared using a t -test, though factors related to data distribution, sample size, and number of rating choices would need to be considered (Clason & Dormody, 1994). This translation of ordinal rating scale data into that of an interval scale for the use of parametric statistical methods underscores the controversy identified by Acock & Martin, 1974). As previously described in the levels of measurement, ordinal scales are ranked with no specific measurable difference or distance between the numerical categories, and statistical calculations used to determine the mean, standard deviation, and patterns of correlation may result in inaccurate research findings (Harwell & Gatti, 2001; Miller, 1998). In the classic Likert scale design numbers to provide order and ranking, but since the actual distance and origin between the numbers is unknown, then it cannot be scored in an additive manner save to indicate the central tendency and range of the responses (Dawes, 2008). This supports the contention that non-parametric test should be used for a quantitative analysis, though it could be argued there are methods that would be more statistically accurate and valid, such as ordinal regression techniques (McCullagh, 1980). Furthermore, the calculation of the mean scores for the ordinal data would need to be established in order to be statistically valid in a analysis using parametric tests (Trochim, 2006), whereas those arguing against this approach contend that the median but not the mean can serve as a measure of central tendency (Trochim, 2006). For this study, the calculations for will exclude the mean, though the mode, median, and range of the score distribution will be presented.

## Data Analysis

A descriptive quantitative research methodology will be used for this study through the use of a survey that was designed to collect data from the direct care providers at the Colorado Mental Health Institute. A survey methodology is often used for data collection from a specific population or from a sample of the population (Robson, 1993). Within the realm of psychology, surveys are used for gathering data about individuals, groups and organizations, and larger social networks (Rossi, Wright, and Anderson, 1983). Most Americans have participated in some form of survey in an either online or in a paper format whether in an educational, social science, political, or marketing research context. Many organizations have employees participate in satisfaction surveys, and surveys in general have become a sort of social barometer for monitoring the public’s attitudes and opinions. Often surveys target a specific population, with a sample survey focusing on subgroups determined by a variety of factors such as gender, race, geography, political beliefs etc. dependent on the nature and focus of the subject areas studied (Rossi, Wright and Anderson, 1983).

The advantages and strengths of using a survey methodology for gathering information are that they can be of a qualitative or quantitative design, and be easily administered to gather information quickly and inexpensively (Leary , 1995). In the past, this was frequently accomplished through the use of paper and mailed surveys, with the latter method used particularly when large populations were being surveyed. With advances in communication and digital technology, surveys can be conduct via electronic formats as varied as those associated with phones and computers. Other frequently used research methodologies such as personal interviews can be more time and cost intensive than with the use of a survey instrument, and anonymity and confidentiality of the respondents can be more difficult to control. These factors strongly influenced the researchers choice of a descriptive research methodology utilizing a survey instrument to assess the Colorado Mental Health institutes’ clinical employees attitudes, feelings, and lived experiences related to the crisis intervention training program the organization has utilized for 25 years. Few attitudinal surveys have been conducted regarding this specific subject area, and a review of the literature did not reveal any studies concerning the long-term impact of this type of training on the culture of a mental health organization.

## The Colorado Mental Health Institute’s 2008 Safety Survey

Another key factor that sponsored this researchers’ interest in designing and administering an attitudinal survey was a review of a safety survey conducted at the Colorado Mental Health Institute in 2008. The Colorado Mental Health Institute currently uses a survey methodology for exit interviews with clients that are discharging from the facility, as well as surveys designed to gauge employee’s job satisfaction. The Institute has only conducted one survey that concerned the subject area of crisis intervention training being focused on in the proposed study for this dissertation. The purpose of the survey was stated thusly: “ In light of the high number of patient to staff assaults in fiscal year 07-08, the hospital focused the annual Failure Mode Effects and Analysis on the Management of Assaultive Patients, and assembled a workgroup to perform the analysis. The hospital director suggested a staff survey specific to perceptions of safety to be done as part of the work on reducing assaults” (Colorado Mental Health Institute Safety Survey, 2008). Although the survey provided some interesting results, there was no data regarding the internal and validity of the survey instrument. Thus, the results are highly questionable in terms of these factors, and the survey cannot be considered quantitatively accurate, though it can provide useful information for developing a more scientifically accurate instrument by analyzing its improper design. It is important to note that in 2008 the Colorado Mental Health Institute had eight treatment units as opposed to the four that currently provide service for adults, with the closure of geriatric, adolescent, and children services in recent years.

## 2008 Survey Methodology

There were approximately 240 surveys distributed to the direct care staff members, with 224 surveys submitted by participants indicating a 93% return rate. The survey focused on the aggregated responses of nurses, Mental Health Clinicians, and Public Safety Officers (security) as the providers with the most direct care contact. The survey used a 5 point Likert scale shown below:

## Table 1.

1

2

3

4

5

Strongly Disagree

Disagree

Neutral

Agree

Strongly Agree

The safety survey was comprised of four questions:

“ I feel safe working with patients on my unit.”

“ I feel safe when I float to another unit.”

“ Crisis management training is useful/helpful in preparing me to deal effectively with agitated and escalated patients.”

“ I feel adequately trained to deal with assaultive patients.”

There was one open-ended question, asking “ What further training do you feel would be beneficial to prepare you to work with the various patient populations at the Colorado Mental Health Institute and the assault risks they present?”

## Table 2. Colorado Mental Health Institute Safety Survey, 2008.

The survey presented tables illustrating findings specific to each treatment unit similar to the one above. They are not included in this study for the sake of brevity, however, it is important to note that there was some disparity between the different populations responding to this survey and a slight difference between the teams and disciplines. These differences were small enough to lend credence to the aggregate results. Overall, the highest scored variable was “ I feel adequately trained to deal with assaultive patients” at 3. 96. The lowest scored variable was “ I feel safe when I float to another unit,” at 3. 21. “ CMT is useful/helpful in preparing me…” registered at a middle value at 3. 21. All of the scores ranged between “ neutral” and “ agree”, and appear to have been derived using a summative method. The following tables further illustrate these findings (Colorado Mental Health Institute Safety Survey, 2008):

## Table 3.

## Table 4.

There were only twenty responses to the open-ended question, “ What further training do you feel would be beneficial to prepare you to work with the various patient populations at the Colorado Mental Health Institute and the assault risks they present?” Some of the responses did not specifically address training, for instance; several respondents said they needed “ more staff.” Other respondents asked for training specific to patient populations, age group, and diagnoses. Several respondents stated that there should be an orientation to any units before floating (i. e. working on unfamiliar units). In addition, some respondents felt that they just needed more training, or more practice.

## Analysis of the 2008 Safety Survey

Considering that this survey uses a 5 point scale the variable indicating 3. 21 when employees “ float” might have to do with the reaction of working in a different treatment setting with patients, staff, and conditions with which the employee is unfamiliar. The 3. 96 rating ascribed to feeling adequately trained is relatively high, though there is a disparity when contrasted with the 3. 21 rating of the perception of crisis management training being useful. The fact that all the readings are neutral or higher would support the contention found in the literature review that some form of crisis management training is more beneficial and preferable to having none at all. The middle score for crisis management training and the tenor of some of the individual comments suggest a performance gap with regards to the training process. This may have had to do with the crisis management training approach, the amount of practice and experience an employee has with using the techniques, or the assessment methods utilized to evaluate employee competency. However, these inferences are merely based on conjecture as the reliability and validity of the research instrument was not established. In addition, using only four Likert-items the survey design seems too limited in scope to gather a comprehensive overview of the employees attitudes and opinions. Therefore, the results are subject to interpretation, lacking in specificity and clarity, and cannot be considered as valid and reliable empirical evidence.

## Data Analysis Methods for the Proposed Survey

The data analysis for the proposed survey for this study will begin by examining the surveys for correctness and completeness, and then numerically coding the responses into a database in the PASW-18 software (Predictive Analysis Software) and performing an analysis of descriptive responses. Surveys that are incomplete will be were discarded from the analysis, and the frequency tables and descriptive statistics will display the results relevant to answering the research questions. As described previously, the calculations will exclude the mean, though the mode, median, and range of the score distribution will be presented. A Chi-square test will conducted to determine if a relationship exists between frequencies, although it will not indicate the strength or positive/negative direction of the relationship. This is consistent with standard statistics guidelines for using the non-parametric tests that will be used in this study as opposed to parametric tests; for as Jamieson (2004) advises, “ the appropriate inferential statistics for ordinal data are those employing non-parametric tests, such as chi-squared, Spearman’s Rho, or the Mann-Whitney U-test1 because parametric tests require data of interval or ratio level” (p. 2113). As the numbers in the survey that will be assigned to Likert-items fall into the ordinal measurement scale, the descriptive statistics will include a mode or median for central tendency and frequencies for variability, and the inferential non-parametric tests will follow Jameison’s (2004) recommendations, including Spearman ‘ s Rho to detect linear relationships and Kendall’s tau-b for any increasing or decreasing relationship between the Likert scale questions.

## Data Collection

Crisis intervention training is mandated for those staff members at the Colorado Mental Health Institute that provide direct care for clients residing in inpatient settings. In addition, those members in the Institutes’ administration that oversee this care are also involved in biannual trainings. Therefore, the sample population will consist of 210 clinical staff members, with the surveys distributed in the upcoming training review for the clinical staff members at the Institute. Although attendance for this training is mandated for employees, it will be made clear that participation in the survey is voluntary, with the anonymity of the participants made clear to addressing the issues regarding confidentiality. The respondents will be assured that the survey is anonymous and confidential, though they will be given the opportunity to provide information regarding their respective disciplines, years of service, level of education, and gender. Once the data has been collected, the results will be recorded and statistically analyzed using PASW-18 software. Prior to the distribution of the surveys during training review, the purpose and importance of the study will be announced in each of the treatment units team meetings, and posted for all clinical staff members via the Institutes’ intranet e-mail service.

## Research Questions/Hypotheses

\_\_\_\_\_ Describes specific research questions and hypotheses (where appropriate) that

\_\_\_\_\_ Are clear and succinct

\_\_\_\_\_ Are congruent with the Statement of Problem

\_\_\_\_\_ Are answerable/testable

\_\_\_\_\_ Correspond to the number of variables of interest

\_\_\_\_\_ Have hypotheses that correspond to research questions

\_\_\_\_\_ Are clearly stated

\_\_\_\_\_ Are open-ended (not yes/no questions)

## Population and Sample

The sampling frame for the survey consists of 210 members of the Institutes’ direct care staff, whose treatment teams multidisciplinary structure are comprised of management staff, nurses, social workers, psychiatrists, psychologists, therapeutic recreation specialists, occupational therapists and mental health clinicians. Although these disciplines function within a team context, there are differences in terms of power and status, as well as their amount of direct client care. Examples of these differences in station are the administrators, psychologists, and psychiatrists that comprise the upper echelon within the teams, with middle managers functioning as liaisons between the lower ranking team members and those higher up in the organization. The social workers on the team have a middling status with regards to power and prestige; with nurses and mental health clinicians representing the lowest status with the greatest amount of patient contact, though having the least contact with the hospital administration. Although this last factor is interesting from an organizational point of view, it is even more relevant in terms of the study as those staff members having the most direct contact with clients are the most likely to encounter and intervene in violent situations occurring on the inpatient units.

## Ethical Concerns

There are ethical concerns inherent in any form of research study, particularly those involving the use of human subjects. To address these, procedures for protecting confidential and anonymity in data collection, analysis, reporting, and storage need to be developed and described. The researcher must be competent to perform the research procedures in order to reduce the risk of causing harm to human subjects, and the reputation of the sponsoring organization as well as the organization were the study is being conducted (Welman et al., 2005: 182). For this proposed study, the researcher has completed the coursework for the Doctoral Program in Organizational Psychology at the University of the Rockies. This course of study has been grounded in the orientation and practice of conducting research, utilizing different kinds of research methodologies, undergoing training relative to conducting legal, ethical, and appropriate research procedures, and applying qualitative and quantitative analysis in a variety of projects under the tutelage of qualified instructors at the doctoral level of education. There are a number of safeguards implemented by the University the Rockies to ensure that ethical concerns are addressed in an appropriate manner, including a requirement that the student completes two courses related exclusively to proper conduct and methodology in the research process, as well as the ethical research guidelines that need to be understood and followed. Throughout this training, the importance of the researcher’s work needing to be authentic and built upon his or hers own ideas have been emphasized, citing the use of other people’s ideas and data to avoid plagiarism. In addition, it has been made clear that providing data that is falsified or misleading is inappropriate and unethical. To ensure appropriate ethical standards are followed, student researchers at the University the Rockies are supervised by a committee of three psychologists, one serving as the committee chairman who works closely with the student. Finally, if the committee supports and approves the research proposal, it is reviewed by the University’s Research Review Board and Institutional Review Board to ensure that the proposed study is original, contributes to the body of work in the field of psychology, and is held to the highest ethical standards.

Factors related to the fairness of the proposed survey pose some ethical considerations. It is important that respondents understand the language used in the survey, and efforts have been made to make sure that each item is clear and easily understood. The specificity and clarity of the survey items was tested in the pilot study to assess if the respondents knowledgeable in the subject area interpreted each item in the intended way, and that this intention is made clear. The survey items are of a relatively simple construction reviewing concepts revisited frequently throughout the training process. Efforts have been made to design a survey that meets the validity and reliability standards discussed

The possibility of ethical concerns arising for the respondents in this proposed study are reduced secondary to the anonymous and confidential nature of the survey; and that the previously, and the survey will be conducted with the same time parameters and conditions for all participants. voluntary participants being surveyed are employees as opposed to the Institute’s clientele. Nevertheless, individuals participating in the survey may question the viability of keeping this information confidential, and worry that an honest and full disclosure of their views and feelings regarding the subject area may negatively impact their status within the organization. Efforts will be made to neutralize this factor by giving the respondents an accurate description of the study, and pointing out the potential benefits for the organization by improving the welfare and safety the Colorado Mental Health Institutes’ clients, as well as those charged with the responsibility of providing their care. The results of the study will be made available to all members of the Colorado Mental Health Institute, including any results that are associated with researcher error.

## Instrumentation

The survey proposed for this study was designed to gather data based on the proposed research questions, and provide information to make appropriate conclusions and inferences based on empirical data. The instrument was designed to measure each respondent’s opinions, attitudes, lived experiences regarding the perceived value and effectiveness of the Crisis Prevention Institutes’ Nonviolent Crisis Intervention Program to manage violence. The attitudinal scale that will be used in the study will be based on a Likert scale, which is often used as a form of summative scale. The design is formatted using a variety of Likert scales with statements that the respondent evaluates by expressing their general level of agreement or disagreement. The scales proposed for this study to assess respondent’s perceptions of the Crisis Prevention Institutes’ Nonviolent Crisis Intervention Program effectiveness will use a 5-point Likert design. The survey is anonymous and voluntary with no identifying information, though clarifying information such as gender, years of service, occupation, and years of education will be requested. The Likert-items for the survey targeted three general areas: 1) training objectives and content, 2) method and training context, and 3) usefulness and knowledge transfer. In the first category, 8 Likert-items refer to specific aspects and components in the training process. The second category contains 7 Likert-items that concerns applications of the training outside of the training event, as well as how it integrates with other training initiatives sponsored by the Colorado Mental Health Institute. The third and final category contains 6 Likert-items that are designed to explore employee attitudes regarding the training in terms of its relevance and utilization in real-life situations. Table 4. shows the rating scale that is used consistently throughout the survey, and for each of the three described categories.

## Table 4. Rating Scale.

1

2

3

4

5

Strongly Agree

Agree

Neither Agree or Disagree

Disagree

Strongly Disagree

The survey has been reviewed by a panel of experts (committee members), who offered recommendations to make the Likert-items more specific and measurable, and then was administered to a screening sample of five people comprised of subjects similar to those who will eventually participate in the survey for this study. The inclusion of the neutral “ Neither Agree or Disagree” category was included to avoid forcing the respondents to choose a response. The numbers of choices on a 5-point scale will present an odd number of choices, allowing respondents to remain neutral. The decision on whether or not to include a neutral category in a survey has been debated, with a decision either way potentially impacting the assessment results (Mogey, 1999). The participants in the screening/pilot sample will be excluded from the final pool of respondents for the survey, which was designed to be completed no more than 10 minutes of time. Some of the considerations in constructing the survey include respondents possibly being influenced by the way they have answered previous questions, in terms of establishing a pattern that they feel the need to break. In addition, some members may desire to take extreme options but instead temper their responses with more moderate ones. This can be sometimes a challenge in terms of testing in a psychologically oriented environment, where participants are adept at finding and arguing dichotomous and often opposing viewpoints.

## Internal and External Validity

Establishing the reliability and validity of the survey instrument proposed for this research study is fundamental as it is created as opposed to a published instrument. Suskie (1996) describes a questionnaire or survey as reliable when it elicits consistent responses from the participants, and this can be accomplished by providin