Improving Institutional Corrections Training Academy:

Analytical Tools for Practitioners for Outcomes-based Training Evaluation

December 2024

The purpose of this resource is to provide analytical tools for practitioners in developing their outcomes-based training evaluation. Examples included in this handout are based on the Bureau of Justice First Line Supervisor Pilot Training developed by The Moss Group.[1] This document is a supplement to the material covered in the webinar series and does not serve as standalone training. Please see the series linked here for more information.

Mini Table of Contents

Replacing Subjectivity with Evidence

Measuring effectiveness is critical for funding decisions, resource allocation, policy and program management, and training development. Data across criminal justice organizations frequently revolve around overarching crime reduction efforts, overtime, and recidivism rates. However, incorporating data into everyday strategic decision-making does not have to be confined to exhaustive research or research taking place over decades. Given the appropriate tools, practitioners can also measure the impact of policies, the effectiveness of programs, and the appropriateness of training programs (thereby ensuring training is achieving what it is designed to accomplish). Calculating averages, frequencies, and percent change, along with assessing survey data and participant feedback, empowers practitioners to engage in real-time functional analysis. With manageable data and user-friendly analytical tools, practitioners can assess whether what is being done is effective or if change is necessary.

Replacing a “gut feeling” or a “hunch” with evidence and results does not have to be overwhelming. There are user-friendly tools available online that provide calculations for many statistical analytics. Excel, for example, is a user-friendly way to analyze data with the resources most staff can readily access. When these tools are accessible, data

[1] The First Line Supervisor training is linked here.

analysis and statistics can and should be incorporated into practice for managers and trainers.

Understanding the Data

To ensure evaluations include the most appropriate and useful metrics, it is important to understand the data. When evaluating training programs, practitioners employ both qualitative and quantitative data analysis. Quantitative data refers to numerical, measurable data (i.e., numbers); qualitative data refers to non-numeric, descriptive data (i.e., narrative). While quantitative data are often used to identify patterns and make predictions through empirical data analysis, qualitative data are generally used to explore and uncover new information via observations, interviews, and case studies, for example. Practitioners use quantitative data to understand "how much" and qualitative data to answer "why."

Outcomes-based training and evaluation can incorporate both types of research. For example, a qualitative analysis involving interviews with staff to determine why morale is low may suggest staff are struggling with emotional regulation. As a result, a human resources unit may implement mandatory emotional intelligence training for all staff. Upon implementation, the training may show to be effective if the number of staff trained on emotional intelligence positively correlates to staff morale and coincides with fewer incidents of internal staff complaints. By using data and information to build the training, trainers replace subjectivity with evidence. Once participants complete the training, trainers are able to validate the training's effectiveness via course tests (designed to assess a participant's understanding of emotional intelligence) and follow-up data (such as comparing the number of internal complaints pre-training versus post-training).

Quick and Easy Analytics

Ultimately, there are numerous analytical techniques and tools available to practitioners. Some of the most user-friendly and efficient include measures of central tendency and percent change.

Measures of Central Tendency

Practitioners frequently use measures of central tendency (mean, median, and mode) when evaluating quantitative data. Concisely, the mean refers to the average within a set of data points, the median is the middle number in a numerically ordered list, and the mode is the number that occurs the most frequently in a data set.

The most frequently used measure of central tendency is the mean. One way the mean can be used is to assess the average competency scores of participants following training to determine if the training is accomplishing the desired outcome. For example, assume first-line supervisors participate in Module 1 of the Bureau of Justice First Line Supervisor Pilot Training developed by The Moss Group. Thirty participants complete an associated skills assessment (e.g., a quiz) evaluating, in part, their understanding of "Lead and Follow."

To determine the average score of the participants, the trainer adds each participant's assessment score and then divides that total by 30 (the total number of participants tested). If 15 participants score 90% on the assessment, one scores 85%, and 14 score 80%, the group average for the assessment is 85.2%.

Consider:	What is the central tendency of the data?
Example:	The average score for the post-test is 85.2%.
The Math:	Mean - sum of all scores divided by the total number of test takers ((15*90)+(1*85)+(14*80))/30 = 85.16 So, participants scored an average score of 85.2. Median - central number in the data set 15 = 90, 1 = 85, 14 = 80 Middle numbers = 85 and 90 So, the median score is 87.5. Mode - most frequently occurring number 15 = 90, 14 = 80, 1 = 85 So, the mode is 90.
Online Resource:	Calculator Soup[1]
Excel Function:	Mean = average(number1, [number2],…) Median = median(number1, [number2],…) Mode = mode(number1, [number2],…)

Analyzing Rating Scales

Another metric that could be assessed is whether participants find training helpful. One question on an assessment could ask: Please advise how much you agree or disagree with the following statement: "I found this training helpful." Trainees can select a response using a Likert scale where 5=strongly agree; 4=agree; 3=nether agree nor disagree, 2=disagree or 1=strongly disagree.

Assume that 27 (90%) of trainees responded to this question. Of those 27, 12 answered strongly agree (5); 8 responded agree (4); 3 said neither agree nor disagree (3); and four replied disagree. When the scores are added up, the total is 109, divided by 27 participants = an average of 4.03. On the scale, 4 means agree, which indicates that, on average, the 27 participants agree the training was helpful.

When analyzing rating scales, such as a Likert scale, it is important to remember to exclude any missing cases. Using the same example, three (10%) of the trainees did not respond to the question regarding how much they agreed or disagreed that the training was helpful. If those responses were included in the analysis, the scores would still total 109; however, the divisor would be 30 (the total number of trainees instead of the total number of participants who responded). As a result, the average becomes 3.63. On the scale, 3 means neither agree nor disagree, this indicates that, on average, the participants neither agree nor disagree that training is helpful. By including the three missing responses, the average is negatively altered and could impact future decision-making. So, to prevent inaccurate evaluation when using rating scales, do not include missing cases in your analysis.

Consider:	What is the central tendency?
Example:	On average, participants agree training is helpful.
The Math:	Sum of the scores divided by total responses [(12*5)+(8*4)+(3*3)+(4*2)]/27 = 4.03 So, the average response was 4 (agree). What if the missing cases were mistakenly included? [(12*5)+(8*4)+(3*3)+(4*2)+(3*0)]/30 = 3.63 Including non-responses, in this case, results in an average response of 3 (neither agree nor disagree).
Online Resource:	Porsline[1] Various tutorials - search “Excel tutorial on analyzing Likert scales.”
Excel Function:	Mean =average(number1, [number2],…)

Frequencies

Another useful metric for practitioners is frequency, or the number of times an event or observation occurs. Calculating a frequency helps practitioners and trainers gain insight into the distribution of data across a group (i.e., class, correctional staff, individuals under supervision).

For example, assume 30 correctional staff participate in the Bureau of Justice First Line Supervisor Pilot Training developed by The Moss Group Module 3: The Skills: Strategies and Techniques Used by Effective Supervisors training. This training is designed to enhance skills such as listening, empathy, adaptability, and discretion. The trainer wants to ensure the curriculum and the associated assessment are

appropriate and is curious as to whether any assessment questions are consistently problematic for participants. To test this, instead of reviewing individual results, the trainer can analyze the data collectively to determine how often, or the frequency by which, participants are accurately or inaccurately responding to each metric. If 24 of the participants respond to a particular question incorrectly, for example, the trainer may decide to review the training to ensure the curriculum appropriately covers the material assessed by that question and the wording of the question to ensure it reflects the material covered in the training.

Consider:	How many times does something occur?
Example: indicator(s) on	# of participants incorrectly answered the same key an assessment
The Math:	# of evaluated outcome/all occurrences 24/30 = 0.80 * 100 = 80% So, 80% of participants responded incorrectly.
Online Resource:	Social Science Statistics - Frequency Distribution Calculator[1] OmniCalculator[1]
Excel Function:	=countif(range, criteria) Where the range refers to the range of cells from which you want to count, and criteria are the conditions that define

Calculating Percent Change

A simple calculation to assess the difference between two quantities is percent change. This calculation requires two sets of data, such as pre- and post-test scores. For example, the aforementioned First Line Supervisor training, “The Basics,” begins with a pre-test to assess the pre-training competency of 30 staff members. Following the training, the trainer issues the same assessment (post-test). Comparing each participant’s pre-test results to the corresponding post-test results allows the trainer to determine if there was any change in each participant’s competency relative to The Basic’s curriculum of mission creation, leadership, boundaries, and communication.

Taking it a step further, trainers can compare the average pre-test scores for the group of 30 participants to the average post-test scores for the same group to determine if, collectively, the training increased the group’s competency overall. Assume the average pre-test score for the group was 50 and, after completing the training, the group’s average post-test score is 85. The differences (between 50 and 85) show the training had a positive impact. To reflect the depth of that change, the percent change can be calculated. In this case, the percent change would be 70%.

Consider:	What is the change in values?
Example:	Average pretest score = 50 Average posttest score = 85
The Math:	((Time2 - Time1)/Time 1)*100 ((Pretest score - Posttest score)/Pretest score)*100 ((85-50)/50)*100 = 70% So, 85 is a 70% increase of 50.
Online Resource:	Calculator.net[1]
Excel Function:	No function available.

Additional Resources and Tools

More information about these and other statistical methods can be found online at https://www.socscistatistics.com/tests/. In fact, Social Science Statistics designed an online tool titled "Which Statistics Test Should I Use?[2]," to help users determine the best statistics test based on the variables.

Today, we have access to numerous data collection and data analysis tools at our fingertips. Many of these tools are free, publicly available resources with sufficient capacity to help managers and trainers make educated decisions about policies, programs, and training curriculums.

[2] https://www.socscistatistics.com/tests/what_stats_test_wizard.aspx
[3] https://www.statskingdom.com/statistical-power-calculators.html
[4] https://www.socscistatistics.com/
[5] https://www.socscistatistics.com/tests/

For data collection, Google Forms, Survey Monkey, Mentimeter, and Otter AI are useful tools. For data analysis, various platforms such as Power BI, Qualtrics, and online statistics calculators make outcomes-based training and evaluation attainable even when resources are scarce.[3] [4] [5]

This project was supported by Grant No. 2019-RY-BX-K002 awarded by the Bureau of Justice Assistance. The Bureau of Justice Assistance is a component of the Office of Justice Programs, which also includes the Bureau of Justice Statistics, the National Institute of Justice, the Office of Juvenile Justice and Delinquency Prevention, the Office for Victims of Crime, and Office of Sex Offender Sentencing, Monitoring, Apprehending, Registering, and Tracking. Points of view or opinions in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.