DRAFT

SFA University Measurement & Evaluation Capabilities

Purpose/Overview

To highlight Performance Measurement and Evaluation capabilities to consider when evaluating training courses during the "Assess It" phase. This document will provide a high-level description of measurement frameworks and methods/tools available, while also mapping their appropriateness for when to use what. To conclude, a detailed process will be shared for designing surveys.

Objectives

- Identify and define types of measurement frameworks
- Identify and define types of evaluation methods/tools
- Identify when to use what framework and what method
- Provide processes for how to build evaluation tools (Surveying)

Why Measure?

Measuring the results of a training program can provide answers to questions about a program's effectiveness as well as identify specific ways for improving training. And depending on the type of analysis that you wish to perform, you must also choose specific frameworks, methods and tools. For example, a trainer may use Surveys to capture a trainee's immediate reaction to training, a supervisor may monitor employee performance to track learning effectiveness and time to competency, or Finance may use financial analysis to measure the cost/benefit effectiveness of various training programs.

SFA University Measuring Capabilities:

- Determine if customers (students, schools and financial partners) are receiving the capabilities they need from training. Is the current delivery method, content, and/or timing effective?
- Determine if customers (students, schools and financial partners) are retaining the capabilities and applying them to the job. Do they have the "right" skill set and does their performance meet expectations?
- Understand if partners/customers are satisfied with the design, development and delivery of learning services.
- Evaluate if current learning services are designed, developed and delivered in a cost-effective manner. Is there a favorable ROI?
- Determine if University Services, External Partner Services and Channels are providing appropriate levels of support to the target audience post-training.
- Evaluate the performance and skill set of subject matter experts and trainers. Do they have the "right" skills to effectively perform their jobs?

CAFFARELLA Evaluation Process (Dick Leatherman: The Training Trilogy pages 214-218)

	Step	Operational Guidelines
1	Identify the individuals to be involved in planning and overseeing the evaluation	An individual or team of individuals should be designated to oversee the program evaluation process
2	Define precisely the purpose of the evaluation and how the results will be used	Purpose should be clearly stated and understood by all parties involved; especially top management and supervisors of those being trained
3	Specify what will be judged and formulate the evaluation questions	• Participants learning (knowledge, skills and attitudes) • Participants job performance • Impact on the Organization (reduced costs, improved quantity, reduced turnover, less absenteeism, etc) • Impact on the training unit's policies, procedures, practices, etc Questions should address the purpose of the evaluation

4	Determine who will supply the needed evidence	From participants, supervisors, training staff, instructors, management personnel, or outside consultants
5	Specify the evaluation design to be used	Choose a design that matches the evaluation purpose, questions, and nature of the program. Examples of quantitative evaluation designs include: 1. One group pretest and posttest 2. One group time series 3. One nonrandomized control group
6	Determine the data collection techniques to be used	The techniques chosen should be based primarily on the purpose of evaluation and design chosen. In addition, the characteristics of the respondents, the expertise of the evaluators, and the time and cost requirements should be considered. (Interviews, written questionnaires, review of performance records, etc)
7	Specify the analysis procedures to be used	Related directly to the evaluation questions, design and kind of data collection techniques used. For quantitative data, they can range from simple numerical counting or percentage reporting to very sophisticated statistical analysis. Qualitative data is usually reported in prose form, though some simple numerical tables are also used
8	Specify what criteria will be used to make judgments about the program	Should indicate the level of performance or change that will be considered acceptable. Criteria should be set for each major evaluation
9	Determine the timeframe and the budget needed to conduct the evaluation	Timeframe may be a set time (before and after a specified program) or be done on a continuous basis (recording of change in the learning of participants for all training programs). Budget should be defined prior to initiating the process

Measurement Frameworks (Amy Day & Elizabeth Hagerty: Accenture)

What Evaluation Frameworks Are Available?

- The Kirkpatrick Model (Donald Kirkpatrick: Evaluating Training Programs 2nd Edition pg 19-67)
  - Level 1 - Reaction: Reaction measures the learner's level of satisfaction with the training course. It is a measure of the feel good factor of a program (e.g., how well did participants like the learning experience)
  - Level 2 - Learning: Learning measures the impact of training on the learner's level of knowledge, skills, and attitude. To what extent did participants change attitudes, improve knowledge, and/or increase skill as a result of attending the program?
    Level 3 - Behavior: Behavior measures the transfer of learning to the real work environment. Level 3 can be the most time consuming and costly to conduct. The climate induced by supervisors can have a positive or negative affect on behavior that is unrelated to learning.
  - Level 4 - Results: Results measure the impact of learning on bottom-line business measures and the duration of those results over time (increased production, improved quality, decreased costs, reduced frequency and/or severity of accidents, increased sales, reduced turnover, and higher profits). Results can be the most difficult to quantify. The data this level provides is insightful but it is difficult to assess which impacts are directly attributable to training versus other change initiatives that are typically occurring in top companies.
- Return on Expectations Model (ROE)
  - Emphasizes evaluation of learning and behavior
  - Three steps:
    - Step 1 - Stakeholders' expectations: Identify the expected outcomes of the training from all relevant stakeholders. Answer the questions: What should the participants learn? How should the participants perform?

- Step 2 - Post-training Outcomes: Determine actual outcomes of post-training. Answer the questions: What did the participants learn? How did the participants perform?
- Step 3 - Degree of linkage between expected and actual outcomes: Compare actual post-training outcomes and expected outcomes. Answer the question: How similar are the expected and actual outcomes?

- Time to Competency
  - How long does it take employees to learn their job and ultimately increase corporate profits (e.g. Call Centers – decreased handling time and improved one-call resolution)
- Time to Market
  - The faster a product gets into the customers hands (decreased time in production cycle)
- Balanced Scorecard
  - Helps to ensure that all critical performance measures are evaluated. Provides a check and balance so that one area isn't overemphasized at the expense of another
  - Tracks the key elements of a company's strategy using financial and operational measures
    - Financial Perspective – how do we look to our shareholders? (Decrease Unit Costs/Increase Profits)
    - Customer Perspective – how do customers see us? (Improve Customer Satisfaction)
    - Internal Perspective – what must we excel at?
    - Innovation and Learning Perspective – can we continue to improve and create value? (Improve Employee Satisfaction)
- ROI (Return on Investment)
  - Comparing the monetary value of the results with the costs of the training
  - Cost savings from instructor salaries, travel and refreshments

Measurement Methods/Tools (Amy Day & Elizabeth Hagerty: Accenture)

- Surveys and questionnaires
  - Captures trainees' immediate reactions
  - Useful for illustrating any attitudinal or executional changes
  - Simple to administer and analyze
  - Relatively low cost
- Focus groups
  - Reduces high cost of interviewing individuals
  - More detailed/honest assessment of training impact
  - Useful for capturing the impact of training on job-related behaviors
- Written and Performance tests
  - Demonstrates if a trainee learned the skills or knowledge designed by the program
  - Easy and quick to compare pre-training and post-training tests
  - Combine with Control Groups
  - Performance tests can be in the form of role playing or skills practicing
- Direct Observation
  - Useful for assessing if the trainee is applying newly learned knowledge/skills
  - Accuracy may be questionable if trainee is aware of observation (halo effect)
  - Highly labor intensive/expensive
- Interviews
  - Interview trainees, supervisors/managers, trainers or colleagues to account for changes in behaviors or gauge reaction of courses
  - More detailed/honest assessment of training impact

- Performance Measures

  Approach to developing a feedback process that identifies and monitors achievement of the goals and objectives of an organization, unit, or individual. It involves defining a balanced scorecard of performance metrics that translates organizational goals and objectives into a concrete set of measurable indicators. It also emphasizes the creation of supporting architecture to enable ongoing performance measurement and ensure linkage with decision-making. Demonstrate improved on-the-job performance and bottom-line business impact
- Financial Analysis
  - Cost/Benefit Analysis
  - ROI (Return on Investment)

- Control Groups
  - Used in conjunction with experimental groups to measure learning, behavior changes and results
  - Often used concurrently with other measurement methods such as performance tests, direct observation, interviews and performance measures to gather comparison data with experimental groups

When should I use what tools?

How do other measurement frameworks and tools link to Kirkpatrick's 4 levels?

Measurement Frameworks	**Level 1 (Reaction)**	**Level 2 (Learning)**	**Level 3 (Behavior)**	**Level 4 (Results)**
ROE (return on expectations)		X	X	
ROI (return on investment)			X	X
Time to Competency			X	X
Time to Market				X
Balanced Scorecard		X	X	X
Measurement Methods	**Level 1 (Reaction)**	**Level 2 (Learning)**	**Level 3 (Behavior)**	**Level 4 (Results)**
Surveys	X		X	
Focus Groups	X			
Written & Performance Tests		X		
Direct Observation			X	
Interviews	X	X	X	
Performance Measures			X	X
Financial Analysis				X
Control Groups		X	X	X

How do I build these tools?

1. **Surveys** (Tad Waddington: Accenture)

Where can I use Surveys/what are the benefits?
Surveys are relatively cheap and quick methods of gathering information to your questions. Surveys can be made useful in situations such as:

- You want to find out if your employees are unhappy (satisfaction survey)
- You want to decide whether to fund or to kill Project Q (program evaluation)
- You want to evaluate the effectiveness of a course in achieving its learning objectives (learning evaluation)
- You want to know what the most pressing issues in your firm are (information gathering)
- You believe that the higher the level of the person in the firm, the more games of golf they play in a week (hypothesis testing)

What should I watch out for when creating surveys?
Surveys suffer from four types of error, but a good survey guards against them while answering the question driving it.

- **Sampling error**: This is the ± 3% or ±5% you see in the Gallup poll that indicates the range of error, or error term, of a sample. If you don't survey everybody in the population and do a sample of the population, then a quick and reasonably accurate way to know how much error you have is, where n is the sample size

$$e = \frac{1}{\sqrt{n}}$$

- **Non-response bias**: Ideally, everybody you sampled responds to your survey so all you need to worry about so far is sampling error. The problem with non-response is not just that it eats into your sample size, but that it introduces an unknown and unknowable source of error into your data. The error occurs when only certain groups of people are responding (e.g., only happy customers or only angry ones). Try to figure out who is responding and who is not, and whether it is biasing your data. A response rate of 30 to 70% is tolerable; above 70% is very good.

- **Measurement error**: Measurement error affects the reliability of the data. Item difficulty, rater severity, the way the questions are written, and the rating scales that are used can affect measurement error. For instance, an item that is very difficult to understand, that is poorly written, or that uses a poorly constructed rating scale will increase the amount of measurement error in the data. Likewise, an item that is moderately difficult to say "yes" to, that is well written, and that uses a well-constructed rating scale will minimize the amount of measurement error.

- **Coverage error:** Can your survey reach every person in the population? Phone surveys of people who don't have phones and E-mail surveys of people who don't have E-mail suffer from coverage error. Sometimes, however, this is not a problem. If you are trying to determine who will be elected president, the phone survey works because people who don't own phones usually don't vote either. Think through the potential coverage error.

What are the steps in building a Survey?

DRAFT

Step 1 – Survey Goal: Ask why you are conducting a survey. What decision(s) do you need to make? What action(s) do you intend to take? How will you use the information? (Use of a Steering Committee)

Step 2 – Determine Time & Budget Constraints: Consider your time and budget constraints. If it is possible for you to administer your survey electronically, you should do so. Electronic media is faster and cheaper. Think about what media you will be using to conduct your survey. The media you choose will have consequences for data management and may produce some biases in the data. Possible choices of media include:

- Electronic web
- Electronic Mail (email)
- Paper-based handout
- Paper-based mail-out
- Telephone
- Voicemail
- IVR (Interactive Voice Recognition)

Step 3 – Choose Survey Type: Define the objectives for what you are evaluating

- Program Evaluation (e.g. Does the current Orientation program help employees adjust to their jobs?).
- Learning Evaluation (e.g. A company wants to know whether an eCommerce course has improved the ability of its employees to manage eCommerce projects. The company wants to know what parts of the course are most valuable and effective for its employees)
- Satisfaction Survey (e.g. A company wants to monitor employee satisfaction in targeted areas of training courses offered over time, in order to improve its services, build its business and maintain its customer base)
- Information Gathering (e.g. A company wants to find out what the two most popular training courses are offered for Customer Sales)
- Hypothesis Testing (e.g. A manager thinks there is a correlation between two things and wants to find out if he/she is right. This would be a survey designed to test your expected result against the actual result)

  Before moving on to Step 4, first you need to define the objectives of each of the activities, events or services in the program that you want to evaluate.

Step 4 – Sampling

Don't bother everybody if you can get the information you need by bothering only some of the people. DO A SAMPLE instead of questioning everybody. The sample is the selected group that you are generalizing to the population in your survey.

- **How do you determine the size of the sample?**

  The size of the sample you need to survey is dependent on: How large a population you want to generalize to, which demographic groups you want to compare against each other, how large an error term you can tolerate (below +- 5%), and the response rate to your survey.

- **How do you select a sample?**

  RANDOMLY!!! The most important aspect of selecting a sample is that it must be random.

Step 5 – Writing Questions: Keeping item difficulties and rater severity in mind, tailor questions to your sample population and survey goals or purposes.

- Item Difficulty: How you phrase your questions can influence how respondents answer. Therefore, you want to phrase the questions in such a way that you can prevent obtaining the same answer from all your respondents.
- Rater Severity: Some people are hard graders and some people are easy graders.

Dos

1. Keep questions fewer than 25 words so they remain short and easy to understand.
2. If at all possible, keep your survey under 30 questions, or 20 minutes duration.
3. Keep questions as concrete and specific to your goals and purposes as possible.
4. Organize your survey into simple sections where your contents and scales flow logically.
5. Keep one voice throughout the survey.
   For example, if you begin the survey with "I agree …", continue to begin the questions with "I agree" not "Do you agree…?"
6. Each question should measure one item.
7. Ask a series of questions that encourages respondents to recall the particular event or time period before they answer the really important part of the question.
   For example, "Thinking back to the last election, did you happen to vote?"
8. Put open-ended questions at the end of the survey. Generally speaking, open-ended responses are harder for respondents and interrupt the flow of the survey.
9. Put your demographic questions at the end of the survey. Studies have shown that these questions influence how the respondents feel about themselves before answering the questions.
10. If you want to ask about the respondents' overall impression, put the question at the beginning of the survey. This alone drives up item difficulty, thereby giving you better measures.

Don'ts

1. Don't use loaded questions.
   For example, "Many doctors recommend exercise as a way of maintaining good health. Do you agree? In this example, nearly 100% of respondents are likely to agree. If so, then you have not gathered any information about the respondents' opinions on the relationship between exercise and health.
2. Don't use double-barreled questions.
   For example, "Are you satisfied with the amount and kind of information you receive from your benefits administrator?" In this case, when you receive the answers, you won't know which part of the question the respondent is answering. Are they responding to the amount of information or the kind of information?
3. Don't change scales unless it is necessary.
4. Don't antagonize respondents, or ask questions that make them immediately defensive. As a general rule, using a "why" question will make your respondents feel defensive. *For example, "Do you often cheat on your payroll?" or "Why are you dissatisfied with your job?"*
5. 5. Don't put open-ended questions at the end of every question on your survey. This will generate a vast amount of qualitative data that you will not be able to analyze as well.
6. If you're going to ask specific open-ended questions, like, "*What was most valuable about the course?"* or "*What aspects of the course could use improvement?"* Don't force respondents to list a specific number of things. Some people may have only one thing, and some may have seven. Let' the respondents decide how many things were important to them.
7. Don't use idioms, slang, or unnecessary jargon.

Step 6 – Designing Scales: A scale is the tool that translates text statements into quantifiable numbers. We use scales in a survey in order to be able to score and compare people's attitudes or beliefs. Think of scale as a ruler

that gives measurements of the answers. One dilemma is that a good scale gives you more reliable data than does using many scales. It allows you to use the power of mathematics, statistics, and measurement. But one scale does not necessarily fit all questions and therefore one scale can be highly constraining. Another dilemma is that the more categories in a scale, the higher your data's reliability *can* be, but research shows that people best process a maximum of five categories in a scale. So it might seem that a 5-point scale would be the best choice.

Points to watch out for when Designing Scales:

- Avoid vague quantifiers to obtain more precise estimates. What is meant by 'regularly" can vary for different respondents.
- Beware of semantic weight of word choices. Words like 'useful' and 'functional' have different implications
- Remain consistent in word choices. 'Delight,' 'satisfaction' and 'terror' are not at all the same sort of thing.
- Not all answer choices will be presented on a scaled continuum. Multiple-choice answers are common for demographic questions, while open-ended questions allow the respondent to provide any answer to the question and are most appropriate for information gathering questions. (Note: open-ended questions should be used sparingly since respondents are less likely to respond and the answers are much more difficult to analyze. A comments area where the respondent can write general feedback has similar advantages and disadvantages as open-ended questions and should be used as a summary tool rather than to evaluate discreet information).

Examples

1. Likert scale

The Likert framework uses a series of similar multiple-choice answers presented on a continuum from positive to negative to measure attitudes and beliefs. Two common scales used in conjuncture with the Likert framework are the "Agree" and "Expert" scales.

Agree - Asks respondents to state agreement or disagreement to a particular statement. Not being able to agree or disagree with a statement is not the same as having no opinion. The former is actually an opinion itself, while using 'no opinion' as a mid-range category makes it difficult to discriminate whether people truly have no opinion or simply want to response somewhere between 'agree' and 'disagree.'

"Agree" Scale
1. Strongly disagree
2. Somewhat Disagree
3. Neither agree nor disagree
4. Somewhat Agree
5. Strongly agree

Ability – Also known as the **Expert** scale. Asks a respondent to suggest a degree of proficiency in executing a specific task. The "Expert" scale gives you a very concrete sense of how well the respondents can perform a skill. The more or less scale can give a comparison but does not provide you with an exact rating of skill.

"Ability" Scale
1. Little/no ability
2. Perform with substantial guidance

3. Perform with minimal guidance
4. Perform independently
5. Expert

As seen above, both scales associate a numerical value with each response later used for statistical analysis and reporting. Therefore, directions asking a respondent to select the most representative answer choice should accompany scales.

2. Other Scales & samples

Sample questions:	***Do***	***Avoid***	***Why?***
Ex 1: How many times do you update your files with the application?	1 = Never 2 = Once a month 3 = 2 to 3 times a month, 4 = Once a week 5 = More than once a week	1 = Never 2 = Rarely 3 = Occasionally 4 = Regularly	Avoid vague quantifiers to obtain more precise estimates. What is meant by 'regularly" can vary for different respondents.
Ex 2: The performance of the application is…	4 = Excellent 3 = Good 2 = Fair 1 = Poor	5 = Very useful 4 = Useful 3 = Functional 2 = Dysfunctional 1 = Very Useless	Beware of semantic weight of word choices. Words like useful and functional have different implications
Ex 3: My impression of the application is…	4 = Very satisfied 3 = Somewhat satisfied 2 = A little satisfied 1 = Not at all satisfied	4= Delighted 3= Mostly satisfied 2= Mostly dissatisfied 1= Terrible	Remain consistent in word choices. Delight, Satisfaction and Terror are not at all the same sort of thing.
Example 4: The application is easy to use. (**AGREE)**	5 = Strongly agree 4 = Somewhat agree 3 = Neither agree nor disagree 2 = Somewhat disagree 1 = Strongly disagree 98 = No opinion	5 = Strongly agree 4 = Somewhat agree 3 = No opinion 2 = Somewhat disagree 1 = Strongly disagree	Not being able to agree or disagree with a statement is not the same as having no opinion. The former is actually an opinion itself, while using "no opinion" as a mid-range category makes it difficult to discriminate whether people truly have no opinion or simply want to response somewhere between "agree" and "disagree."
Example 5: Describe your ability to apply the application after training. **(EXPERT)**	5 = Expert 4 = Perform Independently 3 = Perform with Minimal guidance 2 = Perform with Substantial guidance 1 = Little/ no ability	___ More ___ Same ___ Less	The Expert scale gives you a very concrete sense of how well the respondents can perform a skill. The more or less scale can give a comparison but does not provide you with an exact rating of skill.

Step 7 – Data Structure

Data structuring basically means coding your data and organizing your data into table formats to be used for analysis. This must happen before the survey is sent out, as the questions are written. Because you are using a scale to turn words into numbers, it is crucial that these numbers are easily understandable and easily managed. Put your variables in columns and people in rows. Make sure each cell has only one value. **Think of the results in terms of data that can be collected and easily coded.**

- **Code data intuitively.** For example, assign larger numbers to positive answers and lower numbers to negative answer, e.g., 5 = Excellent…1=Poor.

- **Pre-coding:** You can assign coding values to answers when you construct your survey (pre-coding) or after data collection (post-coding). The advantage of pre-coding is that you have better control of what your returned data will look like, and you know what kind of data you are receiving for analysis.
- **Non-applicable and Don't know:** For "non-applicable," a value of 98, and for "don't know", a value of 99 is assigned so that it is easy to spot. Many programs can only accept numerical data, which means that you must assign a number to "non-applicable" rather than using words or "NA" to identify it. The advantage to coding "non-applicable" with an out-of-range number is that when you do analysis, dubious results can leap at you immediately and force you to check and clean your data manually.
- **Distinguish real missing data from unanswerable data.** Usually, people code missing data with a 0.
- **For hand-entry coding, be sure to check your data frequently while entering.**
- **Make sure that each cell has only one value in any table.** You can ask a question that generates multiple answers. But when you code, you should break the question down into separate items with a Yes/No scale.

Step 8 – Test & Pilot Survey

It is very important that you test your survey before you ever send it to your population. The only thing more irritating than taking a survey is taking it again because of an error. Find errors early and avoid aggravating your population with unnecessary work. The following are some suggestions on how you might go about testing your survey:

- Test it yourself and check your media delivery mechanisms
- Have some colleagues complete the survey to make sure scales and questions are clear, straightforward, and intuitive
- Create a mock report with your data structure. Plot phony averages and percentages for your questions. You may find that some of your questions don't address the survey goal you had in mind. Weed out unnecessary questions. You may also find that there was a question you forgot to ask, which will be crucial to making your decision. Use the mock report to make sure you know what your data will mean to you and how it will influence your decision

Step 9 – Administer Survey

Engaged Participants and Response Rates - The quality of your data depends, in part, on getting more of those selected in the sample to complete your survey. Ideally, participants should feel that they are engaged in a relationship of reciprocity. They will do something for you (e.g., fill out your survey), because you are going to do something for them (e.g., use the results of the survey to make improvements). Response rates will also depend on the ease with which respondents can complete the survey. Do everything you can to get respondents into, through, and out of the survey as quickly and painlessly as possible.

Step 10 – Analysis and Reporting

If you started by clarifying your survey goal at the beginning and designed your survey to answer the questions necessary for your particular decision, your analysis should be clear. You should know from your results what your decision will be even before you have gathered your data.

Analysis Steps

- Data Cleaning - Check for missing data to see if there is a question that is consistently unanswered by checking the Mean, outliers and missing data. Check your data with your general impression obtained from the comments. Also check for duplicate and incomplete surveys

- Descriptive Statistics - To make your analysis strong and convincing, use simple and descriptive graphs and tables to provide simple summaries of the data. You want to create a picture of the data, and do it in a way in which the main point is immediately clear.
    - **Percentage:** is the frequency of individual values for a variable. Summarize your data with a bar chart or a pie chart.
    - **Average/Mean:** is the average of the coding values for your scale. Keep in mind missing data and out-of-range values when you calculate the means. The Mean of a single question or the average of responses to all questions will give you a good sense of the "overall" response of your sample to your survey.
    - **Distribution:** is a summary of the frequency of individual values for a variable. The simplest distribution would list every value of a variable and the number of persons who have each value.
- Slice by Demographics - You need to be aware of how demographics may affect your particular business case (e.g. Age can affect responses received for the best song of all-time). You should also be aware that when you slice your data by demographics, the amount of error in your data changes.
- Check Distributions - When you are performing an analysis of a program, you want to be able to identify specifically those parts of the program that are failing to deliver. For example, if you polled employees to find out if they are satisfied with current training programs, you will get an average score. But, if you broke down those results further by functional department, you may find the Customer Service Reps are highly satisfied while the Sales personnel are highly unsatisfied.
- Non-Response Bias – It is critical to view each situation as unique and determine how different delivery methods will lead to a variance in Non-Response Bias. At times, it will be more appropriate to call rather than send consensus surveys.
- Open-Ended Comments - Analyzing open-ended comments is very difficult and time consuming, so you need to make sure your organizational strategy makes sense in terms of your overall goals and objectives. Consider the following strategies for categorizing comments:
    - **Processes -** The data may be organized to describe important processes (e.g., control, recruitment, decision making, communication)
    - **Issues -** The data may be organized to distinguish key issues, primary evaluation questions, etc.
    - **Key events -** Present the data by critical incidents or major events, in order of importance
    - **People –** If individuals or groups are the primary unit of analysis, then code comments by people or groups
    - **Various settings -** Describe various places, sites, settings or locations
    - **Chronology -** Describe the contents over time, to tell the story from beginning to end
    - **Reporting –** Through the use of Pie Charts, comments can be organized in groups and represented by charts

Step 11 – Make Recommendations

2. Written Tests (Robert Mager: Measuring Instructional Results pages 17 –31)
The goal of creating a written test is to select items that will test achievement and help us to decide whether the instruction actually accomplished what it was intended to accomplish. This means writing or selecting test items that ask students to:

a) Do what the objective asks them to be able to do
b) Under the conditions described by the objective

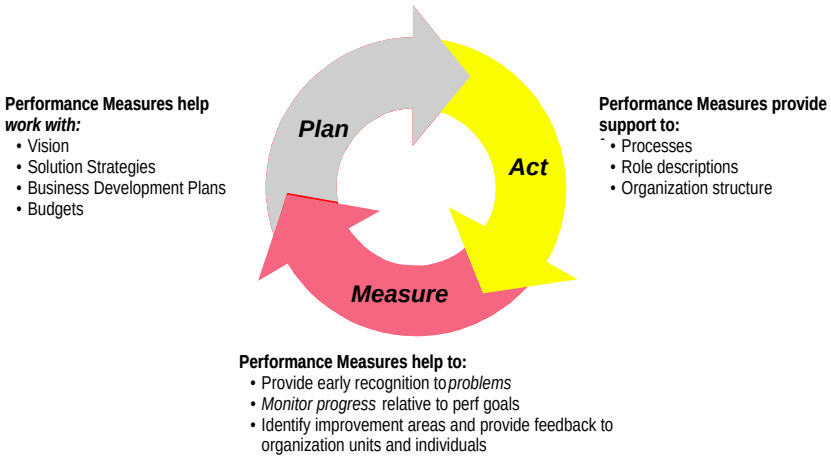
Written & Performance tests can be effective when used concurrently with control groups to measure attitudes, knowledge and/or attitudes before and after a program. A pre-test is given to both the control group and experimental group prior to training. A post-test is given to the same groups following training; however, only the experimental group received training and the test material relates to new material presented during training. Average scores are then calculated from pre-tests and post-test to determine variances. Test scores as a whole help to measure the effectiveness of the instructor in increasing knowledge and/or changing attitudes.

Just as important to the effectiveness of the instructor is the specific information that evaluation of learning provides. By analyzing the change in answers to individual items, the instructor can see where he or she has succeeded and failed. This information is vital when planning future training.

3. Performance Measures

The Role of Performance Measures in the Management Cycle

Performance measures are an integral component in the "Plan, Act, Measure" management cycle. They support the vision, values, core competencies, process and organizational structure.



Where can I use this/what are the benefits?
Performance measurement ensures the alignment of individual actions with organizational goals and objectives. It assigns personal accountability for achieving objectives, and focuses individuals on a balanced set of measures in key performance areas. It reinforces the organizational direction through clearly articulated values and rewards, and creates a shared understanding of the organization and each individual's objectives. It also helps in identifying systemic impediments to successful change and in encouraging ownership of the change program.

There are three broad categories of performance measures:

- Productivity - can track performance changes over time, across units etc.

Examples
(Projects Completed/Projects Attempted)
(Number of Packages Used/Cost of Using the Packages)

- Quality

Example
(Defects/Finished Products)

- Timeliness

Example
(Call Handling Time/Average Call Handling Time)

4. Financial Analysis (Amy Day & Elizabeth Hagerty: Accenture)

Where can I use this/what are the benefits?
Financial Analysis becomes beneficial when trying to compare the monetary return of any value added savings or results with training development costs.

Examples

- Benefits to Cost Ratio (Cost/Benefit Analysis):

$$BCR = \frac{\text{Benefits}}{\text{Total Costs}} = \frac{\$467,000}{\$79,000} = \$5.9$$

For every dollar invested a benefit of $5.9 is returned

- Return of Investment:

$$ROI = \frac{\text{Benefits - Total Costs}}{\text{Total Costs}} = \frac{479,000 - 79,000}{79,000} = 4.9$$

For every dollar invested there was a return of 4.9 in net benefits.

Drawbacks
•Does not capture all of the company's strategic objectives
•ROI is a snapshot in time
•Doesn't adequately address the business complexities
•Difficult to establish causal relationship
•When applied to training evaluation, the ROI of the ROI is low

5. Balanced Scorecard
Step-by-step instructions for building a balanced scorecard:

- Define the vision.
  - The vision must be inclusive enough to provide guidance for each of the four levels
  - An effective vision must be desirable for the long-term interests of the stakeholders, feasible, focused enough to guide decision-making, flexible, and easy to communicate
- Define the critical success factors
  - Areas that will drive performance to the vision
- Identify the specific critical measurements
  - Financial examples: net income, return on net assets, cash flow, etc.
  - Customer examples: percent of repeat customers, percentage of customer complaints, number of new accounts, etc.

- Internal examples: product development time, engineering change notifications, scrap and rework, customer response time, etc.
- Innovation and Learning examples: total process cycle time, number of new products developed, employee satisfaction, etc.

The balanced scorecard is not only used at the organizational level, but also at a more specific level - such as for a particular training course or performance solution

- Collect the department's vision and mission statements
- Ask for any information on strategic initiatives that are in place
    - Ask for critical success factors and measures in all four areas
- Agree upon a set of measures with the customer
    - Measures selected will work best when:
        - You have control over what is measured
        - They are easily quantified
        - Everyone in the customer organization can be involved in the measurement
        - They are measured as close to the front line - the people who do the work - as possible
        - They are viewed as an improvement tool, not as punishment

Conclusion

What are some current trends?

- Consumer demand for innovation, customization, and higher service levels requires continuous training and retraining
- Downsizing forces streamlined workforce to know and do more with current or fewer training dollars
- Increased competition and downward price pressure have heightened the demand for cost-effective training programs

With the current trends described above, there is an increased emphasis on measuring the economic impact of training. Consequently, training dollars are often first to go, because they are considered less critical to the well being of the organization. However, if evaluation methods are in place, then effectiveness can be measured and the results can prove training is relevant and crucial, and can rally credibility for training even from those who are most reluctant.

To help achieve this success, evaluation needs to be integrated with training processes to avoid simply providing a grade on performance at the conclusion of the event. The idea is to move away from the perspective on evaluation as an isolated measurement event conducted at the conclusion of an educational program but to thread evaluation processes throughout the training process.

In the end, evaluating training programs effectively will help you realize the validation of your design and training investment, and build credibility for future training endeavors.