Course: Test Development and Evaluation (6462)
Semester: Autumn, 2020
ASSIGNMENT No. 2

Q.1 Define classroom projects. What is a difference between task and test?

The following are definitions of testing, assessment, and evaluation. In spite of important differences between these terms, they are often used interchangeably by teachers.

Testing, assessment, and evaluation

The following sections attempt to define testing, assessment, and evaluation.

Definition of evaluation

The verb evaluate means to form an idea of something or to give a judgment about something. The term comes from the French word 'évaluer', meaning "to find the value of". The origin is from the Latin term 'valere' meaning "be strong, be well; be of value, or be worth".

In the educational context, the verb 'to evaluate' often collocates with terms such as:

- the effectiveness of an educational system,
- a program,
- a course,
- instruction,
- a curriculum.

According to Weiss (1972), evaluation refers to the systematic gathering of information for the purpose of making decisions. It is not concerned with the assessment of the performance of an individual, but rather with forming an idea of the curriculum and making a judgment about it. This judgment is made based on some kind of criteria and evidence. The purpose is to make decisions about the worth of instruction, a course, or even the whole curriculum. Evaluation is thus larger and may include an analysis of all the aspects of the educational system.

Definition of assessment

The verb assess comes from the French 'assesser', but the origin is from the Medieval Latin 'assessare' meaning "fix a tax upon,". Another derivation of the Latin term is 'assidere' or 'adsidere' meaning "to sit beside" (a judge). Reference is made to the assistant of the judge whose job was to fix the amount of a fine or tax by estimating the value of a property.

Assessment is thus the process of collecting information about students from diverse sources so that educators can form an idea of what they know and can do with this knowledge. While evaluation is concerned with making judgments about instruction, a curriculum, or an educational system, assessment is concerned with the students' performance. In other words, one assesses an individual but evaluates a program, a curriculum, an educational system, etc.

The verb ‘assess’ often collocates with:

- skills,
- abilities,
- performance,
- aptitude,
- competence.

According to Le Grange & Reddy, (1998, p.3)

Assessment occurs when judgments are made about a learner’s performance, and entails gathering and organizing information about learners in order to make decisions and judgments about their learning.”

Assessment is thus the process of collecting information about learners using different methods or tools (e.g. tests, quizzes, portfolios, etc).

Educators assess their students for a variety of purposes:

- To evaluate learners’ educational needs,
- To diagnose students’ academic readiness,
- To measure their progress in a course,
- To measure skill acquisition.

There are different types of assessment:

- Formative assessment:

  It is process-oriented and is also referred to as ‘assessment for Learning’. It is an ongoing process to monitor learning, the aim of which is to provide feedback to improve teachers instruction methods and improve students learning.

- Summative assessment:

  It is product-oriented and is often referred to as ‘Assessment of Learning’. It is used to measure student learning progress and achievement at the end of a specific instructional period.

- Alternative assessment:

  It is also referred to as authentic or performance assessment. It is an alternative to traditional assessment that relies only on standardized tests and exams. It requires students to do tasks such as presentations, case studies, portfolios, simulations, reports, etc. Instead of measuring what students know, alternative assessment focuses on what students can do with this knowledge.

Definition of testing

Simply put, a test refers to a tool, technique or a method that is intended to measure students knowledge or their ability to complete a particular task. In this sense, testing can be considered as a form of assessment. Tests should meet some basic requirements, such as validity and reliability.

- Validity refers to the extent to which a test measures what it is supposed to measure.
- Reliability refers to the consistency of test scores when administered on different occasions.

There are different types of tests:

- Placement tests: It is designed to help educators place a student into a particular level or section of a language curriculum or school
- Diagnostic tests: they help teachers and learners to identify strengths and weaknesses.
- Proficiency tests: they measure a learner's level of language.
- Achievement tests: they are intended to measure the skills and knowledge learned after some kind of instruction.

Q.2 Write different types of interviews and also explain their strength and weakness.

During the interview you will be asked some difficult questions. These questions are designed to see how you will react under pressure and unfamiliar territory. However, by being familiar with the types of questions you may be asked and identifying examples of situations from your experiences where you have demonstrated behaviors or skills, and practicing responses to questions, you will give yourself the opportunity to answer these questions comfortably and confidently. While we describe these interviews separately, your interview may be a combination of these interview types.

Traditional Interview

Although behavioral interviewing (see section below) is being used more frequently, a traditional interview is still very common. The length is usually from a half hour to one and a half hours. You are usually asked many of the questions from the Common Interview Questions section and possibly a few behavioral questions related to the job. After this interview, if you are being considered, you would be invited back for 1-2 more interviews.

Panel Interview

In a panel interview, you will be in the room with several people, usually three to five, but it could be as many as ten. Sometimes each person will ask you a series of questions, and other times a few people will ask questions while the rest listen. The purpose of having many people sit in on the interview is sometimes so several staff members can meet the candidates and also to keep the number of interviews a candidate has limited to one or two. The questions could be traditional, behavioral or a combination.

Serial Interview

This type of interview consists of a series of interviews on the same day. For example, if you are scheduled for an interview from 1-3:30, you may have 5 half-hour interviews with individual people. Each interviewer will most likely ask you different questions. Question types may vary. If your interview is during a mealtime, like 10-2, most likely, someone will take you to lunch. Do not let down your guard; this is still part of the interview. A "lunch interview" (see section below) is more common for upper-level positions or positions in which you would have frequent "lunch meetings" with clients/colleagues.

Behavioral Interviews

Behavioral interviews are designed to provide the interviewer with information about how you think and behave in situations. Hiring managers are interested in your method of thinking and going about doing a task as well as how you respond in difficult situations. An effective way to answer behavioral questions is the to provide the interviewer with four pieces of information known as the STAR Method: Situation, Task, Action, Results.

Case Interview

This is really a type of question. Your whole interview could be made up of case interview questions, or you could be asked 1-2 case questions. You will be given a problem or scenario related to the industry and asked to solve it or describe the process in detail. The interviewer is not necessarily looking for the "right" answer, but wants to evaluate your problem solving skills. If you do not have adequate information to determine a solution, you may need to ask the interviewers questions. Case interviews are more commonly used by investment banks, consulting firms, technical positions, and for upper-level management positions. You may be asked to provide a detailed financial analysis, to plan an event or to develop a marketing campaign.

Examples of Case Interview Questions:

- A chain of grocery stores currently receives its stock on a decentralized basis. Each store deals independently with its suppliers. The president of the chain is wondering whether the company can benefit from a centralized warehouse. What are the key considerations in making this decision?
- A magazine publisher is trying to decide how many magazines she should deliver to each individual distribution outlet in order to maximize profits. She h as extensive historical sales volume data for each of the outlets. How should she determine delivery quantities?
- A personal finance and loan company that currently targets low-middle class people is expanding to the Philadelphia market. They need to know how many financial centers the area can support and where they should be located. What kind of information would you need in order to determine the number of centers?

Phone/Skype Interview

To save the company time and money, many interviewers will schedule phone interviews as an initial screening of potential candidates. This may be an interview with a human resources representative or a hiring manager. Since this is typically your first interaction with the company, it is also their first impression or you.

- Set up a specific time for the call so you will be fully prepared.

- Treat the phone interview just as you would a face-to-face interview. Just because it is over the phone does not mean that you should take it less seriously!
- Have your resume, company research, notes, and questions to ask spread out for the interview. Make sure your information is well-organized and legible. You certainly don't want to noisily go rummaging through papers on the desk while you say, 'ummmmm' because you are trying to find your list of questions.
- Make sure the phone/computer you are using works well and avoid cell and portable phones if possible.
- If you have call-waiting, turn it off if possible. At least, DO NOT answer it during an interview.
- If you have a bad connection or have difficulty hearing the interviewer, offer to call them back.
- Turn off radio and television to limit distractions. Plan to sit in a quiet room where you won't get interrupted.
- Put a 'Do Not Disturb' sign on your door.
- Get yourself water beforehand.
- Stand up during the interview (for phone interview). You will sound more expressive and alert.
- Smile. Your enthusiasm and happiness will show through in your voice.
- Establishing rapport is an important aspect of interviewing. On the phone, one way to establish rapport is to "match" the interviewer's tone of voice and rate of speech. This will make you seem "like" him or her.
- Before you get off the phone, ask about a time to meet face-to-face. Do not be afraid to ask! The worst thing that can happen is they tell you 'no' or say they will get back to you. Have your calendar ready so you can schedule and interview easily.
- Employers will be calling you to set up interviews. If you have an answering machine or voice mail on your home and cell phone, make sure it is professional.

Lunch Interview

Job interviewers take candidates to lunch to evaluate your business etiquette, communication and interpersonal skills, social skills and table manners. DO NOT let your guard down. This is no time to act like you would with your college buddies. This is still a part of your interview. Following are tips to help you while dining.

- If you are a picky eater, check out the restaurant ahead of time. That way you'll know exactly what's on the menu, what you might want to order and where the rest rooms are located.
- Make sure your cell phone is turned off
- Don't order messy food (pasta with lots of sauce, chicken with bones, ribs, big sandwiches, and lobster).
- Don't order the most expensive entree on the menu. Opt for moderate priced items like chicken or fish.
- Avoid foods that have a strong or unpleasant order (garlic and onions).
- Consider staying away from things that are crunchy. If may affect how well you hear someone.
- Do order food that is easy to cut into bite-size pieces.
- If you can, avoid soup. It is harder to eat than you think.
- Eat your entire meal if possible. It may seem disrespectful if you leave a large amount of food behind.
- Consider water. Most other beverages stain if spilled.

- Interviewing is hard enough without adding alcohol to the mix.
- Be polite. Remember to say “please” and “thank you” to your server as well as to your host.
- During conversation, remember to listen and participate.
- Stay away from sensitive subjects like religion and politics.
- Do not talk with your mouth full.
- Do not complain about the food.
- Keep your elbows off the table, sit up straight and don’t talk with your mouth full.
- You are not expected to pay for your meal or tip when invited to lunch. It would be inappropriate to offer.NEVER ask for a doggy bag.
- Remember to say thank you and tell the interviewer you enjoyed the meal.

Assessing your weaknesses

Let's get the hard part out of the way first—your weaknesses. This is probably the most dreaded part of the question. Everyone has weaknesses, but who wants to admit to them, especially in an interview?

Some examples of weaknesses you might mention include:

- Being too critical of yourself
- Attempting to please everyone
- Being unfamiliar with the latest software

The best way to handle this question is to minimize the trait and emphasize the positive. Select a trait and come up with a solution to overcome your weakness. Stay away from personal qualities and concentrate more on professional traits. For example: "I pride myself on being a 'big-picture' guy. I have to admit I sometimes miss small details, but I always make sure I have someone who is detail-oriented on my team."

Assessing your strengths

When it comes time to toot your own horn, you need to be specific. Assess your skills to identify your strengths. This is an exercise worth doing before any interview. Make a list of your skills, dividing them into three categories:

- Knowledge-based skills: Acquired from education and experience (e.g., computer skills, languages, degrees, training and technical ability).
- Transferable skills: Your portable skills that you take from job to job (e.g., communication and people skills, analytical problem solving and planning skills)
- Personal traits: Your unique qualities (e.g., dependable, flexible, friendly, hard working, expressive, formal, punctual and being a team player).

Some examples of strengths you might mention include:

- Enthusiasm
- Trustworthiness
- Creativity

- Discipline
- Patience
- Respectfulness
- Determination
- Dedication
- Honesty
- Versatility

When you complete this list, choose three to five of those strengths that match what the employer is seeking in the job posting. Make sure you can give specific examples to demonstrate why you say that is your strength if probed further.

Q.3 what is a role of invigilator? Describe his/her duties in detail.

Examination invigilators are an integral part of the exams team in every school and college. The role of the examination invigilator is to ensure that examinations are conducted according to the current JCQ instructions for conducting examinations.

The invigilator is the person in the examination room responsible for conducting a particular examination session in the presence of the candidates. Invigilators have a key role in upholding the integrity of the external examination/assessment process.

What's involved

The JCQ details the invigilator's role as:

- ensuring all candidates have an equal opportunity to demonstrate their abilities
- ensuring the security of the examination before, during and after the examination
- preventing possible candidate malpractice
- preventing possible administrative failures
- liaising with the examination officer and senior invigilator

JCQ regulations also state that invigilators must:

- be familiar with JCQ's 'Instructions for Conducting Examinations', the 'Warning to Candidates', Information for Candidates' and 'Mobile Phone' posters, and any specific instructions relating to the subjects being examined
- give all their attention to conducting the examination properly
- be able to observe each candidate in the examination room at all times
- inform the head of centre if they are suspicious of the security of the examination papers

Invigilators **must not** carry out any other task (for example, reading a book or marking) in the examination room.

Skills and experience you will need

- good communication skills

- good IT and administrative skills
- good organisational skills
- able to work as a member of a team
- an ability to adhere to rules and regulations
- to think clearly and make correct decisions in the event of an emergency in the exam room

Entry requirements

The Joint Council for Qualifications (JCQ) sets out in their 'Instructions for Conducting Examinations' booklet the requirement that 'a training session must be held for any new invigilators'. Therefore, schools and colleges decide what qualifications (if any) and experience applicants need to have, so it is important to check locally. Some may employ people based on their experience and potential alone.

Q.4 what is classical test theory? Explain.

Classical Test Theory (CTT), sometimes called the true score model, is the mathematics behind creating and answering tests and measurement scales. The goal of CTT is to improve tests, particularly the reliability and validity of tests.

Classical test theory is a collection of many statistics, including the average score, item difficulty, and the test's reliability.

1.Correlation

Correlation: shows how two variables X and Y are related to each other. Different measures are used for different test types. For example, a dichotomously scored test (e.g. yes/no answers) would be correlated with point-biserial correlation while a polytomously scored test (one with multiple answers) would be scored with the Pearson Correlation Coefficient.

Classical test theory (CTT) is an approach to measurement that considers the relationship between the expected score (or "true" score) and observed score on any given assessment. The word classical is used in the sense that the theory is considered to be the first practical application of mathematics to describe this relationship. CTT offers a relatively parsimonious, elegant, and intuitive way to scale individuals according to some theorized latent construct. This entry further describes CTT and its basic principles and estimation procedures, then discusses its framework for determining a measure's proportion of true score variance, standard error of measurement, item analysis, and validity. Finally, it looks at the limitations to the theory.

2. Covariance

Covariance is a measure of how much two random variables vary together. It's similar to variance, but where variance tells you how a single variable varies, co variance tells you how two variables vary together.

3. Discrimination Index

Discrimination Index: the ability of the test to discriminate between different levels of learning or other concept of interest. A high discrimination index indicates the test is able to differentiate between levels.

4. Item difficulty

Item difficulty: a measure of individual test question difficulty. It is the proportion of test takers who answered correctly out of the total number of test takers. For example, an item difficulty score of 89/100 means that out of 100 people, 89 answered correctly.

5. Reliability Coefficient

Reliability coefficient — a measure of how well the test measures achievement. Several methods exist for calculating the coefficient include test-retest, parallel or alternate-form and internal analysis. Rules of thumb for preferred levels of the coefficient:

For high stakes tests (e.g. college admissions), > 0.85.

For low stakes tests (e.g. classroom assessment), > 0.70.

6. Sample Variance / Standard Deviation

The sample variance and sample standard deviation are measures of how spread out the scores are.

7. Standard Error of Measurement

Standard Error of Measurement (SEm): a measure of how much measured test scores are spread around a "true" score.

Q.5 Write a comprehensive note on Kilpatrick four level Model.

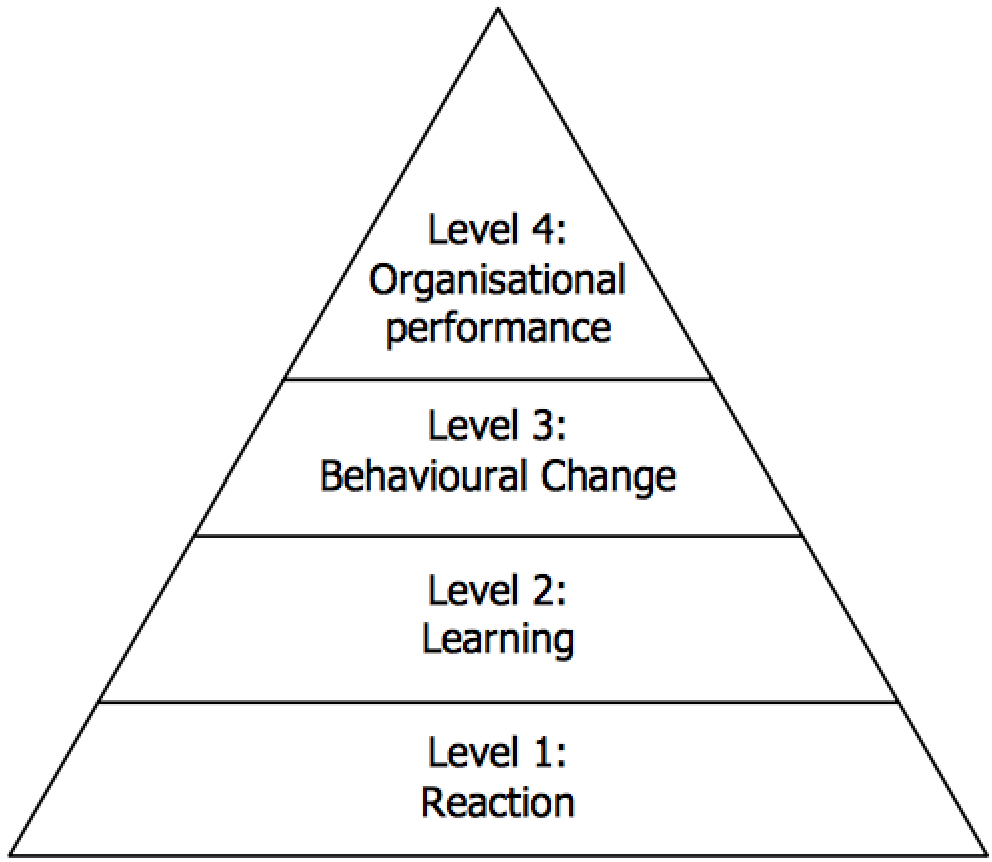
The Kirkpatrick Model is probably the best known model for analyzing and evaluating the results of training and educational programs. It takes into account any style of training, both informal or formal, to determine aptitude based on four levels criteria.

Level 1 Reaction measures how participants react to the training (e.g., satisfaction?). **Level 2 Learning** analyzes if they truly understood the training (e.g., increase in knowledge, skills or experience?). **Level 3 Behavior** looks at if they are utilizing what they learned at work (e.g., change in behaviors?), and **Level 4 Results** determines if the material had a positive impact on the business / organization.

This model was developed by Dr. Donald Kirkpatrick (1924 – 2014) in the 1950s. The model can be implemented before, throughout, and following training to show the value of training to the business.

As outlined by this system, evaluation needs to start with level one, after which as time and resources will allow, should proceed in order through levels two, three, and four. Data from all of the previous levels can be used as a foundation for the following levels' analysis. As a result, each subsequent level provides an even more accurate measurement of the usefulness of the training course, yet simultaneously calls for a significantly more time-consuming and demanding evaluation.

Undoubtedly, the most widely used and in-demand method for the assessment of training in businesses nowadays is Kirkpatrick's system based around the four levels as guidelines. The Kirkpatrick model has been used for over 30 years by many different types of companies as the major system for training evaluations. It is evident that Kirkpatrick's vision has made a positive impact to the overall practice of training evaluation.



Source: from Kirkpatrick, 1996

Listed below is an in-depth look into the four levels of the Kirkpatrick Model:

Level 1 Evaluation – Reaction

The objective for this level is straightforward, it evaluates how individuals react to the training model by asking questions that establishes the trainees' thoughts. Questions will figure out if the participant enjoyed their experience and if they found the material in the program useful for their work. This particular form of evaluation is typically referred to as a "smile sheet."

As outlined by Kirkpatrick, each program needs to be assessed at this level to help improve the model for future use. On top of that, the participants' responses is essential for determining how invested they will be in learning the next level. Even though an optimistic reaction does not ensure learning, an unfavorable one definitely makes it less likely that the user will pay attention to the training.

Examples of resources and techniques for level one:

- Online assessment that can be graded by delegates/evaluators.
- Interviews
- Can be done immediately after the training ends.
- Are the participants happy with the instructor(s)?
- Did the training meet the participant's needs?
- Are the attendee's happy with the educational tools employed (e.g., PowerPoint, handouts etc)
- Printed or oral reports provided by delegates/evaluators to supervisors at the participants' organizations.
- "Smile sheets".
- Comment forms determined by subjective individual reaction to the training course.
- Post-training program questionnaires.
- Verbal responses that can be taken into consideration and considered.
- Especially encourage written comments
- Try to get honest responses and feedbacks

Level 2 Evaluation – Learning

Evaluating at this level is meant to gauge the level participants have developed in expertise, knowledge, or mindset. Exploration at this level is far more challenging and time-consuming compared to level one.

Techniques vary from informal to formal tests and self-assessment to team assessment. If at all possible, individuals take the test or evaluation prior to the training (pre-test) and following training (post-test) to figure out how much the participant comprehended.

- Measurement and evaluation is simple and straightforward for any group size.
- You may use a control group to compare.
- Exams, interviews or assessments prior to and immediately after the training.
- Observations by peers and instructors
- Strategies for assessment should be relevant to the goals of the training program.

- A distinct clear scoring process needs to be determined in order to reduce the possibility of inconsistent evaluation reports.
- Interview, printed, or electronic type examinations can be carried out.
- An interview can be carried out before and after the assessment, though this is time-consuming and unreliable.

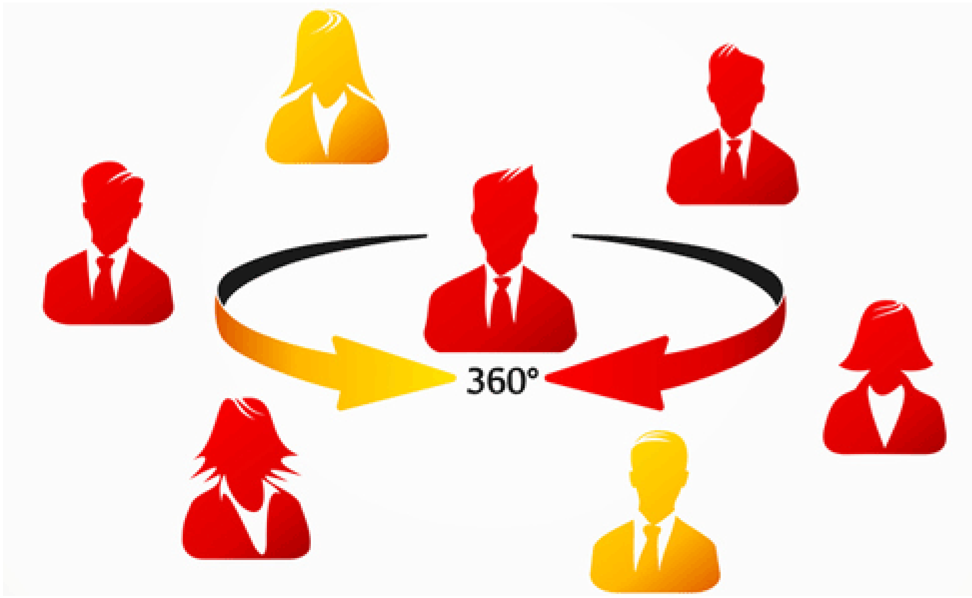
Level 3 Evaluation – Transfer

This level analyzes the differences in the participant's behavior at work after completing the program. Assessing the change makes it possible to figure out if the knowledge, mindset, or skills the program taught are being used the workplace.

For the majority of individuals this level offers the truest evaluation of a program's usefulness. Having said that, testing at this level is challenging since it is generally impossible to anticipate when a person will start to properly utilize what they've learned from the program, making it more difficult to determine when, how often, and exactly how to evaluate a participant post-assessment.

Examples of assessment resources and techniques for level three:

- This can be carried out through observations and interviews.
- Evaluations have to be subtle until change is noticeable, after which a more thorough examination tool can be used.
- Were the learned knowledge and gained skills used?
- Surveys and close observation after some time are necessary to evaluate significant change, importance of change, and how long this change will last.
- Online evaluations tend to be more challenging to integrate. Examinations are usually more successful when incorporated within present management and training methods at the participant's workplace.
- Quick examinations done immediately following the program are not going to be reliable since individuals change in various ways at different times.
- 360-degree feedback is a tool that many businesses use, but is not necessary before starting the training program. It is much better utilized after training since participants will be able to figure out on their own what they need to do different. After changes have been observed over time then the individual's performance can be reviewed by others for proper assessment.



360-degree feedback

- Assessments can be developed around applicable scenarios and distinct key efficiency indicators or requirements relevant to the participant's job.
- Observations should be made to minimize opinion-based views of the interviewer as this factor is far too variable, which can affect consistency and dependability of assessments.
- Taking into consideration the opinion of the participant can also be too variable of a factor as it makes evaluation very unreliable, so it is essential that assessments focus more defined factors such as results at work rather than opinions.
- Self-assessment can be handy, but only with an extensively designed set of guidelines.

Level 4 Evaluation – Results

Commonly regarded as the primary goal of the program, level four determines the overall success of the training model by measuring factors such as lowered spending, higher returns on investments, improved quality of products, less accidents in the workplace, more efficient production times, and a higher quantity of sales.

From a business standpoint, the factors above are the main reason for the model, even so level four results are not usually considered. Figuring out whether or not the results of the training program can be linked to better finances is hard to accurately determine.

Types of assessment strategies and tools used for level four:

- It should be discussed with the participant exactly what is going to be measured throughout and after the training program so that they know what to expect and to fully grasp what is being assessed.
- Use a control group
- Allow enough time to measure / evaluate

- No final results can be found unless a positive change takes place.
- Improper observations and the inability to make a connection with training input type will make it harder to see how the training program has made a difference in the workplace.
- The process is to determine which methods and how these procedures are relevant to the participant's feedback.
- For senior individuals in particular, yearly evaluations and regular arrangements of key business targets are essential in order to accurately evaluate business results that are because of the training program.