
TRAINING EVALUATION – APPLICATION BOTTLENECKS

Training evaluation has a chequered history centered around Donald Kirkpatrick's four level evaluation proposed in 1959. After 40 years, level 5, Return On Investment (ROI), proposed by Jack Phillips has gained wide currency. Workers like Bakken and Bernsteinin, Jack Phillips, Dan Spitzer, Greg Wang, Donald Clark and many others have added meaningful volumes to T & D measurement movement. Though Kirkpatrick's model is simple to understand, in practice we find the need to answer many haunting questions. Answering these questions are critical for us, T & D professionals, as we have to demonstrate value to stake-holders.

Key Words: 1. Training evaluation 2. Donald Kirkpatrick 3. Level five, return on investment.

When Donald L. Kirkpatrick published his set of training evaluation ideas in 1959, in a series of articles in the US Training and Development Journal, it's reasonable to assume he had no idea this would become a mile stone in training measurement. In 1975, in his book "Evaluating T & D" published by ASTD, all these ideas were explained and discussed in depth. His 4 level training evaluation model became a standard in training circles; a starting point for any researcher or student in training evaluation. In 1998, Kirkpatrick updated his ideas in another book "Evaluating Training Programmes : the 4 levels."

The debate that Kirkpatrick's early article described a "taxonomy or a classification scheme of evaluating training" (Alliger & Janak, 1989, Holton, 1996) will continue. The model provides a communication tool for practitioners to communicate their evaluation efforts. In reality, the ASTD handbook by Kirkpatrick gives many guidelines, formats and specific ways to evaluate each training level and so elevates his four-level model much beyond and above a classification system. He classified training evaluation into 4 levels.

Level One - Reaction.: Kirkpatrick^{1b} describes this as “how participants feel about the various aspects of the training programme”. Does the learning experience generate positive feelings about the facilitator, the instructional design materials and was the whole experience happy? The purpose of the learning experience should be to generate motivation to learn and enjoy.

In practice, we have found that whenever our learners share positive experience with prospective learners they are motivated to attend the programme and the enrollments for subsequent programmes increase. The feedback formats taken at the end of the programmes has given us valuable insights into what the learners valued and what is not; our faculty members scramble for these sheets after every programme. Those who receive low grades from the trainees receive quick feedbacks from seniors and helps improve effectiveness of our T & D interventions. We agree with Kirkpatrick that these “smile-sheets” are customer satisfaction metrics and gives us valuable inputs to improve instructional design and is a credible indicator of level one effectiveness.

Level Two - Learning: Kirkpatrick describes this level as “knowledge acquired, skills improved, or attitudes changed due to training”. He feels that majority of the training will give some tangible gains in this area. Practitioners need to answer dominant questions here; as per the training objectives, did the learner learn anything demonstrably better; can he perform new / old tasks in improved and more effective ways?

In practice we find that at the entry level, before every session, if we ask the learners to write what is their concept of the topic and subsequently allow them to compare their notes at the end of session, the learners realize the distance they have traveled due to the learning intervention. Our experience with tests are very satisfactory; a quick, short and objective type test helps anchor learnings. The learners themselves recommend that we do this more often. Presentation of learning concepts also give us insights into the extent of the gains; this is limited by the number of participants and other variables. If time is a constraint and if there are a large number of participants, containing the participants interest throughout the presentations and to measure each

participant against any bench mark is a challenge. Of all the assessment methods, role-plays and case study are the most popular and effective to measure level two effectiveness; to a large extent this simulates their work-environment, helps them deploy the knowledge and skills they have just acquired and helps them to experiment in a relatively risk-free atmosphere.

Level Three – Behavior : Kirkpatrick explained it as “measure of the extent to which participants change their on-the-job behavior because of training”. Here is where the transfer of training takes place, where the rubber hits the road. Kirkpatrick mentioned the challenge of trainees shining in training and subsequently on-the-job retreating back to old behaviors. We agree with Kirkpatrick whole-heartedly, transfer of learning and skills inside the class-room is no guarantee of transfer on the work place. When the trainee hits his natural environment a whole series of variables influence his application. It is critical that the environment re-enforces and supports application of learning inputs and facilitates a fear-free atmosphere. If we involve the immediate supervisor from the training need analysis, then we find that they are more inclined to participate in the implementation of learning concepts. We also find that managers who are exposed to regular training in the past have more faith in T & D interventions. They also give more enthusiastic feedbacks to the T & D team and seem to view training interventions more seriously.

Level Four - Results ; Kirkpatrick indicated this as the last level in measurement and included items like reduced costs, higher quality, improved productivity and lowered absenteeism and turnover. In practice, we find that level four is where the water gets murky. It's a snake-pit where the hissing starts, everybody wants their pound of flesh. Marketing, finance, logistics, manufacturing, HRD; everyone demands recognition for their contribution in improved business metrics. We also find that measuring impact of trainee's new behavior on organizational performance is difficult. If we have to measure the benefit to the organization which can be apportioned to training, we need to plan much before the intervention, involve all the stake-holders in deciding the metrics to be measured, methodologies of delivery and fixing tasks. In a dynamic business environment where the business needs change and demand quick turn-arounds in

developing and delivering impactful training interventions, planning early becomes idealistic.

KIRKPATRIC'S CLASSIFICATION – APPLIED

| LEVEL | MEASUREMENT FOCUS | MEASUREMENT AREAS |
|--------------|--|---|
| 1 - REACTION | Participants satisfaction with the programme. How did they like it? | End of the programme Evaluation sheets (Smile Sheets) |
| 2 – LEARNING | Additions in knowledge, skills and attitude because of learning (Class room Transfer of KSA)) What did they learn? | Pre-programme evaluation tests & post programme evaluation, Simulations, Role-Plays, Case studies & presentations |
| 3 – BEHAVIOR | Changes in the on-the-job behavior (Shop-floor transfer of KSA). What did they use? | Periodic reports from supervisors, Interviews with line managers, surveys on application oriented behaviors. |
| 4 – IMPACT | Changes in the business metrics as a results of changes in behavior due to learning. How did it affect the business? | Sales, activity indices, market shares, cycle times, profitability, customer satisfaction, shareholder values. |

Level 5 - Return On Investment (ROI). ROI, an accounting formula was used in the early 1900 by Du Pont Powder Company; this was complex and took the entire enterprise into account. Compared to present day ROI calculations this was more complicated; at best it helped to explain the relationships with sales, profits and other assets. There was no effort to identify ROI for different functions in an organization. The fifth level of training evaluation was proposed by Jack Phillips and found wide acceptance in learning circles; ROI consulting and practice has grown to become a hugely successful enterprise. ROI aims to calculate if the financial returns exceed the cost of training. Mathematically expressed, ROI is net training benefits divided by total training cost. It is a ratio of net benefits to costs, expressed in a percentage.

ROI exponents are emphatic there is a real need to isolate training's effect on performance. It is not enough to link learning to business results, we need to convert the training impact into monetary value in comparison to the cost of training. ROI gives balanced figures to isolate training results, gives a step-by-step process and

offers a flexible solution to all types of programmes. ROI also acts as a bridge between academic research work with practical applicability. ROI ensures that training demonstrates contribution to business like other functions. ROI analysis also lends credence to successful programmes and helps weeds out the inefficient programmes. An analysis validates further investments in programmes that have higher RIO and supports the training professional in his investment decisions. It's probable that senior management and decision makers will support and respect the training function more if ROI metrics are available and when comparison with other functions can be made. Based on the metrics, senior management also can substantiate their investment logics measurably to all stake-holders; they would be able to make more informed decisions and support training visibly. Jack Phillips⁶ advocates several ways to isolate training effect on performance.

A credible approach to differentiate training effect is the use of control groups in pilot design. In the pilot, the experimental group receives training; the control groups is not given the training. It is important to ensure that both groups are homogenous in nature; they should be similar in demography, randomization and from the same environment. A performance comparison of both groups after the training will isolate the impact of the training. Use of pilot group is especially important when timelines and resources are tight and the training intervention is expensive. Based on the outcome of the control group, it would be possible to take investment decisions on a broader roll-out or for an organization-wide intervention. Positive ROI from regional settings will give the green signal for a national level.

Another approach to ROI is the forecasting method. This is more mathematical and analytical in nature. A linear equation is used to calculate the value of the proposed improvements. This is possible only if one variable affects the results; where multiple variables are involved, complicated statistical models are necessary. This model is dependant on correct and relevant data being available to calculate the impacts.

Participant's feedback is another reliable model to calculate training impact. Participants are asked to measure the amount of performance improvement due to the training programme. Since they are the primary stake-holders in the intervention, they should have an idea of the extent of application of the training inputs and the measure

of change that has brought in the results. This is a valuable and credible tool since the participants themselves were the vehicles of learning delivery and the training intervention. They may be asked the following questions to get reliable answers.

1. Has the application of knowledge and skills brought about any improvement in your performance, if yes, how much?
2. What metrics supports your beliefs?
3. What is the % of confidence on this estimate?
4. Who else can make these estimates?
5. What other variables / inputs have contributed to this improvement in performance?

Jack Phillips recommends that we need to factor in a confidence level. If a learner estimates that 50% improvement is because of training and is only 70% confident about the estimate, we need to multiply the confidence percentage by the improvement of percentage and divide by 100; that gives us 35%. Then we need to multiply that figure by amount of improvement to isolate the portion connected to training. If we convert that portion to a monetary value, we will have the ROI. ROI workers need to appreciate the fact that it is an estimate; all estimates can be challenged. The competence of the participant to make an estimate and their attitude to estimations can affect this process. This method is inexpensive, fast and easy to understand by all stake-holders.

Supervisor estimations can be added on to the participant's evaluations. Superior officers have more insight into variables that have given additional boost to outcomes. The same questions given to participants need to be given to superiors also in order to give uniform inputs and to get reliable results. This has all the advantages and disadvantages as taking estimates from participants themselves. However, when a combination of participant's and their superior officer's inputs are combined together, the credibility of data increases.

Management estimate is yet another tool to give estimates of impact due to training. There is an element of subjectivity to this and is often debatable. However, since

management has funded the venture in the first place they would be more liable to view the exercise seriously and hence this estimate is also another option.

Customer inputs, customer satisfaction scores are yet another metric to isolate impact of training. This would also be a logical step from the fact that training programmes are many times designed to improve the customer interface and outcomes.

Expert estimates are also used for the purpose of isolating training impact and calculations. Third-party estimations, subject experts and industry specialists can also be used for estimations.

Subordinate estimates are another source of isolating measurables. They can be given inputs on the training received by their bosses and asked to assess what and how much has changed after the training. Subordinates are usually aware of changed behaviors, improved efficiencies of their bosses and the measurable the changes this has on the outputs.

Multiple approaches combined with a conservative estimate lend more accuracy and credibility to the ROI calculation.

ROI professional can use hard data to substantiate their efforts. On the output front, units produced, units assembled or sold, forms processed or tasks completed are some examples. On the quality front, defects, rejections, scrap, waste and rework can provide solid metrics. On the time front, equipment downtime, employee overtime, time to complete projects and training time can give more measurables. On the cost front, overheads, variable costs, accident costs and sales expenses will give hard data..

Soft data also will give some handles to the ROI professional. Where work habits are concerned, absenteeism, visits to hospitals / dispensaries, carelessness and safety rule violations will give clear indicators. On the work climate front, employee grievances, employee turnover, discrimination cases and job satisfaction indices are indicators of measurable outcomes. Attitudes can be gauged by loyalty and changes in performance. New skills can be measured through decisions made, problems solved, conflicts avoided / resolved and frequency of use of new skills. Development and advancement can be measured through number of promotions or pay increases,

number of training programmes attended, requests for promotions and performance appraisal ratings.

While doing ROI analysis, it's ideal to stick to some guiding principles:

1. When a higher level evaluation is conducted, data must be collected at lower levels. When you are at level 4 of evaluation, you need level 3 data on how a learner has used his new skills and knowledge on the job. This would result from the fact that learner has learned new skills and knowledge which could be measured at level 2. It is recommended that data is made available for all lower levels when a higher level evaluation is mandated.
2. When an evaluation is planned for a higher level, the previous level of evaluation does not have to be in-depth. It is enough to get the core numbers that will determine the effectiveness of training at each level so these numbers can be tied to the next level.
3. For data analysis use only the most credible sources. Direct estimates from participants themselves and their supervisors or a combination of both lends credence to the data.
4. When analyzing data, choose the most conservative among alternatives. This ensures that credibility of the whole process is enhanced in the eyes of all stake-holders and also protects the ROI output data against tough scrutiny.
5. If no data is available of improvement from a target audience, it is assumed that no improvement has taken place.
6. Extreme data items and non-credible claims should not be used. Widely fluctuating data, self-supporting claims without statistical support or huge deviations are to be eliminated from the calculations.

Critics of the ROI concept have been baying for blood some time now. They feel it is an imperfect science where estimates are made about potential savings. Some advocate doing away with ROI altogether and replacing it with a sound front-end training need analysis. Professionals working in this area argue that doing an after-the-event analysis, after the money has already been spent on a training programme, is illogical; a proper before-the-event analysis is more productive. After-the-event-evaluation is fine for academicians, researchers and scholars but, from a business point

of view, it's too late. Front-end analysis starts with the performance indices of the organization as a focus and not with the training course or training department. Historical data suggests that very few companies do evaluation at level 4 or level 5. In order to calculate ROI two factors are critical; isolating the net cost of training and isolating the net monetary impact of the training. The second part can prove to be tricky.

One can tie training needs to business results by isolating critical knowledge and skill areas that impact results directly, thus establishing linkages between training efforts and results. If you can identify these few critical components, is it not logical to focus on building these competencies so they will have a causal link with results? If your analysis identifies measurable objectives, performance gaps, critical tasks that need immediate attention and helps focus your training energies on them, then you have a guiding thread throughout all levels of measurement. The more these competencies improve; stronger will be the causal link between training and business results. There are many companies who do little or no training evaluation; they concentrate on front-end-analysis and see the impact on business outcomes. If the outcomes improve, training is spot-on and there is no need to do extensive ROI work. Reportedly, Toyota which has a market cap of more than \$ 130 billion, do little evaluation of training.

Those who have no patience with ROI in training feel that business result is the outcome of many dynamic variables working on the tangibles. Is it possible to prove beyond all doubts that the increase in output was the result of the one or many learning interventions? In the real world, any training practitioner will be laughed at if he claims that his training was solely responsible for that 30% raise in sales. There would be other contenders for the pound of flesh; marketing will claim that their newly churned out strategy has worked, logistics will claim that their new policy has been responsible for the improved efficiencies, HR will claim their fare share, and manufacturing will stake their claims and so on. Even if you are somehow able to isolate the financial impact of training, it's going to be time-consuming, expansive and not so clean. On the other hand, starting with established metrics of a before-the-event evaluation will ensure you will be dealing with known variables and bench-marks which will guide you throughout all levels of evaluation.

Practicing T & D professional will appreciate that when we evaluate something we are arriving at the value; value is different when viewed from different perspectives. More valuable or less valuable compared to what? ROI standards differ industry-wise. If we compare returns from manufacturing, trading, service industry and R& D, the ROI's tend to vary widely. Inside the same organization if we compare marketing, HR, finance, manufacturing and R&D, the ROI values will yet again vary widely. T & D workers need to understand the present ROI metrics being used in various functions in their own setting and then set-out to do ROI calculations for training. One white paper published on this subject pegs ROI from T & D to be as high as 25%; in today's world where margins are under constant strain and ever increasing competition and input costs, this is an eye-popping figure for any management to look very seriously at investments in Training.

Other issues we need to understand are, what is the scope of the training intervention, is it short term or long term? Who are the champions of the programme; the senior management team which is looking for improved efficiencies which will impact bottom line, the HR / PR team who want to blitz the number of programmes done in a year, the line-managers who are genuinely concerned with efficiency and productivity numbers against increasing competition and targets, the trainees who want to have a nice time / those who are interested in increasing their competencies or the training department personnel who have to necessarily do some training throughout the year. Answer to any one of them or combination of them will dictate the training strategy and tactics.

In conclusion, we have found that a combination of common sense Kirkpatrick approach and a no-nonsense ROI concept application is needed to maximize training results in order to establish training credibility. Every T & D worker also needs to honestly answer some basic questions before he/she sets out to do any training.

1. What are the 3 critical knowledge and skill gaps that impact business results of your particular industry? (You will be strongly tempted to go for more; you'll lose training focus)
2. Who in the organization, among the stakeholders, endorse these 3?

-
3. If you plug these 3 what will be the estimated improvements in outcomes and who else says so (other than you) ?
 4. Who are the trainees who need improvement IMMEDIATELY on these identified 3 areas and who says so?
 5. What is the instructional design and who are the partners in the development of this design ?
 6. How will you use level 1, 2 and 3 evaluation data? Specially, do you have the buy-in from the supervisors of the trainees to reasonably measure changes in behavior (level 3) or application of training concepts ?
 7. Is the application of new learning improving outcomes (level 4) ?
 8. What programmes qualify for ROI calculations & how will you use this to enhance credibility of T& D in your organization?
 9. What resources do I have for the interventions, are they enough?

In a chaotic world where change is speeding up, we are never going to have all the answers or ideal set up to start any training intervention. All we have are accepted practices within our industry, what we can experiment with from outside our industry and the ability to take risks in altogether new directions. If we keep in mind the True North principle, "T & D exists only to build human competencies that improve business metrics", we will be bowling to the stumps.