Evaluation of Training: Need to Focus

Manik M. Wankhede, Student, Pune University
Dr. Rajashree Gujarathi, Ph D, Pune University

ABSTRACT
Training is one of the important or we can say that most pervasive method for enhancing the ability of individuals and communicating organizational goals to new personnel, given importance and impact of training on organization and expenditure associated either the development or implementation of training.
This is also important that training is used as continuously run process, further modern technology and its related charges, automation, require updating the skill and knowledge, but it is also important to justify the financial investment in the training and development programs.
The paper attempts to study the variety of sources on evaluation of training and its subsystems in the municipal councils by using statically methods. The study is based on literature on the subject. The paper also analyze the evaluation of training

Key words: Evaluation, Assessment, Objectives, Training Program, Monitoring.

Introduction
In today’s rapid age the increased accountability, the training evaluation process is critical component of an organization’s training program. Organizations administering the program not only are accountable for what employees learn, they also are accountable for ensuring that employees transfer their knowledge to their work performance. While traditional training evaluation methods focus on using the assessment process to improve training delivery, information should also be collected to determine whether training is assisting the organization to improve its business performance.

Evaluation is hardly a new subject for discussion in the HR and Training system and it is increasingly becoming a critical one. Several researchers recently reviewed the topic; among the conclusions drawn was that, “The enabling conditions for evaluation across UNDP remain weak. These include the variable quality of frameworks for monitoring and evaluation, the lack of systematic results monitoring, the strategic allocation of human and financial resources, and the mechanisms for quality assurance. Evaluation itself continues to be poorly organized and funded across the organization.”
It was also noted that evaluation was not adequately carried out properly during training. The major objects for evaluations of training are as under

a) To undertake a literature search, document analysis and a critical review of methods and processes of training impact evaluation, with a view to determine the relevance of these approaches to the assessment of the impact of UN Fellowships programmes;

b) With reference to the review of the literature the scope, dimensions and core indicators for evaluating the impact training programmes;

c) To identify necessary conditions and supportive measures to enable implementation of the impact evaluation framework in the context of training programmes.

The present report, derived from a variety of sources on evaluation, takes prime responsibility for the review and analysis of the literature on the subject and will also attempt to identify organizational “measures” which could support and enhance an evaluation framework for various agencies. Its sources are identified through considerable number of footnotes. The report should be viewed as a compilation and synopsis of the work of evaluation specialists. An evaluation refers to a wide range of activities, processes, products, etc. An astounding volume of literature is devoted to training evaluation in the area of reaction, learning, behavior, change, etc.
II. What is evaluation?

1. Definition

Systematic valuation can provide the information needed for continuous improvement. Moreover, today managers are no longer satisfied with knowing how many fellows underwent training, how they liked it, and what they learned. Increasingly managers want to know if the fellows are using what they learned, and –most importantly – what if any institutional results were improved. In any review of evaluation it is first essential to define the term itself, as well as its stakeholders and its goals. Then an analysis can be made of the various types of evaluation and major models/methodologies commonly applied to measure impact.

Probably the most frequently given definition is:

"Evaluation is the systematic assessment of the worth or merit of some object"  

The definition is hardly perfect. There are many types of evaluations that do not necessarily result in an assessment of worth or merit – descriptive studies, implementation analyses, and formative evaluations, to name a few. Better perhaps is a definition that emphasizes the information-processing and feedback functions of evaluation. For instance, one might say:

"Evaluation is the systematic acquisition and assessment of information to provide useful feedback about some object."

Another definition sees evaluation as the systematic collection and analysis of data needed to make decisions (see Lana Muraskin, “Understanding Evaluation, the Way to Better Prevention Programs”).

With all the slight differences in each definition, several distinct “steps” are usually followed in any evaluation:

- **STEP 1** Get an overview of the programme
- **STEP 2** Determine why you are evaluating
- **STEP 3** Determine what you need to know and formulate research questions
- **STEP 4** Figure out what information you need to answer questions
- **STEP 5** Design the evaluation
- **STEP 6** Collect information/data
- **STEP 7** Analyze information
- **STEP 8** Formulate conclusions
- **STEP 9** Communicate results
- **STEP 10** Use results to modify programme

It should be noted that most definitions emphasize acquiring and assessing information rather than assessing worth or merit because all evaluation work involves collecting and sifting through data, making judgements about the validity of the information and of inferences derive from it, whether or not an assessment of worth or merit results.

This feedback is provided to the stakeholders, i.e. anyone without whose input an organization would be unable to function. In evaluation the term stakeholder can be more broadly defined as one who may have no formal role in an organization’s functions but still be positively or negatively affected by its functioning. The goals of evaluation are to influence decision-making or policy formulation through the provision of empirically driven feedback. It is imperative to ensure that all stakeholders are included in the evaluation process before a particular evaluation type or model/methodology is chosen. Indeed, one evaluation model (or method as some prefer) focuses essentially on the stakeholders, arguing that to adequately evaluate training it is necessary to assess the extent to which all stakeholder groups are satisfied with what they have given to and received from the training.

---

1 Evaluation literature has a jargon of its own but even among the specialists terminology can differ. What is “purpose” for one may be “type” for another. To provide a perspective on evaluation, somewhat different than the present synopsis, Paul Duignan’s “Introduction to Strategic Evaluation” at http://www.strategicevaluation.info/documents/104.htm is recommended reading (8 pages) for its definition of evaluation approaches, purposes, methods and designs. also see Zinovieff, and Rotems “Review and analysis of Training, Impact evaluation methods, and proposed measures to support a United Nations system Fellowships Evaluation Framework”, 2008.

For a very useful on line “course” on the general subject see also “Introduction to Evaluation”, the course on evaluation and change by the Instructional Systems Technology Dept. of Indiana University’s School of Education at www.indiana.edu/~istr561/knuth06sum/unit1a.shtml.

2 See www.socialresearchmethods.net/kb/intreval.htm

3 See William Trochim, currently President of the America Evaluation Association, at www.eval.org/part_1.asp

4 See www.evaluationwiki.org/wiki/index.php/Evaluation_Definition

5 See Fred Nichols “A Stakeholder Approach to Evaluation"
2. Dimensions

Evaluation literature refers to the “dimensions of evaluation” as process, outcome and impact. These concepts are fundamental and we will return to them in other contexts more fully.

- **Process evaluations**

Process Evaluations describe and assess programme materials and activities. Establishing the extent and nature of programme implementation is an important first step in studying programme outcomes; that is, it describes the interventions to which any findings about outcomes may be attributed. Outcome evaluation assesses programme achievements and effects.

- **Outcome evaluations**

Outcome Evaluations study the immediate or direct effects of the programme on participants. The scope of an outcome evaluation can extend beyond knowledge or attitudes, however, to examine the immediate behavioral effects of programmes.

- **Impact evaluations**

Impact Evaluations look beyond the immediate results of policies, instruction, or services to identify longer-term as well as unintended programme effects. Very useful reports on this subject have notably been made by the Center for Global Development\(^6\), and by Deloitte Insight Economics\(^7\). For a comprehensive review of the three dimensions – process, outcome and impact

3. Goals\(^8\)

The generic goal of most evaluations is thus to provide useful feedback to a variety of audiences including sponsors, donors, client-groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived as “useful” if it aids in decision making. But the relationship between an evaluation and its impact is not a simple one -- studies that seem critical sometimes fail to influence short-term decisions, and studies that

4. Approaches

An evaluation approach is a general way of looking at or conceptualizing evaluation; the main evaluation approaches according to Paul Duignan (“Introduction to Strategic Evaluation”) include notably:

- **Utilisation-focused evaluation** – determines methods on the basis of what is going to be most useful to different audiences

- **Empowerment evaluation** – emphasises that the evaluation process and methods should be empowering to those who are being evaluated

- **Stakeholder evaluation** – looks at the differing perspectives of all of a programme’s stakeholders (those who have an interest in it)

- **Goal-free evaluation** – in which the evaluator’s task is to examine all of the outcomes of a programme, not just its formal outcomes as identified in its objectives;

- **Naturalistic or 4th generation evaluation** – emphasises the qualitative uniqueness of programmes and is a reaction against the limitation of quantitative evaluation approaches;

- **Theory based evaluation** – puts an emphasis on detailing the assumptions on which a programme is based (intervention logic) and follows those steps to see if they occur;

- **Strategic evaluation** – emphasises that evaluation design decisions should be driven by the strategic value of the information they will provide for solving social problems.

5. Purposes

There are various ways of describing various purposes of evaluation activity, e.g. design, developmental, formative, implementation, process, impact, outcome and summative. The evaluation purpose is best understood as identifying what evaluation activity is going to be used for. Recent years have seen evaluation move to develop types of

---

\(^6\) See “When Will We Ever Learn: Improving Lives through Impact Evaluation”, May 2006 publication by the Center for Global Development.


\(^8\) See inter alia www.socialresearchmethods.net/kb/intreval.htm. above.
evaluation that are of use right across a programme lifecycle. It should be noted that any particular evaluation activity can have more than one purpose. The range of evaluation terms are used in various ways in the evaluation literature.

- **Design, developmental, formative, implementation** – evaluative activity designed to improve the design, development, formation and implementation of a programme;

- **Process** – evaluation to describe the process of a programme. Because the term process could conceivably cover all of a programme from its inception to its outcomes, it is conceptually useful to limit the term process evaluation to activity describing the programme during the course of the programme, i.e. once it has been initially implemented;

- **Impact, outcome and summative** – looking at the impact and outcome of a programme, and in the case of summative, making an overall evaluative judgment about the worth of a programme. The purposes of evaluation also relate to the intent of the evaluation.9

- **Gain insight** – provide the necessary insight to clarify how programme activities should be designed to bring about expected changes;

- **Change practice** – improve the quality, effectiveness, or efficiency of programme activities;

- **Assess effects** – examine the relationship between programme activities and observed consequences;

- **Affect participants** – use the processes of evaluation to affect those who participate in the inquiry. The systematic reflection required of stakeholders who participate in an evaluation can be a catalyst for self-directed change. Evaluation procedures themselves will generate a positive influence.

Some organizations work towards the Investors in People (IIP) initiative which defines the purposes of evaluation as follows:10

**IIP Indicator 4.1** The organization evaluates the impact of training and development on knowledge, skills and attitude

**IIP Indicator 4.2** The organization evaluates the impact of training and development actions on performance

**IIP Indicator 4.3** The organization evaluates the contribution of training and development to the achievement of its goals and targets.

In the general discussion on the purposes of evaluation a note of caution is often introduced with respect to cost vs. benefit. This is known as “rationalizing evaluation”. It is important not to get “carried away” with an evaluation effort which is disproportionately greater than the investment made, or the benefit likely to be achieved. The IPD (Independent Professional Development, London) study “Making Training Pay” (1997) suggests that the scope of an evaluation strategy should be carefully weighed against the following considerations:

- The size of the training investment
- The number of staff involved
- The likelihood that the training will be repeated
- The criticality of the training to the business
- The “newness” of the training methods used.

**6. Types**11

The most basic difference is between what is known as the formative and the summative types of evaluation. In more recent years the concepts of confirmative and meta evaluation have received much attention as well.

**Formative & Summative types of evaluation**

Formative evaluations strengthen or improve the object being evaluated – they help form it by examining the delivery of the programme or technology, the quality of its implementation, and the

---

9 See “University of British Columbia e-learning course”.
10 See inter alia www.dba.co.uk/tips/vol3/vol3iss5.htm.
11 See also www.evaluators-webring.net/Independent_evaluators_webring_definitions_May06.pdf
assessment of the organizational context, personnel, procedures, inputs, and so on. Summative evaluations, in contrast, examine the effects or outcomes of some object – they summarize it by describing what happens subsequent to delivery of the programme or technology; assessing whether the object can be said to have caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes; and, estimating the relative costs associated with the object.

i. Formative

Formative evaluation includes several evaluation types:

- **Needs assessment** determines who needs the programme, how great the need is, and what might work to meet the need;
- **Evaluability assessment** determines whether an evaluation is feasible and how stakeholders can help shape its usefulness;
- **Structured conceptualization** helps stakeholders define the programme or technology, the target population, and the possible outcomes;
- **Implementation evaluation** monitors the fidelity of the programme or technology delivery;
- **Process evaluation** investigates the process of delivering the programme or technology, including alternative delivery procedures.

ii. Summative

Summative evaluation can also be subdivided:

- **Outcome evaluations** investigate whether the programme or technology caused demonstrable effects on specifically defined target outcomes;
- **Impact evaluation** is broader and assesses the overall or net effects – intended or unintended – of the programme or technology as a whole;
- **Cost-effectiveness and cost-benefit analysis** address questions of efficiency by standardizing outcomes in terms of their dollar costs and values;
- **Secondary analysis** re-examines existing data to address new questions or use methods not previously employed;
- **Meta-analysis** integrates the outcome estimates from multiple studies to arrive at an overall or summary judgment on an evaluation question.

There is an abundance of literature on the subject of formative and summative evaluations; the questions and methodologies addressed under these types.

To a number of evaluation specialists these two fundamental types of evaluation do not suffice to constitute what is referred to as “Full-scope evaluation”\(^\text{12}\)

Full scope evaluation systematically judges the merit and worth of a long-term training programme before, during, and after implementation. Full scope evaluation is appropriate only for training programmes that are designed to run for one year or more; it is not appropriate for a one-time training event, such as a single-session workshop to introduce a new product to sales representatives.

Full-scope evaluation integrates four types of programme evaluation – formative, summative, confirmative and Meta – into the training programme evaluation plan. Working together, the four types of evaluation help to determine the value of a long-term training programme and develop the business case or rationale for maintaining, changing, discarding, or replacing the programme. Full-scope evaluation introduces the concepts of the confirmative and the Meta type of evaluation. These concepts may not be of priority interest to other than evaluation practitioners but deserve at least the brief mention below.

iii. Confirmative

Confirmative evaluation goes beyond formative and summative evaluation; it moves traditional evaluation a step closer to full-scope evaluation. During confirmative evaluation, the evaluation and training practitioner collects, analyzes, and interprets data related to behaviour, accomplishment, and results in order to determine the continuing competence of learners or the continuing effectiveness of instructional materials and to verify the continuous quality improvement of education and training programmes.

While formative and summative evaluations comprise two initial levels, confirmative evaluation assesses the transfer of learning to the “real world”:

\(^{12}\) See notably Joan Dessinger and James Morley, ”Full-scope Evaluation: Raising the bar”, 2003
a) **Level one**: evaluate programmes while they are still in draft form, focusing on the needs of the learners and the developers;

b) **Level two**: continue to monitor programmes after they are fully implemented, focusing on the needs of the learners and the programme objectives;

c) **Level three**: assess the transfer of learning to the real world.

Even level four of Kirkpatrick's four levels of evaluation is confirmative evaluation by another name. Level four measures the results of training in terms of change in participant behaviour and tangible results that more than pay for the cost of training.  

iv. **Meta**  

Formative, summative, and confirmative evaluation are all fodder for Meta evaluation. Meta evaluation is all about evaluating the evaluation. The evaluator literally zooms in on how the evaluation was conducted. The purpose of Meta evaluation is to validate the evaluation inputs, process, outputs, and outcomes. It serves as a learning process for the evaluator and makes the evaluators accountable.

There are two types of Meta evaluation: type one and type two. Type one Meta evaluation is conducted concurrently with the evaluation process. It is literally a formative evaluation of evaluation.

Type two Meta evaluations is the more common approach. It is conducted after formative, summative, and at least one cycle of confirmative evaluation is completed. Some evaluation specialists have also defined the several types of evaluation more thematically, as below.

Goal-based  

Goal-based evaluations are evaluating the extent to which programmes are meeting predetermined goals or objectives. Questions to ask when designing an evaluation to see if the goals have been reached include:

- How the programme goals (and objectives, if applicable) were established? - Was the process effective?
- What is the status of the programme’s progress toward achieving the goals?
- Will the goals be achieved according to the timelines specified in the programme implementation or operations plan? If not, then why?
- Do personnel have adequate resources (money, equipment, facilities, training, etc.) to achieve the goals?

vi. **Process-based**

Process-based evaluations are geared to fully understanding how a programme works – how does it produce that results that it does. These evaluations are useful if programmes are long-standing and have changed over the years, employees or customers report a large number of complaints about the programme, there appear to be large inefficiencies in delivering programme services and they are also useful for accurately portraying to outside parties how a programme operates. There are numerous questions that might be addressed in a process evaluation. These questions can be selected by carefully considering what is important to know about the programme. Examples of questions include:

- On what basis do employees and/or the customers decide that products or services are needed?
- What is required of employees in order to deliver the product or services?
- How are employees trained about how to deliver the product or services?
- How do customers or clients come into the programme?
- What is required of customers or client?

vii. **Outcomes-based**

Evaluation with an outcomes focus is increasingly important for nonprofits and asked for by funders. An
outcomes-based evaluation tries to ascertain if the organization is really doing the right programme activities to bring about the outcomes it believes to be needed by its clients. Outcomes are benefits to clients from participation in the programme. Outcomes are usually in terms of enhanced learning (knowledge, perceptions/attitudes or skills) or conditions, e.g. increased literacy, self-reliance, etc. Outcomes are often confused with programme outputs or units of services, e.g. the number of clients who went through a programme. The United Way of America (www.unitedway.org/outcomes/) provides an excellent overview of outcomes-based evaluation, including introduction to outcomes measurement, a programme outcome model, why to measure outcomes, use of programme outcome findings by agencies, eight steps to success for measuring outcomes, examples of outcomes and outcome indicators for various programmes and the resources needed for measuring outcomes.

7. Programme evaluation

Before reviewing the main thematic “models” of evaluation, a few words on the interlinking of “goals”, “objectives” and “programme”. In evaluation literature “programme evaluation” is often used. It is more than an evaluation “type” but neither is it a “model” per se. Michael Patton in particular has written extensively on the subject\(^\text{18}\). Broadly, programme evaluation is a comprehensive form of ascertaining to what extent goals/objectives have been achieved. To effectively conduct programme evaluation one first needs to have “programme” (a strong impression of what customers/clients actually want and need).

Programme evaluation is carefully collecting information about a programme or some aspect of a programme to make necessary decisions about it. Programme evaluation can include any or a variety of at least 35 different types of evaluation (according to Patton, others have identified even more). The type and model of evaluation one undertakes to improve one’s programmes depends on what one wants to learn about the programme.

One should worry less about what type/model of evaluation one needs and worry more about what one needs to know to make the programme decisions one needs to make, and worry also about how one can accurately collect and understand that information.

Patton notes that among the key questions to consider when designing a programme evaluation the following are of priority:

1) For what purposes is the evaluation being done, i.e. what do you want to be able to decide as a result of the evaluation?
2) Who are the audiences for the information from the evaluation, e.g. customers, bankers, funders, board, management, staff, customers, clients, etc?
3) What kinds of information are needed to make the decision you need to make and/or enlighten your intended audiences, e.g. information to really understand the process of the product or programme (its inputs, activities and outputs), the customers or clients who experience the product or programme, strengths and weaknesses of the product or programme, benefits to customers or clients (outcomes), how the product or programme failed and why, etc.
4) From what sources should the information be collected, e.g. employees, customers, clients, groups of customers, or clients and employees together, etc.
5) How can that information be collected in a reasonable fashion, e.g. questionnaires, interviews, examining documentation, observing customers or employees, conducting focus groups among customers or employees, etc.
6) When is the information needed (so, by when must it be collected)?
7) What resources are available to collect the information?

The writings of Patton stress the need to focus on “programmes” (goals/objectives) before initiating an evaluation. To undertake evaluation one must also choose an appropriate “type”, as identified above, and then determine if anyone “model” or “method”, and/or combination of such, best fits one’s situation. We will therefore now describe some major “models” thematically.

8. Models/methods

The thematic categorization of evaluation types is echoed in the literature with respect to the various evaluation “models” (some prefer the term “methods”) which is of prime importance in the effort to identify an evaluation framework.

A recent article on “approaches to evaluation” by Deniz Eseryel identifies six such general approaches:

- Goal-based evaluation
- Goal-free evaluation
- Responsive evaluation
- Systems evaluation
- Professional review
- Quasi-legal

The Indiana University website above defines these approaches as follows:

- Goal-based evaluation begins with goals in mind and seeks to determine if those goals were achieved;
- Goal-free evaluation does not seek to confirm or deny a pre-determined outcome or goal. Rather, it seeks to discover any benefits that result from the intervention;
- Responsive evaluation is an approach that is based on client requirements. This can present unique challenges for the evaluator, but it is a common approach;
- The systems approach to evaluation focuses on whether the intervention was efficient and effective;
- Professional review evaluation uses external expert appraisal to evaluate instead of other commonly used and accepted methods;
- The quasi-legal approach is infrequently practiced, but is uses an actual court-of-inquiry format to present evidence, take testimonials, and evaluate an intervention or product.

Generally, however, the literature focuses essentially on goal or objective-based vs. systems-based models. Goal-based models (such as the “bible” of evaluation models, Donald Kirkpatrick’s “Evaluation Training Programs”) may help. Practitioners think about the purposes of evaluation ranging from purely technical to covertly political purpose. However, these models do not define the steps necessary to achieve purposes and do not address the ways to utilize results to improve training. The difficulty for practitioners following such models is in selecting and implementing appropriate evaluation methods (quantitative, qualitative, or mixed). Naturally, many organizations do not use the entire model, and training ends up being evaluated only at there action, or at best, at the learning level. As the level of evaluation goes up, the complexities involved increase. This may explain why only levels 1 and 2 are used. More on the Kirkpatrick’s model below.

On the other hand, system-based models (e.g. CIPP, IPO and TVS) seem to be more useful in terms of thinking about the overall context and situation but they may not provide sufficient granularity. Systems-based models may not represent the dynamic interactions between the design and the evaluation of training. Few of these models provide detailed descriptions of the processes involved in each steps. None provide tools for evaluation. Furthermore, these models do not address the collaborative process of evaluation, that is, the different roles and responsibilities that people may play during an evaluation process. More on these models below.

i. Objective/goal – based

a) Donald Kirkpatrick’s 4 levels

Discussion on the subject of evaluation types may appear somewhat academic. However, in evaluation literature this discussion inevitably leads to the very concrete examples of evaluation models and schemes. The most famous – and applied – evaluation model was developed by Donald J. Kirkpatrick (notably in his “Evaluating Training Programs”). Kirkpatrick described 4 levels of training evaluation: reaction, learning, behavior and results.

He identified the four levels as:

- **Reaction** – a measure of satisfaction (what the trainees/fellows thought and felt about the training); evaluation here focuses on the reaction of individuals to the training or other improvement intervention;
- **Learning** – a measure of learning (the resulting increase in knowledge or capability); evaluation here assesses what has been learned as measured with end of course tests;
- **Behaviour** – a measure of behaviour change (extent of behaviour and capability improvement and implementation/application); evaluation here measures the transfer of what has been learned back to the workplace;
- **Results** – a measure of results (the effects on the institutional environment resulting from the fellows’ performance); evaluation here measures (at least tries to) the impact of the training on overall organizational results (in the private sector on business results).

In the framework of the above summary of “types” of evaluation levels 1 and 2 are normally seen as part of formative evaluation, whereas levels 3 and 4 are typically associated with summative evaluation. There have also been attempts to establish a level 5 by measuring the impact at a societal level (in business terms, by calculating return on investment (ROI)). Levels 4 and 5 are associated with normative and/or Meta evaluation to achieve an ideal “full-scale” evaluation.

Although most organizations have some form of level 1 evaluations for their training programmes, the number diminishes as the levels increase. And very few organizations take advantage of the rewards of level 4 evaluations. The following should be borne in mind:

Level 1 (reaction) and level 2 (knowledge and skills) evaluations can lead to a false sense of security; there may be no relationship between how participants feel about the training and improved individual and organizational performance; level 3 evaluations can be used to refine the training provided, but level 4 will determine whether it has value. It may not be desirable, practical, or necessary to do all levels of evaluation. Each organization needs to select the level that will produce the information required to evaluate the training programme.

Again, Kirkpatrick’s model consists of 4 levels that progress in difficulty from 1 (the easiest to conduct) to 4 (the hardest). When choosing the appropriate model to include in an organizational assessment, it is essential first to identify the questions the evaluation needs to address. Kirkpatrick expressed this in a tabulation:

<table>
<thead>
<tr>
<th>Level Measurement focus Questions addressed</th>
<th>Level 1 (Reaction)</th>
<th>Level 2 (Learning)</th>
<th>Level 3 (Behavior)</th>
<th>Level 4 (Results)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction Trainees’s perceptions what did trainees think of this training?</td>
<td>Trainees’s perceptions what did trainees think of this training?</td>
<td>Knowledge/skills gained Was there an increase in knowledge or skill level?</td>
<td>Worksite implementation Is new knowledge/skill being used on the job?</td>
<td>Impact on organization what effect did the training have on the organization?</td>
</tr>
</tbody>
</table>

Level 1 (Reaction) is the most commonly-used method of evaluation, probably because it is the easiest to administer and evaluate. This level produces what has been dubbed the “smile sheet”, which measures how well the trainees like the training. Level 2 (Learning) is not as well-used in business settings as an evaluation technique; public sector/academic settings are more likely to use level 2 techniques; these are most reliable when pre- and post-evaluations are utilized. The fact is that in both private and public sectors there is today an increasing need to show concrete evidence that training/fellowships are achieving their goals of changing behavior on the job (level 3) and are also contributing to the institutional “bottom” line. The problem is that trainers will probably not do levels 3 and 4 evaluations unless they are told to do so. Level 3 evaluations are difficult because human behavior needs to be measured. Some believe level 4 evaluations may actually be easier to accomplish than level 3, since level 4 is (at least ideally) tied to measurable information. Some trainers therefore believe that a positive level 3 evaluation implies success at level 4.

For an excellent synopsis of Kirkpatrick’s model, inclusive of comparative “grids” and tabulations, see which go beyond the first two levels. If it is, they would then need to decide on the most appropriate evaluation model/methodology. In fact, research from the American Society of Training and Development shows that over 75% of organizations measure only the level of reaction through the use of questionnaires, the “smile” or “happy sheets”. The ASTD study also showed (2002) that 44% of US corporations measured how much people learned, while only 21% how much people changed their behavior as a result of training; indeed a mere 11% measured whether training affected organizational results. This is a danger for the quality of training, of course, for if participant reaction is the only measure of performance too much energy, on the part of trainers, could be devoted to obtaining favorable ratings and not helping people to learn, grow, and change (for the better). In this context critics of the Kirkpatrick model (more on this later) note that just as there is no proven causal link between reaction and performance too much energy, on the part of trainers, could be devoted to obtaining favorable ratings and not helping people to learn, grow, and change (for the better). In this context critics of the Kirkpatrick model (more on this later) note that just as there is no proven causal link between reaction and performance.

---

20 For UN agencies, the issue is whether they are prepared to spend the money to carry out genuine evaluation exercises.

21 See the ASTD “Training & Development Handbook”, edited by Robert Craig, which inter alia contains a chapter by Kirkpatrick on “Evaluation” and a chapter by Jack Phillips on “Measuring the Results of Training” but beware – the Handbook runs well over one thousand pages.
learning, there is no such link between learning and behavior change. Just because a participant learned something does not mean anything will be done with the learning hence the importance of behavior change evaluation well after the training and, to some, recourse to different evaluation models/methods.²²

b) Jack Phillips Return on Investment (ROI)²³

In his many books and articles Phillips has gone beyond even Kirkpatrick’s level 4 to focus on real measurement of ROI (justification of the cost of training based on the return on investment and organizational impact). Training in this sense has thus moved from satisfying trainees to improving organizational performance. Training are carried out to have a positive impact on the organization. This is obviously a far cry from the “smile sheets” forming the basis of level 1 evaluation. Today, many evaluators point out that while the Kirkpatrick’s model is useful to evaluate a) whether learners liked their instruction, b) whether they learned something from it, and c) whether it had some positive effect for the organization, its weakness is that it cannot be used to determine the cost-benefit ratio of training (ROI). These modern evaluators have consequently recommended adding the so-called fifth level to Kirkpatrick’s model, at least for some programmes. This may be too much for a UN agency since it requires collecting level 4 data, converting the results to monetary values, and then comparing those results with the cost of the training programme. Research into the training evaluation models that have been proposed over the last 45 years since Kirkpatrick’s framework, show that many have used the four levels as a basis for their thinking, though Phillips has also had “impact”.

A few of these (goal-based) models/methods are identified below.²⁴

c) Hamblin’s 5 levels²⁵

Hamblin was one of the first to modify Kirkpatrick’s model. The first three levels in his model correspond closely to Kirkpatrick’s model. However, the final level is split into two: organization and ultimate value. The five level model is therefore:

- **Level 1**: Reactions
- **Level 2**: Learning
- **Level 3**: Job behaviour
- **Level 4**: Organization – the effects on the organization, from participant’s job to performance changes
- **Level 5**: Ultimate value – the financial effects, both on the organization and the economy.

d) Guskey’s critical levels

Thomas Guskey (2002) has also elaborated Kirkpatrick’s 4 levels into 5: his levels may be of relevance as he had “students” and educational environments in mind (see the University of Minnesota website “Education Minnesota”²⁶)

- **Level 1**: Participant reaction
  - Purpose: to gauge the participants’ reactions about information and basic human needs
  - Technique: usually a questionnaire
  - Key questions: was your time well spent? Was the presenter knowledgeable?
- **Level 2**: Participant learning
  - Purpose: examine participants’ level of attained learning
  - Technique: test, simulation, personal reflection, full-scale demonstration. Key question: did participants learn what was intended?
- **Level 3**: Organizational support and learning
  - Purpose: analyze organizational support for skills gained in staff development
  - Technique: minutes of district meetings, questionnaires, structured interviews or unobtrusive observations
  - Key questions: were problems addressed quickly and efficiently? were sufficient resources made available, including time for reflection?
- **Level 4**: Participant use of new knowledge and skills

²³ See notably 2003 edition of his “Return on Investment in Training and Performance Improvement Programs”.
²⁴ See listings published and analyzed by the UK Institute for Employment Studies.
²⁶ www.educationminnesota.org/profdev/tall/Pages/5levels.aspx.
• Purpose: determine whether participants are using what they learned and using it well
• Technique: questionnaires, structured interviews, oral or written personal reflections, examination of journals or portfolio, or direct observation
• Key question: are participants implementing their skills and to what degree?

- Level 5: Student learning outcomes
- Purpose: analyze the correlating student learning objectives
- Technique: classroom grades, tests, direct observation
- Key question: did student show improvement in academic, behaviour or other areas?

**e) Indiana University taxonomy**

Indiana University developed an evaluation taxonomy-based on six strata, which were not intended to be a hierarchy of importance. The first and last strata provide additions to Kirkpatrick’s framework:

- Stratum 1: activity accounting – which examines training volume and level per participant
- Stratum 2: participant reactions
- Stratum 3: participant learning
- Stratum 4: transfer of training
- Stratum 5: business impact
- Stratum 6: social impact

The sixth stratum examines the impact of changed performance on society, and as such is similar to Hamblin’s ultimate value.

**f) Industrial Society stages**

The Industrial Society (now the Work Foundation) developed a six stage circular model which starts with a planning phase. The stages are:

- Stage 1: identify the business need
- Stage 2: define the development objectives
- Stage 3: design the learning process.
- Stage 4: experience the learning process
- Stage 5: use and reinforce the learning
- Stage 6: judge the benefits to the organization (quality measures, customer satisfaction and financial benefits provide the main measures at this level).

The Industrial Society differentiated between stages 3 and 4 which aim to validate the training, and stages 5 and 6 which aim to evaluate it. True evaluation needs to take place long before and after training has taken place and the process of identifying the business need is an essential component of the evaluation model.28

**g) Kearns and Miller KPMT model29**

Kearns and Miller’s KPMT model has many similarities to Phillips work. They argue that clear objectives are an essential component of a training evaluation model. Where they differ is in their aim to provide a sort of toolkit to help evaluators work through the process of identifying bottom-line objectives by means of questioning techniques, evaluating existing training, and using process mapping to identify the added value to organizations.

They argue that training can only bring added value to organizations if the business is not performing effectively or there is a market opportunity which can be exploited. To identify bottom line benefits, pre-training measurements need to be in place. Only where the training is to bring someone up to the standards of the job is this not necessary.

The four-stage KPMT model starts at the beginning of the training cycle by identifying the business need rather than the training need. The emphasis is on clarifying objectives from a business perspective rather than that of the trainees. Despite this, the evaluation levels look very similar to Kirkpatrick’s:

- Reaction to training and development
- Learning
- Transfer to the workplace/behaviour
- Bottom line added value, measured in relation to the base level measures taken

---

Where Kearns and Miller differ from some of the other models is in their belief that return on investment can only be looked at in hard terms. They state that if a business objective cannot be cited as a basis for designing training and development, then no training and development should be offered.

**h) Nine outcomes model**

Also worth mentioning is the “Nine Outcomes” model which aims to measure whether training has been successful. In identifying the 9 outcomes, Donovan and Townsend pose 9 questions with the training participants in mind:

- **Reaction** to training – did they like it?
- **Satisfaction** with the organization of a training event (facilities, logistics, meals, etc.)
- **Knowledge** acquisition – did they learn anything?
- **Skills** improvement – can they do something new or better?
- **Attitude** shift – have they changed their opinions about something?
- **Behaviour** change – have they changed their way of doing things following the training?
- **Results** – how did the training impact on the organization’s key success factors?
- **Return on investment** – to what extent did the training give back more than it cost?
- **Psychological capital** – how did the training affect corporate image?

All 4 of Kirkpatrick’s levels, incidentally, are included among the 9 outcomes.

**i) Organizational elements model**

Kaufman and Keller (1994) argue that Kirkpatrick’s model was intended for evaluating training, and that as organizations now seek to evaluate other types of development events, the framework needs to be modified. They expanded Kirkpatrick’s model to include societal contribution as an evaluation criteria. They argue that manufacturing organizations in particular are increasingly being called to account for societal consequences such as pollution and safety.

The model also included some additions at the other levels, such as the inclusion of needs assessment and planning in the evaluation, an examination of the desired or expected results, and a review of the availability and quality of resources. They contend that evaluation at all levels should be planned and designed prior to the implementation of any intervention. With the additional help of Watkins in 1995, the team reclassified the criterion in their model into the following six levels:

- **Level 1: Input** – similar to Kirkpatrick’s reaction level, but has been expanded to include the role, usefulness, appropriateness and contributions of the methods and resources used;
- **Level 2: Process** – this level also has similarities to the reaction level, but is expanded to include an analysis of whether the intervention was implemented properly in terms of achieving its objectives;
- **Level 3: Micro (acquisition)** – is similar to the learning level and examines individual as well as small-group mastery and competence;
- **Level 4: Micro (performance)** – links closely to the behaviour level and examines the utilization of skills and knowledge. The focus is on application rather than transfer of skills and knowledge;
- **Level 5: Macro** – relates to the results level and examines organizational contributions and payoffs;
- **Level 6: Mega** – an additional level which looks at societal outcomes. They argue that costs can be examined at each stage, from efficiency measures at the input level to utility costs at the highest level.

**j) Contemporary ROI models**

A range of “contemporary” models/methods for assessing outcomes have been developed, elaborating on Phillips ROI in an effort to somehow better evaluate the ROI. On the face of it these methods have little relevance to public sector international organizations but are identified below if only to give an idea of the myriad of assessment methods available:

- Methodology
- Benefit/cost ratio
- Payback period

---

32 See http://scholar.lib.vt.edu/ejournals/JITE/v41n3/brauchle.html
• Return on true value of dollars
• Present value of dollars and future value of dollars
• Utility analysis
• 360-degree feedback
• Performance teams satisfaction
• Balanced scorecard
• HRD benefit forecasting

ii. Systems-based

Among these the CIPP, IPO and TVS models are perhaps the best known, though the distinction between “goal-based” and “systems-based” is sometimes ambiguous.

a) CIPP (Context, Input, Process, Product)

The CIPP model was developed by Daniel Stufflebeam from 1971 onwards (his latest work is “Evaluation Theory, Models, and Applications”, 2007). It distinguishes four types of evaluation (which we have encountered tangentially in earlier pages):

• **Context evaluation** – which helps in planning and developing objectives
• **Input evaluation** – which helps to determine the design by examining capability, resources and different strategies
• **Process evaluation** – which helps to control the operations by providing on-going feedback
• **Product evaluation** – which helps to judge and react to the programme attainments in terms of outputs and outcomes.

Corresponding to the letters in the acronym CIPP, this model’s core parts are context, input, process, and product evaluation. In general, these four parts of an evaluation respectively ask. What needs to be done? How should it be done? Is it being done? Did it succeed? In this checklist, the “Did it succeed?” or product evaluation part is divided into impact, effectiveness, sustainability, and transportability evaluations. Respectively, these four product evaluation subparts ask. Were the right beneficiaries reached? Were their needs met? Were the gains for the beneficiaries sustained? Did the processes that produced the gains prove transportable and adaptable for effective use in other settings?

As we will see later the “subparts” of the “product evaluation”, i.e. impact, effectiveness and sustainability, are also included in the World Bank’s Independent Evaluation Group evaluation framework and may be of particular relevance to public sector international organizations.

b) IPO (Input, Process, Output)\(^{33}\)

Bushnell developed the IPO model (input, process, output) which focuses more on the inputs to training. The IPO model is used by IBM and helps to monitor employee progress by setting performance indicators at each stage. The stages are:

• **Input** – such as the instructor experience, trainee qualifications, resources
• **Process** – the plan, design, development and delivery of the training
• **Outputs** – the trainees reactions, knowledge and skills gained and improved job performance
• **Outcomes** – profits, customer satisfaction and productivity.

c) TVS (Training Valuation System)\(^{34}\)

Fitz-enz (1994) developed a Training Valuation System (TVS) which is a fourstep process similar to Kirkpatrick’s framework at steps 3 and 4 but has been categorized as “system-based”:

• **Step 1**: Situation analysis – this is similar to an in-depth training analysis. Like Kearns and Miller, he suggests that the manager’s answers are continuously probed until some visible, tangible outcome is revealed and that the questions initially focus on the work process rather than the training;
• **Step 2**: Intervention – this involves diagnosing the problem and designing the training;
• **Step 3**: Impact – this examines the variables that impact on performance
• **Step 4**: Value – this step places a monetary worth on the changed performance.

d) Pulley’s responsive evaluation model \(^{35}\)

\(^{34}\) See J. Fitz-enz “Yes, you can weigh training value”, 1994.
Another system-based evaluation model focuses on the purpose of evaluation, the “responsive evaluation” model developed by Pulley (1994). Responsive evaluation is a tool for communicating evaluation results more effectively by tailoring it to the needs of the decision-makers. Pulley argues that the objective of the evaluation should be to provide evidence so that key decision-makers can determine what they want to know about the programme.

The stages involved are:

- Identify the decision-makers so as to ascertain who will be using the information and what their stake in it is;
- Identify the information needs of the decision-makers – what do they need to know and how will it influence their decisions?
- Systematically collect both quantitative and qualitative data. Pulley argues that the qualitative data is normally relayed in the form of stories or anecdotes and gives life to the numbers;
- Translate the data into meaningful information
- Involve and inform decision-makers on an on-going basis.

e) E-Learning models

More recently, a range of system-based models have been elaborated for evaluating “new technology delivery” – such as online learning and e-Learning, as part of a portfolio of training options available to human resources managers less interested in instruction-led training (see for instance Pollard Hillage “Exploring e-Learning”, 2001). One of many examples is the “Continuous Evaluation of Training Systems Based on Virtual Reality”. These models often seem to have been specifically adapted to evaluate technical scientific achievements.

f) Duignan’s framework for outcomes systems

Paul Duignan, to whom reference has been made in earlier pages, is also the main designer of a very complex system which develops intervention or programme logics using the Outcome Hierarchies diagramming approach. Intervention logics set out the connections between the outcomes an individual, organization, or group of organizations are trying to achieve and the steps, stages or intermediate outcomes which are needed to achieve this. It is a complex and highly technical approach which culminates in the identification of 7 high-level outcome attribution evaluation designs. Duignan’s approach is also referred to as the OIIWA Systematic Outcomes Analysis, OIIWA standing for Outcomes Is It Working Analysis.

iii. Additional evaluation methods, including Contribution Analysis

The models/methods noted above do by no means exhaust the field. Several other methods or approaches exist, some of which may have less relevance for public sector international organizations. A few of these are still identified below:

- Approaches
  - Dixon’s six steps
  - Brinkerhoff’s six stages
  - Bramley’s goal-based
  - Wade’s high impact
  - Shapiro’s matrix
  - Pershing’s perspectives
  - Warr, Bird and Rackham’s CIRO (Context, Input, Reaction, Outcome)
  - Preskill and Torres’ Evaluative Inquiry
  - Kraiger’s Learning Outcomes
  - Kaplan and Norton’s Balanced Scorecard

Contribution Analysis is an innovative approach which may consider dimensions sometimes overlooked by more traditional methods; it can however be used in conjunction with other models and requires considerable intellectual precision.

The evaluation of any training programme has certain aims to fulfill. These are co-related with the determination of change in the organizational behavior and the change needed in the org structure. Hence evaluation of any training programme must inform us whether the training programme has been able to deliver the goals and objectives in terms of cost incurred and benefit achieved. On the basis of above literature, comparison one should understand the importance of evaluation of training programme.

37 See www.oiiwa.org