

USING THE KIRKPATRICK MODEL OF TRAINING EVALUATION

USING THE KIRKPATRICK MODEL OF TRAINING EVALUATION

Halдар, K.¹ and Chandurkar, D.²

¹ Research Manager, Sambodhi Research & Communications, kaveri@sambodhi.co.in

² Chief Knowledge Officer, Sambodhi Research & Communications, dharmendra@sambodhi.co.in

ABSTRACT

Investments in capacity building are a key feature of development interventions underscoring the role of capacity development in enhanced developmental outcomes and sustained change. Also, trainings have become the mainstay of these interventions such that to be synonymous to capacity building. And therefore evaluation of trainings to comment on efficacy of the interventions, efficiency of the investments and contribution to envisioned results becomes imperative. However, the practice of training evaluation is limited with methodological challenges, skills, resource and time constraint being the key contributors.

This paper provides an overview of the various approaches to training and discusses the four-level Kirkpatrick model of training evaluation in detail. It describes the use of the model in training evaluation of a health sector intervention. It describes the methodology, tools and analyses protocols employed for the evaluation and highlight the challenges in application of the model to comprehensively evaluate training interventions.

Keywords: Capacity development, training evaluation, Kirkpatrick model

INTRODUCTION

Capacity development has been an essential component of the development projects and programmes across the globe. The role of capacity development in enabling better developmental outcomes as well as sustainability has kept investments in capacity development as a key ingredient of developmental interventions. The UNDP defines capacity development as, ‘the process through which individuals, organizations and societies obtain, strengthen and maintain the capabilities to set and achieve their own development objectives over time’ (UNDP, 2008). The approach *per se* is multidimensional as well not limited to individuals but extends to organizations and societies. However, training has become the mainstay of capacity building interventions and the primary instrument for capacity development. More often than not, the terms capacity development and trainings are interchangeably and trainings have become synonymous to capacity development with the scope of the interventions being limited to individuals. While there has been increased focus on capacity development, measurement of effectiveness of the interventions has been a challenge (LaFound and Brown 2003, Simister and Smith 2010).

The glossary of training terms defines training as “a planned process to modify attitude, knowledge or skill behavior through a learning experience to achieve effective performance in any activity or range of activities. Its purpose, in the work situation, is to develop the abilities of the individual and to satisfy current and future manpower needs of the organization”(MSC 1981). Therefore investments in trainings can only be justified if it results in achievement of desired performance. And thus it becomes necessary to evaluate trainings so as to assess whether the intended objectives of the trainings have been achieved, resource use has been efficient and learning’s been internalized.

Evaluation of training is defined as “*the assessment of total value of training system, training course, or programme, in social as well as financial terms. Evaluation differs from validation in that it attempts to assess the overall value of the course or programme, and not just the achievement of its laid-down objectives. The term is also used in the general judgmental sense of the continuous monitoring of a programme or of the training function as a whole*”. The definition emphasizes that the overall process of the training is to be evaluated and not only just the laid down objective of the training. This will substantiate further improvements in the overall training process.

Despite capacity building playing such an important role, evaluation of trainings is not being done consistently. This might be due to factors such as lack of expertise,

insufficient budget and time allocation, lack of tools and methods etc. (Eseryel 2002) and evaluation of training practices are still new and experimental (LaFond, 2003). This paper details the various approaches to training evaluation and describes the use of Kirkpatrick model of training evaluation in a health sector intervention. It describes the methodology, tools and analyses protocols employed for the evaluation and highlight the challenges in application of the model to comprehensively evaluate training interventions.

APPROACHES TO TRAINING EVALUATION

Starting from the generic Analyze, Design, Development, Implement and Evaluate-ADDIE model of instructional design (Branson et al 1975), evaluation of training is essential component of all instructional design models. Evaluation of trainings encompasses various components such as evaluation of participant's learning and reaction to the training, evaluation of training materials, transfer of training, returns of investment etc. (Eseryel, 2002). Based on these components various evaluation models have been established. Some of the approaches to evaluation of training summarized by Eseryel 2002 are:

- Goal based evaluation: Which begins with a goal in mind and determines to seek whether the goal was obtained
- Goal-free evaluation does not determine any outcome or goal but seek whether any benefit resulted from the intervention
- Responsive evaluation is based on client requirements
- Systems evaluation focuses on whether the intervention was effective and efficient
- Professional review which practices external expert to evaluate
- Quasi-legal

Of all these approaches, Goal based evaluation and systems evaluation are the most commonly utilized for evaluating trainings. Some of the common goal based models are described hereby:

Donald Kirkpatrick's 4 level

This is the most commonly utilized model for evaluation of training since 1950s. This model has four essential levels that are described in the next section where we elaborate on the model and further showcase an evaluation study using the Kirkpatrick's model.

Jack Phillips Return of Investment (ROI)

Phillips have modified the Kirkpatrick's model by adding the component of cost benefit ratio of training to measure the monetary gains of the organization as a result of training programme (Phillips 1997).

Nine Outcome Model

The Nine Outcome model of Donovan and Townsend is an adaptation of the Kirkpatrick's model (Donovan and Townsend 2004). It assesses the training on the following nine outcomes:

- Reaction to training- did the participants like the training?
- Satisfaction with the organization of training events- was the logistics satisfactory?
- Knowledge acquisition- did the participants learn?
- Skills improvement- did they acquire a new skill?
- Attitude shift- did the opinion of the participant change?
- Behavior change- did the trainees modify their behavior as a result of training?
- Organizational result- did trainees' performance improved organizational results?
- Return on Investment- was there any cost benefit as a result of the training?
- Psychological capital- did the training effect corporate image?

Some of the systems based models are:

Context, Input, Process and Product (CIPP) evaluation

This model was developed by Daniel Stufflebeam in 1971 which includes the following four levels:

- *Context evaluation*- which helps in planning and developing objectives
- *Input evaluation*- which determines the design by assessing the capability, resources available and different strategies
- *Process evaluation*- which improves the training process by giving simultaneous feedbacks
- *Product evaluation*- which measures the training achievement in terms of output and outcome.

Similar to CIPP model, the Input, Process and Output (IPO) model and Training Valuation System (TVS) model of evaluation of training addresses needs assessment or situation analysis along with outcome of the training programme in terms of monetary gains. These models are an adaptation or an extension to the Kirkpatrick's model as well.

THE KIRKPATRICK MODEL OF EVALUATION

Kirkpatrick's model of training evaluation have set the gold standard foundation for many following training evaluation models. Kirkpatrick's model of evaluation was formulated by Donald Kirkpatrick in 1959. This model was based on the goal-based evaluation approach and is described by four levels of training evaluation (Kirkpatrick and Kirkpatrick 1994). The four levels are:

Level 1-Reaction: This measures the satisfaction and the perception of the trainees on the overall training process, classroom and practice sessions conducted, knowledge of the trainers, peer learning and logistics arranged. Reaction of the participants is assessed to improve the quality of the forthcoming training sessions.

Level 2-Learning (Knowledge and Skills): This measures the learning of the participants as a result of the training session. This is imparted as a pre-post test assessments on various training topics to assess the learnings gained due to the training.

Level 3-Behavior: This measures transfer or behavior change of the trainees in terms of capability improvement or application of techniques learned during training into their workplace. This level will help refine the training process further.

Level 4-Result: This measures the impact of the training on the organization/institution as a result of performance improvement of the trainees.

The first two level of Kirkpatrick's evaluation model, Reaction and Learning are a part of formative evaluation focusing on improving the training process whereas the last two levels of Behavior and Result are part of summative evaluation which helps understand the effect and impact of the training conducted.

Kirkpatrick's model of evaluation is widely and commonly used to evaluate training as it's simple, easy to administer and not time consuming. As discussed above, Level 1 and Level 2 are practiced commonly by many organizations, but the results based on these two levels are not concrete in measuring whether the training programme was effective and achieved the intended goal.

Level 3 of the Kirkpatrick's module is the most important in terms of utilizing the knowledge and skills imparted during the training, but the model does not identify the mechanism of capturing this data. Also, the results at level four are not only because of the trainings and therefore distilling the effect of trainings is difficult.

APPLICATION OF THE KIRKPATRICK MODEL: CASE STUDY OF BTAST TRAINING ASSESSMENT

The Government of Bihar (GoB) has launched the “Sector Wide Approach to Strengthening Health” (SWASTH) programme with financial support from DFID (UK). The goal of *SWASTH* is ‘to improve the health and nutritional status of people in Bihar, particularly the poorest and excluded’. DFID support includes provision of technical assistance and a Bihar Technical Assistance Support Team (BTAST) has been set up. BTAST is managed by a consortium of CARE (UK), Options Consulting and IPE Global and consists of national and international consultants to provide technical and managerial support to the state government.

As part of the imperatives, trainings – especially the modules that emphasise on skill building and augmentation – are expected to lead to capacities which in turn should have a positive impact on health service delivery mechanisms. In this regard, Sambodhi Research and Communications was engaged by BTAST for independent assessment of the quality of training in the areas of seven key training programmes conducted by GoB. These being:

- Skilled Birth Attendant (SBA) training for Auxiliary Nurse and Midwife (ANM) and Local Health Volunteer (LHV)
- SBA training for Staff Nurses (SN)
- Integrated Management of Neonatal and Childhood Illness (IMNCI) Training for Health Workers
- Facility-based IMNCI training for Medical Officers (MO)
- Basic Emergency Obstetric Care (BEmOC) training for MOs
- Comprehensive Emergency Obstetric Care (EmOC) training for MOs
- Life Saving Anesthetic Skills (LSAS) for MOs

It was envisaged that the findings of the study would provide the GoB with much needed evidence-based documentation on quality of training and help take appropriate actions at all levels. The study was conducted in June-October 2013.

Methodology

The overarching framework for the study was the Kirkpatrick model of training evaluation. However, scope of the study entailed assessment of the training implementation process and training results—in terms of Reaction and Learning. The temporal scope of the study was of training imparted by GoB during the year 2011-12, 2012-13 and being imparted during the study year i.e. 2013-14. This entailed both on-site and off-site assessment. As regards the spatial scope, of the total 38 districts in the state, the study covered 20 districts.

Of the total 38 districts in the state, 20 districts were selected for the study. The first criterion for selection of districts was—the district being a High Focus District. The state has 10 districts classified as High Focus and therefore all of these were first selected. The second criterion for selection was representation of the nine regions of the state. The selection of districts is describes in the table 1.

Table 1: Selection of districts

Region	No. of Districts	No. of Districts selected	Name of district selected
Patna	6	3	Patna, Bhojpur, Kaimur
Tirhut	6	3	East Champaran, Sitamarhi, Seohar
Saran	3	1	Gopalganj
Darbhangha	4	2	Darbhangha
Kosi	3	1	Samastipur
Purnia	4	4	Purnia, Katihar, Araria, Kishanganj
Bhagalpur	2	1	Bhagalpur
Munger	5	3	Munger, Jamui, Lakhisarai
Magadh	5	2	Gaya, Aurangabad
TOTAL	38	20	

Based on the number of trainings conducted for each programme by the Government of Bihar, sample size for data collection was decided for each of the training. The sample size for the various training programmes is given in table 2 below.

Table 2: Overall and per district sample

Type of training	Sample per District				Total for all districts
	Trainees		Non-trainees	Total	
	On site	Off site			
SBA Training (ANM & LHV)	5	10	5	20	400
SBA Training (Grade A Nurse)	5	10	5	20	400
IMNCI for health workers	5	10	5	20	400
F-IMNCI for MO	1	1	1	3	60
BEmOC for MOs	1	1	10 (total)	2.5	50

Type of training	Sample per District			Total for all districts	
	Trainees		Non-trainees		Total
	On site	Off site			
CEmOC for MOs	8 (total)	10 (total)	10 (total)	-	28
TOTAL	30	60	30	84	1338

As is seen, with and without assessment was employed as mandated by the Kirkpatrick model so as to evaluate the net effect of training.

The following tools were employed for the study.

- a) Participant Questionnaire: The same comprised of 3 sections:
 - Reactions: For assessing trainee reaction to the training
 - Knowledge: For assessing transfer of knowledge with respect to training objectives
 - Skills: For assessing in change in skill levels incidental to the trainings
- b) Trainers Checklist: For in-depth interview (IDI) of the trainers
- c) Training Manager Checklist: For IDI of district nodal responsible for organizing trainings

The draft tools were developed and discussed with BTAST, soliciting inputs. Upon incorporation of the inputs from BTAST, the tools were piloted so as to assess their sufficiency in answering the research questions and then finalized.

Assessing Level 1 of Kirkpatrick's model- Reaction

The first level of the model is participant reaction and the level was assessed using a structured questionnaire having 20 questions. The response was captured in bipolar scale. The parameters on which the participants reactions were captured included: expectations being met, alignment of contents and objectives of training, logical sequence of contents, adequacy of material given, knowledge of resource persons, trainings methodology employed etc. For each of the parameter, a bipolar scale was constructed using most positive and most negative statements describing the parameter. Given below is a section of the questionnaire that describes the employed scale while capturing reactions. The same questionnaire was used for all the trainings.

Were the contents of the training given in a logical sequence?

5	4	3	2	1
Flowed in a sequence				Totally haphazard

Were the resource persons facilitating the training sessions knowledgeable?

5	4	3	2	1
Had thorough knowledge				Not at all knowledgeable

Did the resource persons encourage participation of trainees?

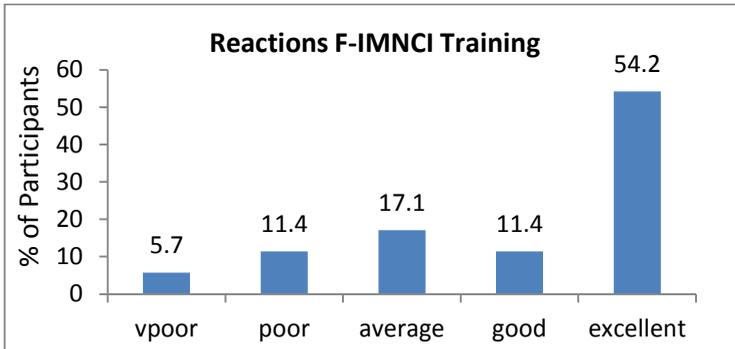
5	4	3	2	1
Very much participatory				Not at all participatory

Did you get sufficient time to practice the skills focused in the training?

5	4	3	2	1
Got sufficient practice				No practice at all

To provide a summative picture of the reactions, Principal Component Analysis³ (PCA) was employed. The index values developed by PCA were categorized into five categories. The lowest 20 percent was categorized as very poor and the highest 20 percent as excellent, with other quintiles classified as poor, average and good. The first two quintiles were considered as participants who well received the training. The analysis of reactions from the F-IMNCl training is given in the figure below. As is seen the frequency distribution of the various categories reflects the reactions of the participants on the training.

³ Calculation of the index value required aggregation of measurements of various impact indicators defined for a particular activity. The most reliable method for developing index is the statistical technique of Principal Component Analysis (PCA). PCA is essentially a data reduction technique that is widely employed to develop summative measures in multi-indicator scenarios. PCA has been used worldwide across various developmental sectors for construction of indices. PCA summarizes variability among a set of variables at different levels and assign statistically derived weights to each indicator following under a particular criteria. It is proposed to employ the same for developing the index. PCA seeks to describe the variation of a set of multivariate data in terms of a set of uncorrelated linear combination of the original variables, where each consecutive linear combinations is derived so as to explain as much as possible variation in the original data, while being uncorrelated with other linear combinations. The index is typically assumed to be the first principal component, i.e. the first linear combination.



Assessing Level 2 of Kirkpatrick's model- Acquired knowledge

The knowledge component of the trainings was assessed by structured test having 20 questions. Multiple-choice questions were used while assessing the knowledge. A section of the test for SBA training is presented here as example. For each of the trainings to be assessed, separate tests were developed based on the key knowledge elements described as objectives of a given training.

- Susheela is 24 years old. She comes to you in March and tells you that she is 5 months pregnant. She says that her last period started a day before Diwali (October 18). Her due date is:

- a. July 17
- b. July 23
- c. July 24
- d. July 25

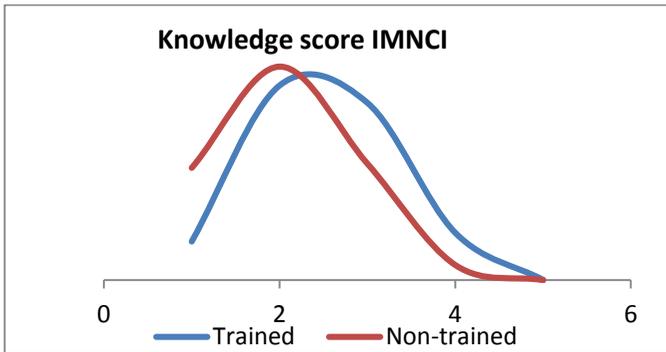
- The second stage of labour begins and ends with:

- a. Onset of labour pains and half dilatation of cervix
- b. Onset of labour pains and full dilatation of cervix
- c. Full dilatation of cervix and delivery of baby
- d. Full dilatation of cervix and delivery of placenta

- The dose and route of oxytocin for the initial management of PPH, before you refer the woman to the FRU, are:

- a. 20 IU, intramuscular stat
- b. 15 IU in 500 ml of Ringer lactate, intravenously
- c. 5 IU, intramuscular stat
- d. 20 IU in 500 ml of Ringer lactate, intravenously

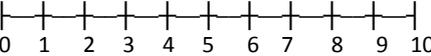
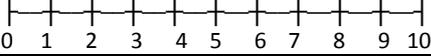
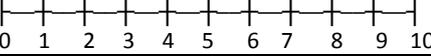
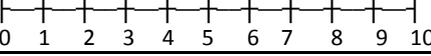
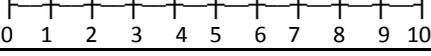
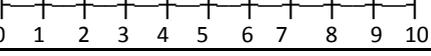
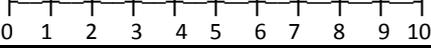
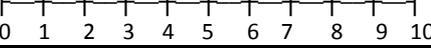
Frequency distributions of the scores were developed for trained and non-trained respondents. The difference in the distribution aided in assessment of training effectiveness. The same obtained for IMNCI training is given in the figure below.



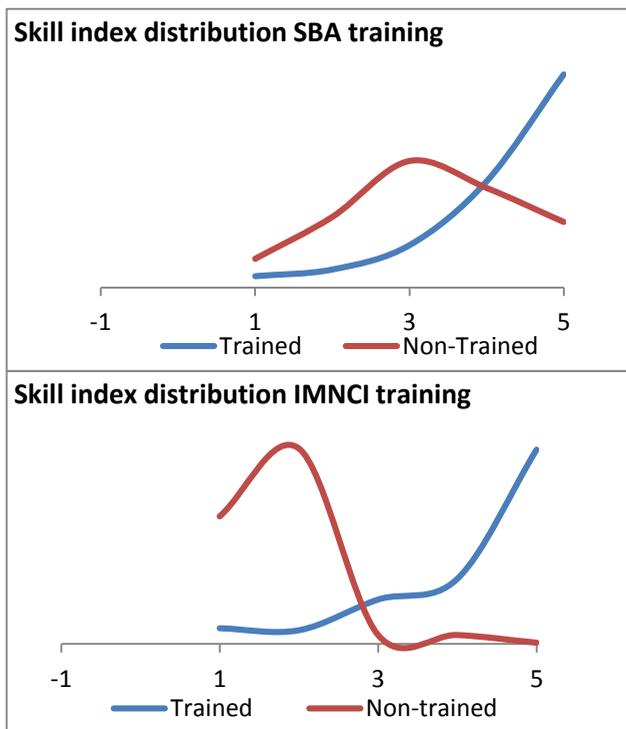
Descriptive of the score along with the distribution were calculated to comment on the level of knowledge and t-test for difference of means done to comment on statistical significance of the observed difference. The knowledge mean score for the IMNCI trained and non-trained was calculated and a difference of 1.7 was observed between the two groups though the dispersion is identical. The difference in the score was also found to be statistically significant (t-test, one tail, $p=0.000$) at 95 percent confidence interval.

Assessing Level 2 of Kirkpatrick's model-Acquired skill

The skill levels of the participants were ascertained through self-administered course review questionnaire. Based on the skills focused in a given training, a questionnaire was developed where the respondent has to rate its effectiveness in performing the skills on a scale of 1 to 10. The rating was also done for the pre-training level. The number of skills probed varied from 15-25 depending on the trainings. A sections of the course review questionnaire employed for BEmOC training of MOs is given here.

<p>Needs Practise Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>	<p>Identify and management of differnt complications of pregnancy</p>	<p>Not at all Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>
<p>Needs Practise Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>	<p>Preparing delivery trolley/ equipment</p>	<p>Not at all Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>
<p>Needs Practise Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>	<p>Perform PV examination</p>	<p>Not at all Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>
<p>Needs Practise Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>	<p>Monitor labour, plot & interpret Partograph</p>	<p>Not at all Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>
<p>Needs Practise Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>	<p>Conduct normal delivery</p>	<p>Not at all Moderately Very Effective</p>  <p>0 1 2 3 4 5 6 7 8 9 10</p>

A summative index, similar to reaction was developed employing PCA and quintiles developed, the top two quintiles reflecting skilled participants. The summative comparison was also made between trained and non-trained so as to comment on the effect of training. Shown below are the comparative analyses from SBA and IMNCI trainings. The difference in frequencies of the trained and non-trained in various skill categories (4 and 5 depicting skilled) reflect on the effect of training on the targeted skills.



DISCUSSION

As discussed in the preceding sections, it is evident that the assessment of the first two levels of the model is quite easy as the same could be done during the training and at the site only. Well-designed instruments can comprehensively capture the participants reactions and learning in a given training programme. Also, having a non-trained group for comparison of acquired learning, inferences can be comprehensively drawn on the quality of training and therefore learning.

While the learning assessment is simple in terms of knowledge elements of a training programme, assessment of skills is difficult. As assessment of skills would require actual demonstration of the skills and in many cases the demonstration may not be possible. Further, the nature of skill e.g. inter personal skills may make the assessment difficult.

Further, assessment of third level of the model-behavior would require workplace observation of actual performance of the task in real work condition. And for this the evaluator has to be present at the work site. Also, it may be possible that the task may not get performed during the presence of the evaluator. Thus, actual assessment of behavior becomes difficult.

However, the most challenging is the measurement of results, the foremost challenge being the time of training and time at which the results would be visible. While knowledge and skills and practice may be necessary conditions for a particular results, it may not be the sufficient conditions. Other things may also be required for practice of skills to translate into organizations results. And therefore, even if the results are measured at a later time, attribution of the results to the trainings is difficult even after having a with-without scenario.

Thus, it can be said that the model is relatively efficient for measuring the on-site effects of training in terms of perception of training and transfer of learning's in classroom conditions. However, its use for comprehensively assessing translation of training to workplace behavior and subsequent results is limited because of involved complexities.

CONCLUSION

Given the focus on trainings as a vehicle of effective and sustained change, training evaluation becomes quintessential so as to justify the investments made therein. There are various methodological options and approaches for evaluation training each with their own applicability, advantages and limitations. Of the various models of training evaluation, Kirkpatrick's model is the most subscribed as it covers the entire continuum-from participants reaction to translation of training into organizational results. While the model presents a robust framework for assessment of training, assessment of all the four levels to comprehensively comment on the effectiveness of training using the model in practice is challenging. While the first two levels of assessment are relatively easy, the complexities involved in assessing the third and fourth level make it difficult to use especially in large scale interventions. Use of the model for on-site assessment of quality of

trainings at the training site is recommended. However, given the complexities in assessing third and fourth level of behavior and results, other methodological options or combinations may be effective alternatives.

REFERENCES

Branson, R. K., Rayner, G. T., Cox, J. L., Furman, J. P., King, F. J., Hannum, W. H. (1975). *Interservice procedures for instructional systems development*. (5 vols.) (TRADOC Pam 350-30 NAVEDTRA 106A). Ft. Monroe, VA: U.S. Army Training and Doctrine Command, August 1975. (NTIS No. ADA 019 486 through ADA 019 490).

Donovan, P. and Townsend, J. (2004). *The Training Evaluation Pocketbook; Management Pocketbooks*; Hants, UK; ISBN 1903776236.

Eseryel, Deniz (2002). *Approaches to Evaluation of Training: Theory and Practices*; Educational Technology & Society, v5 n2 p93-98 Apr 2002; Retrieved from <http://ifets.ieee.org/periodical/issues.html>.

Kirkpatrick, D.L., and Kirkpatrick, J.D. (1994). *Evaluating Training Programs*, Berrett-Koehler Publishers; Oakland, U.S.A .

LaFound, Anne and Brown, Lissane (2003). *A Guide to Monitoring and Evaluation of Capacity Building Interventions in Health Sector in Developing Countries*; MEASURE Evaluation Manual Series, No.7, Carolina Population Center, University of North Carolina at Chapel Hill

Manpower Services Commission (1981). *Glossary of Training Terms*. London: HMSO

Nigel Simister with Rachel Smith (2010). *Monitoring and Evaluating Capacity Building: Is it really that difficult?* Praxis Paper 23. INTRAC. Retrieved from <http://www.intrac.org/data/files/resources/677/Praxis-Paper-23-Monitoring-and-Evaluating-Capacity-Building-is-it-really-that-difficult.pdf>.

Phillips, J. Jack (1997). *Return On Investment in training and performance improvement programs (A step-by-step manual for calculating the financial return)*. 1st Edition. PROPress. South Wimbledon, UK.

UNDP (2008). *Capacity Development Practice Note*. United Nations Development Programme 304 East 45th Street New York, NY 10017 USA. Retrieved from http://www.unpdc.org/media/8651/pn_capacity_development.pdf.

Notes



South Asia

C – 126, Sector 2,
Noida – 201301,
Uttar Pradesh, India
Phone: 0120-4056400

South-East Asia

#132C, Street 135, Sangkat Thkov,
Khan Chamkarmorn, Phnom Penh,
Cambodia, Phone: +855 81738017

Sub-Saharan Africa

1st Floor, Acacia Estates
Building, Kinondoni Road,
Dar-e-Salaam, Tanzania
Phone: +255 787894173