RUNNING HEAD: MCNAMARA - EVALUATION JOB AID

Evaluation Job Aid

Produced for the Benefit of Instructors at

National College of Business & Technology, Knoxville

Prepared for

EDUC 527

Tusculum College MA018

Prepared by

Billie R. McNamara

November 7, 2005

Contents

List of Figures

List of Tables

Identifying & Addressing Performance Gaps

The gap between vision and current reality is also a source of energy.
If there were no gap, there would be no need for any action to move
towards the vision. We call this gap creative tension. ~ Peter Senge

Organizations determine a need for employee behavioral modification as a direct result of employees' actions that deviate from organizational protocols and procedures. Blanchard & Thacker (2004) refer to this determination point as a "triggering event." The "triggering event" gives rise to a formal or informal Needs Assessment to determine what action should be taken, if any, to modify the employees' behavior. Blanchard & Thacker (2004) show the effects of the "triggering event" as a stimulus to a Needs Assessment in their *Training Processes Model*, presented in Figure 3 (p. 22).

Once it identifies a performance gap through a Needs Assessment, an organization has three options:

1. If the performance gap is not important or not worth modifying, ignore it.

2. If the performance gap **is** the result of a deficiency in knowledge, skills, or abilities that can be addressed by training, arrange a training solution.

3. If the performance gap **is not** the result of a deficiency in knowledge, skills, or abilities, apply other Human Resource Development solutions, such as creation and implementation of a Job Resource Aid.

Blanchard & Thacker (2004) have a slightly modified viewpoint. They describe the basic method for determining a behavioral or learning gap as a more-specific series of questions and responses, as shown in Figure 1. Their recommendations for addressing a gap in knowledge, skills, or abilities are shown in Figure 2 (p. 118).

Figure 1. Method for Determining a Behavioral or Learning Gap

Performance Discrepancy

↓

YES

↓

Is it worth fixing?

↓

YES

KSA Deficiency	Reward/ Punishment Incongruence	Inadequate Feedback	Obstacles in the System
YES	YES	YES	YES
Choose Appropriate Remedy	Change Contingencies	Provide Proper Feedback	Remove Obstacles

Figure 2. Addressing KSA Gap

Job Aid

Training

Practice

Change the Job

Transfer or Terminate

Figure 3. Blanchard & Thacker's *Training Assessment Model*

Needs Analysis Phase
Input → Process → Output

Design Phase
Input → Process → Output

Development Phase
Input → Process → Output

Triggering Event

Implementation Phase
Input → Process → Output

Process Evaluation Data

Outcome Evaluation Data

Evaluation Phase
Input → Process → Output

Planning and Implementing Training

> *If you don't know where you are going, you will probably
> end up somewhere else. ~ Lawrence J. Peter*

The first step in planning training is design: identify the goals and objectives.

The next step is development: map out the lessons, identify appropriate teaching methods, and acquire or create instructional media. Once training has been planned, it can be implemented.

Blanchard & Thacker (2004) mapped the planning and implementation process, as shown in Figure 4 (p. 179). They emphasize the necessity for setting objectives for evaluation of the training event, not just evaluation of the transfer of learning.

Throughout the planning and implementation phases, curriculum designers and instructional technologists must evaluate and revise at each major step. Goals and objectives may need revision if learners are unable to assimilate the material and apply it ("transfer of learning"). Lesson plans, methods, and media may require modification if they fail to deliver the necessary information and/or skill development opportunities to learners. Training implementation may require revision if the training location, instructor, or other classroom-specific feature fails.

Figure 4. Planning & Implementation Process (Blanchard & Thacker, 2004)

Input	Process	Output
Learning Theory		
Training Needs	**Develop Training Objectives**	**Determine factors that facilitate learning & transfer**
		Identify alternative method of instruction
Organizational Constraints		**Evaluation of objectives**
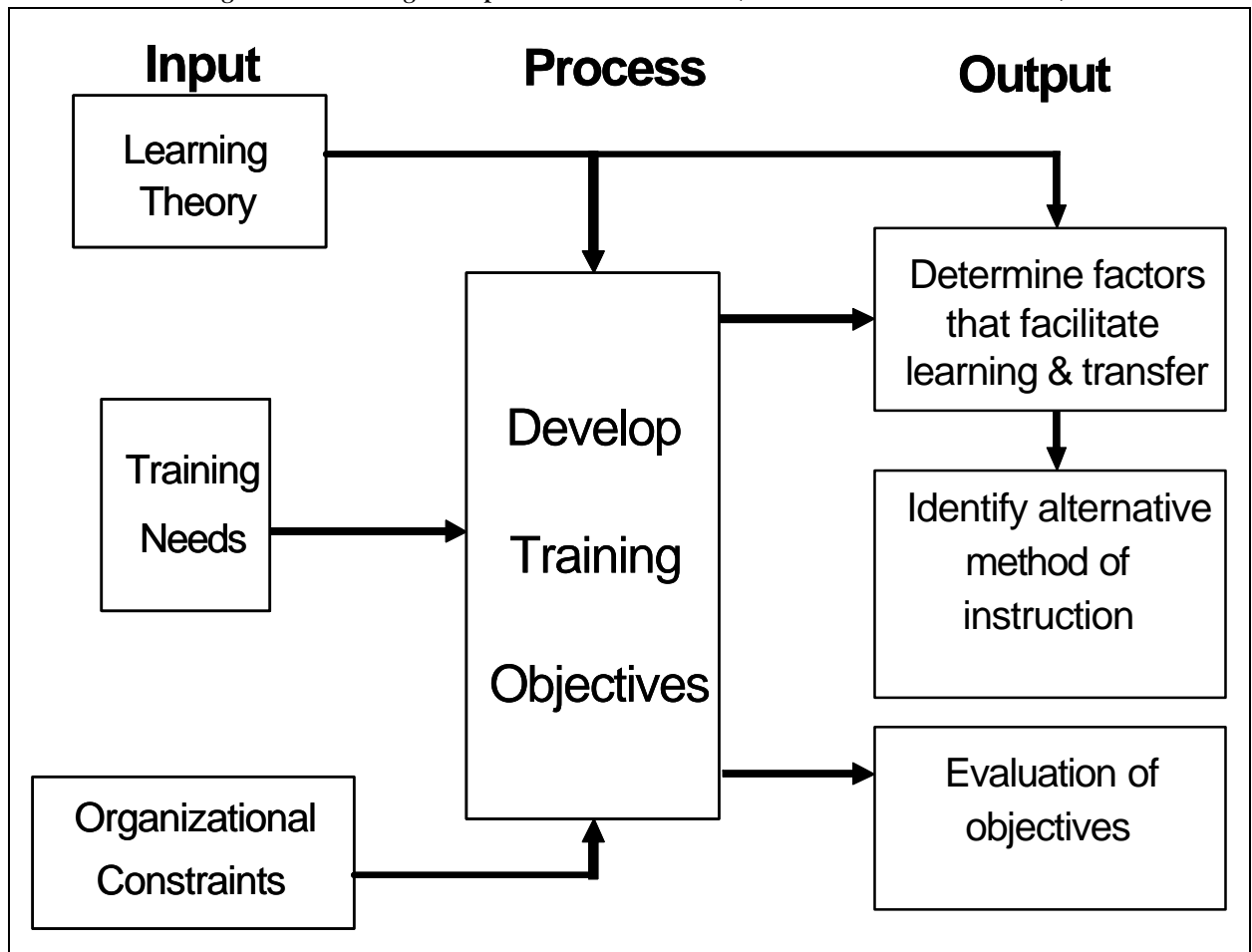
Evaluating Training

> *I know not any thing more pleasant, or more instructive, than to compare experience with expectation, or to register from time to time the difference between idea and reality.* ~ *Samuel Johnson*

Evaluators can not fully evaluate a training program unless they know what the training was developed, designed, and implemented to accomplish. Training objectives, derived from goals, are established to

- describe the performance learners should exhibit after training;

- measure learner performance and competency;

- determine what gets taught and how training occurs; and,

- provide methods to evaluate success of the training.

Training objectives are action statements containing descriptive words open to limited interpretation and conveying explicit intent. Objectives must be specific, measurable, attainable, realistic, and time-specific. A training objective has four parts:

1. Performance - a statement describing what the learner should be able to do;

2. Conditions - circumstances under which the performance occurs;

3. Criteria(on) - definition of an acceptable level of performance; and,

4. Audience - the targeted learners.

   Example: At the conclusion of training, trainees will prepare a standard departmental office supply order form containing five items with 100% accuracy in less than ten minutes.

Most training courses have a number of objectives, and the overall success or failure of a training course is based on measuring whether learners meet those objectives. Evaluators must ask questions such as these:

- Were training program objectives met?

- Were learners' personal objectives met?

- What specific information or skills did learners receive or reinforce?

- What level of commitment did learners show to applying new skills and information on the job?

- What level of retention by learners was evident from their behavior (modified or not) upon returning to the job?

- What successes or failures in behavioral modification were noted by supervisors when employees returned to their jobs?

- What level of Return on Investment (ROI) in the training activity has the organization experienced , either in terms of increased satisfactory performance or actual financial gain?

Evaluating the Training Process

As stated above, all evaluative determinations can not wait until training is completed. Clark (2005) maintains the entire training process must be evaluated at each step of the planning, preparation, delivery, and even evaluation of training. He describes two types of evaluation: "Assessment is the measurement of the practical results of the training in the work environment; while validation determines if the objectives of the training goal were met."

Blanchard & Thacker (2004) describe an "open systems design," which is graphically represented in Figure 5. Effective training evaluation is an open system, because the evaluation is constantly under revision, as put forth by Clark (2005).

Figure 5. "Open Systems Design"(Blanchard & Thacker, 2004)
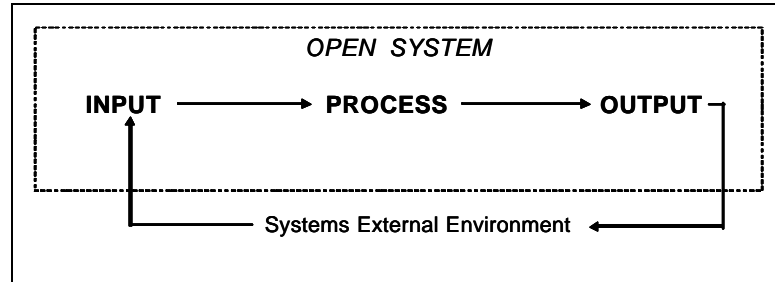
The Open System approach differs from the one proposed by Donald Kirkpatrick (1975) and cited widely as a classic evaluation paradigm. In Kirkpatrick's model, shown in Figure 6, evaluation of training only begins at the Implementation (delivery) phase (Marshall, 2005). Evaluation at this point is summative, or external to the training event. It gives no consideration to the preliminary phases of Analysis, Design, or Development. Rather, such evaluation is only concerned with measuring transfer of learning as evidenced by learner responses, observable changes in job behavior, and organizational Return on Investment.

Figure 6. Kirkpatrick's Four Levels of Evaluation (Marshall, 2005)
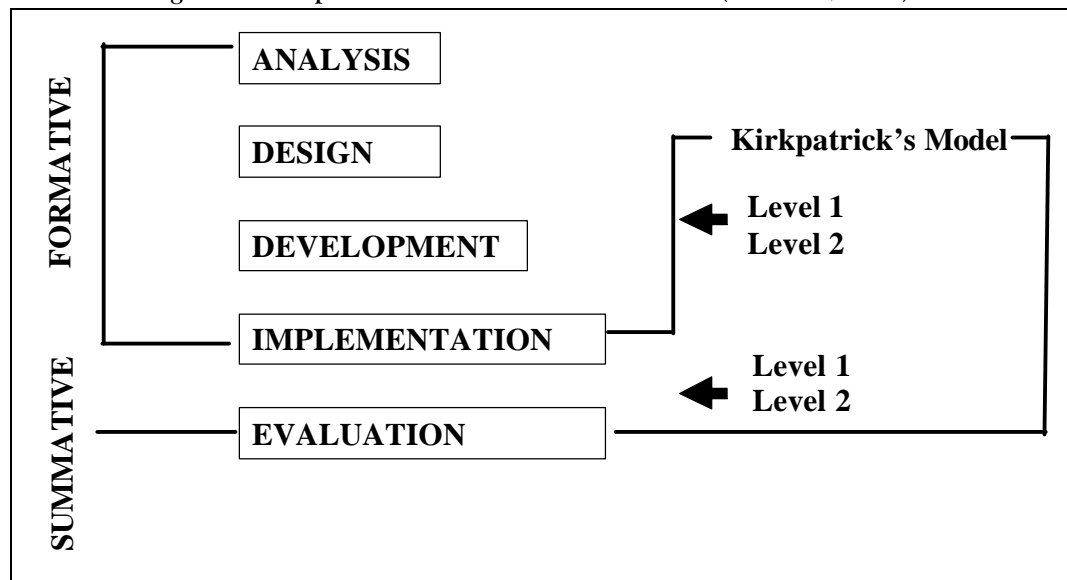
Marshall (2005) supports Clark's position that formative (internal) evaluation is necessary at each phase of the training event in order to ensure the questions raised previously (page 10, *supra*) are answered timely and in the proper sequence. Blanchard & Thacker (2004) published a graphic representation, shown in Figure 7, identifying the influence of formative evaluation on the analysis, design, development, implementation, and evaluation of a training event (p. 344).

Figure 7. Influence of Formative Evaluation on Training Event (Blanchard & Thacker, 2004)
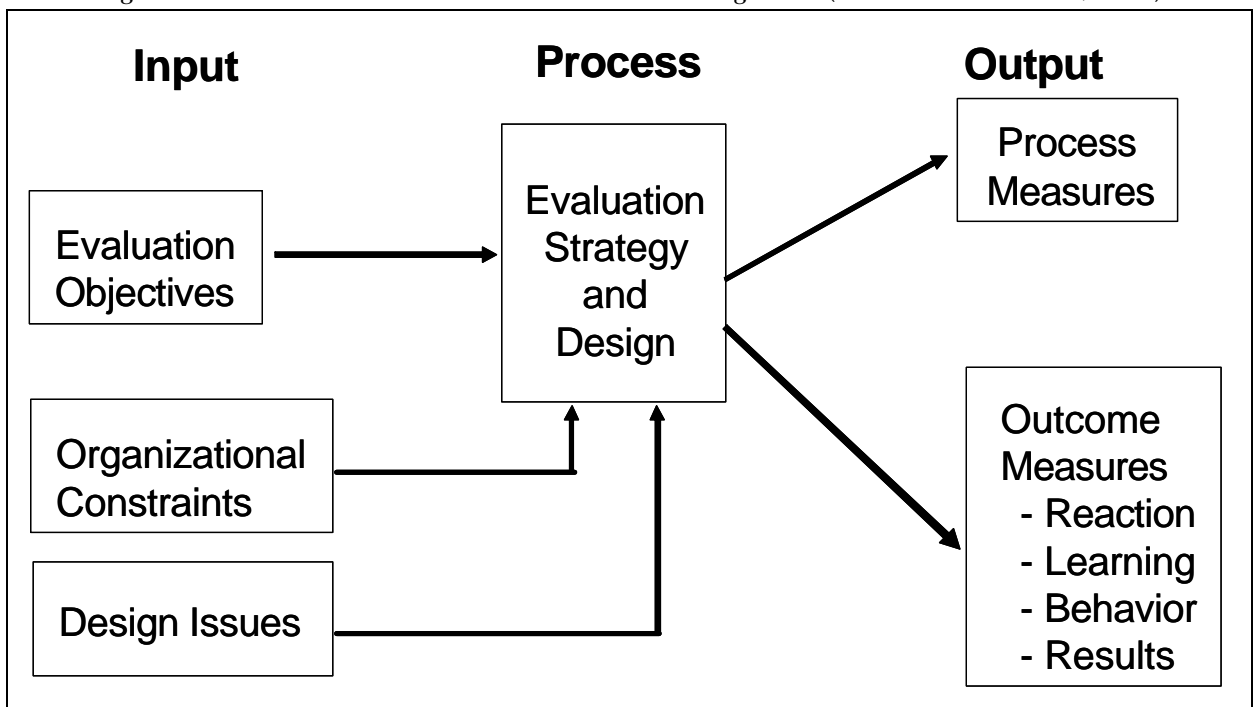
Phillips, Phillips & Hodges (2004) discuss time requirements, but they only address summative evaluations performed after the training event. Figure 8 shows a sample timetable for evaluation these authors derived from their research. They write,

> "Planning for a major study should take no more than a day, and the design of the instruments can vary depending on the number of instruments and method of data collection. Data collection is the most variable part of the analysis. Simple data collection instruments, such as questionnaires, are inexpensive and involve less time to implement. Other instruments, such as observation, interview, and focus groups are more time consuming and can add considerably to the time estimates to this table."

Figure 8. Approximate Time Needed for Evaluation (Phillips, Phillips & Hodges, 2004)

Level of Evaluation	Approximate Time Requirement (in hours)	
	Manual	Automated
Reaction and Satisfaction (Level 1)	1-2	<1
Learning Assessments and Simple Tests (Level 2)	2-6	<1
Application and Behavior Changes (Level 3)	18-24	3-5
Impact Analysis (Level 4)	36-54	4-8
Return-on-Investment (Level 5)	48-80	12-18

Notes:
Assumes that higher-level evaluation includes data at lower levels
Excludes data collection, analysis, and reporting
Excludes time necessary for planning, data integration across programs,
   and meetings to present data

Phillips, Phillips & Hodges (2004) also discuss the time required to establish evaluation

protocols and procedures. Their observations include the following:

•   Evaluation processes should be automated as much as possible.

•   Many computer software applications are available for managing large

         projects.

•   Designs can require an inordinate amount of time and resources.

•   The design process should be automated, if possible.

•   Evaluators should use standardized tools and templates, if possible.

•   Organization impact and ROI analyses should be limited to training

         considered "significant, critical, expensive, highly visible, and time

         consuming," thus reducing the possibility of extensive drains on

         organizational resources.

Phillips, Phillips & Hodges (2004) also propose a structured *Timeline*, shown in Figure 9, for developing an organized, functional evaluation system within an organization or department. Their recommendation is to spend one year developing the system. Unfortunately, most small organizations do not have the resources to spend that much time developing a model because evaluation results are needed relatively quickly.

Figure 9. Timeline for Establishing an Evaluation System (Phillips, *et al.*, 2004)

Evaluation Activity	1	2	3	4	5	6	7	8	9	10	11	12
Design and Pilot Reaction Survey	■	■										
Construct and Validate Learning Measures	■		■									
Research Vendor Database Support		■										
Customize Vendor Tool		■										
Design Internal Database				■	■	■						
Integrate with Internal System	■	■	■	■								
Select Studies for Impact Analysis				■								
Conduct Impact Studies					■	■	■			(average time per study)		
Design Reporting Format								■				
Plan for and Communicate Results										■	■	■

Month

As a result of their research, Phillips, Phillips & Hodges (2004) also developed a budget calculation form for training Evaluators' use, shown in Figure 10. These authors note that total cost of measurement and evaluation could be as much as 3% to 5% of the training budget, especially in those cases where data are collected from every course and several programs are carried through to a comprehensive analysis of Return on Investment. This is the typical training scenario at NCBT.

Interestingly, Phillips, Phillips & Hodges (2004) estimate the total direct and indirect costs incurred in processing one student's responses can be as much as $2.00 per questionnaire. For a small organization, such costs generally prohibit extensive data analysis. Yet, the authors conclude the "value of an effective, efficient, state-of-the-art evaluation system is priceless."

Figure 11 shows a breakdown of estimated, representative costs associated with training evaluation using Phillips, Phillips & Hodges' (2004) model.

Figure 10. Budgeting for Evaluation (Phillips, Phillips & Hodges, 2004)

Evaluation Item	Cost
Salaries and Employee Benefits for L&D Staff (No. of staff x avg. salary x employee benefits factor x no. of hours on evaluation project)	
Meals, Travel, and Incidental Expenses	
Participant Costs (for time involved and evaluation)	
Fees and Licenses	
Office Supplies and Expense of Printing Materials	
Outside Services	
Technology	
General Overhead Allocation	
Other Miscellaneous Expenses	
Total Evaluation Costs	

Figure 11. Typical Costs for Selected Evaluation Items (Phillips, *et al.*, 2004)

Evaluation Item	Approximate Cost
Total cost of measurement and evaluation as a percentage of learning and development budget--best practice	3%-5%
External cost of processing one reaction questionnaire	$2.00
Application/behavior change study (internal)	$2,000-$3,000
Application/behavior change study (external)	$15,000-$20,000
Impact/ROI study (internal)	$5,000-$10,000
Impact/ROI study (external)	$15,00-$50,000
Cost of an impact/ROI study as a percentage of project costs	5%-10%
Fee to attend ASTD two-day workshop on measurement, evaluation, and ROI	$850*
Fee to attend ASTD one-day workshop on bottomline on ROI	$450*
Cost of SurveyPRO (for questionnaire design and adminstration)	$1,500
Fee for certification in the ROI Methodology	$2,995
Cost to the join the ASTD ROI Network	$75*
Cost of study/reading materials	$62.50
Value of an effective, efficient, state-of-the-art evaluation system	Priceless

* Prices reflect ASTD membership

Four Levels of Training Evaluation

"What is now proved was once imagined." ~ William Blake

Clark (2005) describes training evaluation as having two distinct forms: "Assessment is the measurement of the practical results of the training in the work environment, while validation determines if the objectives of the training goal were met."

In 1959, Donald Kirkpatrick published his seminal dissection of training evaluation. Kirkpatrick's model has four levels, identified below (Clark, 2000):

1. Reaction – measures how participants feel about the training event, including the material, instructor, and environment. Many Level 1 Evaluations now contain demographic questions, so marketing statistics can be gathered in conjunction with training evaluation. Level 1 Evaluation occurs near the end of, or immediately following, training. At NCBT, formal Level 1 Evaluations are given to students twice each year. Many instructors perform informal Level 1 Evaluations in the classroom to determine how students like a particular teaching resource or method.

2. Learning – measures how, and how much, participants change their attitudes, improve their knowledge, and increase their skills and/or abilities as a result of the training event. Level 2 Evaluation occurs during or immediately after the training event. At NCBT, Level 2 Evaluations are exams, quizzes, and competency assessments given to students.

3. Behavior – measures observable changes in behavior resulting from the training event. Level 3 Evaluation traditionally occurs in the workplace several weeks to a few months following training to determine if learners retained and now apply their new or refreshed knowledge, skills, and abilities. In most workplaces, this is done through a performance evaluation. At NCBT, however, students receive a quasi-Level 3 Evaluation on their externships and in clinical environments where they are performing skills without direct instructional assistance.

4. Results – measures organizational results arising from learners' participation in the training event. This is generally referred to as a determination of the Return on Investment an organization can calculate as directly or indirectly resulting from training.
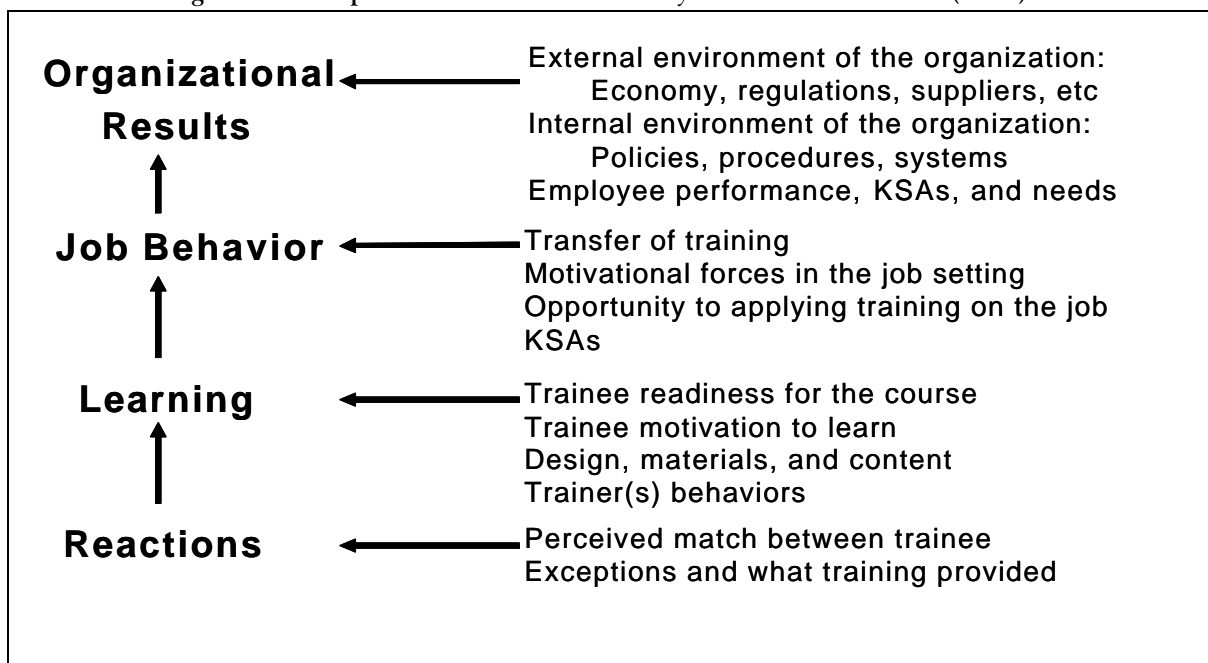
Kirkpatrick's original model is still, more than forty-five years later, the *de facto* standard for evaluating training activities. As discussed previously, Kirkpatrick's model does not provide for formative evaluations, because all of the evaluations in his four levels take place during or after the training event.

Modern models, including those published individually or jointly by authorities such as Scriven, Flagg, Seels, Glasgow, Dick, Carey, and Tessmer (Dabbagh, 2005) and Marshall (2005), incorporate formative evaluations during the analysis, design, development, and implementation phases of training. Marshall (2005) makes an excellent case for the use of formative evaluations at each production step of the training to ensure that all stakeholders – directors, managers, supervisors, trainers, subject-matter experts, and even a sample of the targeted learner population – "buys-in" and endorses the training. Otherwise, the curriculum

designer or instructional technologist may discover the developed product is not in line with

the interests and needs of every stakeholder.

Figure 7 (page 12, *supra*) contains Blanchard & Thacker's (2004) graphical

representation of the training evaluation process. They incorporate Kirkpatrick's model, as

shown in Figure 12 (p. 366).

Figure 12. Kirkpatrick's Model as Refined by Blanchard & Thacker (2004)

Organizational Results ← External environment of the organization:
  Economy, regulations, suppliers, etc
Internal environment of the organization:
  Policies, procedures, systems
Employee performance, KSAs, and needs

↑

Job Behavior ← Transfer of training
Motivational forces in the job setting
Opportunity to applying training on the job
KSAs

↑

Learning ← Trainee readiness for the course
Trainee motivation to learn
Design, materials, and content
Trainer(s) behaviors

↑

Reactions ← Perceived match between trainee
Exceptions and what training provided

Since his original four-level thesis was published in 1959, Donald Kirkpatrick has

continued to revise and update his model. Hundreds of articles, books, and Internet Web

sites discuss the details of Kirkpatrick's four levels. Table 1 contains a short description of

the levels that expands on Blanchard & Thacker's (2004) graphic in Figure 8. Table 1 also

contains representative advantages and disadvantages for each of Kirkpatrick's levels.

Table 1. Kirkpatrick's Four Levels of Evaluation

Level	Definition	Advantages	Disadvantages
1 – Reaction	Measures how well participants have mastered the course objectives. Can include tests of performance before and after the course.	Gives fast feedback on participants' satisfaction with training material, facilitators, and environment.	Measures participants' opinions about the course. This is the most common way to evaluate a course and provides a measure of customer satisfaction.
2 – Learning	Measures how well participants have mastered the course objectives. Can include tests of performance before and after the course.	Compared to reactions, this provides more compelling evidence of whether a training program works.	Compared to Level 1, requires more time, funding, and expertise to develop and implement valid measures of learning.
3 – Transfer to the job	Measures how the knowledge, skills, and values from a course are used on the job. Typically measured 3 to 6 months after training.	Provides stronger evidence that a training investment has the desired impact.	Compared to Levels 1 and 2, requires more time, funding, and expertise to develop and implement valid measures of transfer.
4 – Organizational impact	Measures performance improvement, quality improvements, and cost savings to an organization.	Provides the strongest possible evidence that a training program has the desired impact on an organization.	Substantial levels of investment and expertise are required to implement this level of evaluation successfully.

What Level of Training Evaluation Do You Need?

There are no simple rules for determining which level of training evaluation to use for a particular program. For complex evaluations, the best guidance usually comes from experts in training evaluation. However, the following considerations are an excellent starting point (National Institute of Occupational Safety & Health, 2005):

Question 1: Who will be interested in the results?

- (Critical) Identify all stakeholders – those who will be interested in the results.

- Examples: trainers, managers, organizations, government agencies

Question 2: What questions will be answered?

- Identify questions of particular importance for the specific evaluation – questions one can expect to answer upon completion of the evaluation.

- Examples:

    1. Have people increased their knowledge and/or skills?

    2. Did the information that was learned in training transfer to the workplace?

Question 3: What resources are available for evaluating the training program?

- Determine what resources are currently available and/or what kinds of resources can be obtained easily, if necessary.

- Examples: money, time, personnel, equipment, materials

Question 4: What method(s) will be used to gather information?

- Use a simple form, such as the one below, to organize data as it is collected.

Data Collection Method	Performed (circle choices)
1.	Before, During, or After Training
2.	Before, During, or After Training

Level 1 (Reaction) Evaluation

Background

A Level 1 Evaluation measures participants' emotional and intellectual reactions to training activities and events. At NCBT, as in most training environments, Level 1 Evaluations are commonplace, both in classroom environments and venues such as presentations made to community groups. Often, especially in the context of a for-profit organization such as NCBT, Level 1 Evaluations also request demographic data (*e. g.*, age, ethnicity, education level, zip code) for use in marketing.

In *Evaluating Training Programs: the Four Levels*, Kirkpatrick (1994) provides the following guidelines for creating a Level 1 Evaluation:

- Determine what you want to find out.

- Design a form that will quantify reactions.

- Encourage written comments and suggestions.

The benefits of a Level 1 Evaluation to a curriculum designer, instructional technologist, or training evaluator include the following:

- Get 100 percent immediate response.

- Get honest responses.

- Develop acceptable standards.

- Measure actions against standards, then take appropriate action.

- Communicate reactions as appropriate.

In a training analysis prepared for the U. S. Department of Transportation, Arthur Andersen Co. recommended the following protocols for Level 1 Evaluations that are universally applicable (DOT, 2005):

- 100 % of courses should be evaluated at this level.

- Systematically assess the reactions of training participants and faculty.

- Develop a standard set of questions.

- Develop norms.

- Develop a format for course-specific questions.

- Balance quantitative and qualitative sections of the training questionnaire.

Creating a Level 1 Evaluation

Table 2 contains six steps for developing a Level 1 Evaluation adapted by the U. S. Department of Transportation (2005) from *Diversity Training Evaluation Toolkit*, published in 1994 by the U. S. Federal Aviation Administration.

Table 2. Six Steps for Developing a Level 1 Evaluation	
Step	*Procedure*
Step 1 **Plan**	• Most Level 1 evaluations use a rating scale of 1-5. • This makes it easier to score and compare with other individuals and classes. • The form can also include space for personal reactions to the course or instruction.
Step 2 **Select Tool**	• Based on how much information is needed, select an existing Level 1 evaluation instrument, or create a new form.
Step 3 **Adapt Tool**	• Modify content: • Reword, delete, or add items to accomplish evaluation goal(s). • Modify the format: • Change rating scale items to open-ended questions, or *vice versa*, depending on kind of information desired.

Table 2. Six Steps for Developing a Level 1 Evaluation	
Step	*Procedure*

	• Student reactions:
	  • To obtain general reactions, ask open-ended questions.
	  • To obtain specific reactions to an objective or type of instruction, ask detailed questions.

Step 4
Implement
	• Make enough copies for every student.
	• Assure students that the evaluations will be handled professionally and that student comments remain anonymous.
	  • Students must understand that their answers should be frank and honest, and results will be used to plan future training programs.
	  • When practical, someone other than the trainer should distribute and collect these forms.
	• Make sure every student completes the form.

Step 5
Analyze
	• Rating scales:
	  • For each item, count the number of responses to each rating and find the average.
	• Short answer questions:
	  • Review the comments, categorize them based on similarities, and write a brief summary.

Step 6 **Report**	• Write a short "Class Evaluation Report" summarizing the results (*see* page 40). • Always compute mathematical averages of students responses, even if it is necessary to convert them to numbers. • If someone gives a rating outside the limits (*e.g.*, 10 on a scale of 1 to 5), convert it to the scale (in this case to 5, since that is the highest score). • Always include all student comments, either by attaching copies of the forms themselves or, if time permits, by retyping them. • Send copies of the report to the trainer, the trainer's supervisor, and to the organization that funded the training.

Sample Level 1 Evaluation

Table 3 contains an example of a Level 1 Evaluation form, based on a document created by the U. S. Department of Transportation (2005).

Table 3. Example Level 1 Evaluation Form

Your Name (*Optional*)	Course Name
Trainer's Name	Course Date

Your comments are important to us. Please rate the course you have just completed by responding to the statements in the spaces below. Use 1 for lowest and 5 for highest. Your individual observations will remain anonymous, but the collective observations will be used to plan future training.

Course Evaluation Form					
Observation	*1*	*2*	*3*	*4*	*5*
My level of understanding of the objectives and job relevance of this course before attending it.	o	o	o	o	o
Level of difficulty of the course.	o	o	o	o	o
Pace of the course.	o	o	o	o	o
The degree to which the course objectives were accomplished.	o	o	o	o	o
The degree to which the course met my expectations.	o	o	o	o	o
Quality of class discussion.	o	o	o	o	o
Quality of group activities.	o	o	o	o	o
Relevance of the course to my job.	o	o	o	o	o
Degree to which management and co-workers will support my use of skills and values from the course.	o	o	o	o	o
Consistency of materials and media with course objectives.	o	o	o	o	o
Overall quality of the materials (e.g., written materials, videos, etc.).	o	o	o	o	o
Instructor's ability to deal with conflict in a productive manner.	o	o	o	o	o
Instructor's ability to deal with disruptive behavior.	o	o	o	o	o
Quality of individual feedback given to me.	o	o	o	o	o
Instructor's encouragement of my active participation.	o	o	o	o	o
Clarity of instructor's explanations and instructions.	o	o	o	o	o
Instructor's overall effectiveness.	o	o	o	o	o
Comments					
The most beneficial part of the course was					

The least beneficial part of the course was

My suggestions for improvement are

Level 2 (Learning) Evaluation

Background

The following guidelines were adapted by the U. S. Department of Transportation (2005) from Donald Kirkpatrick's, *Evaluating Training Programs: the Four Levels*:

- Use a control group, if practical.

- Evaluate knowledge, skills, and/or attitudes both before and after the program.

- Use a paper and pencil test to measure knowledge and attitudes, and use a performance test to measure skills.

- Get 100 percent response.

- Use the results of the evaluation to take appropriate action.

In an analysis of non-technical, non-military training performed by the U. S. Department of Transportation entitled *U. S. DOT Training Program Review, Final Report*, Arthur Anderson Co. recommended the following protocols for Level 2 Evaluations that are universally applicable (DOT, 2005):

- 50% of courses should be evaluated at this level.

- Utilize learner assessment on a more limited basis.

- Embed test items into Level 1 questionnaires.

Creating a Level 2 Evaluation

Table 4 contains six steps for developing a Level 2 Evaluation adapted by the U. S.

Department of Transportation (2005) from *Diversity Training Evaluation Toolkit*, published in

1994 by the U. S. Federal Aviation Administration.

Table 4. Six Steps for Developing a Level 2 Evaluation	
Step	*Procedure*
Step 1 **Plan**	• Level 2 evaluation is used to determine if students have reached a satisfactory level of understanding and mastery of the course objectives. • Indications for revisions to the course materials are also revealed. • Talk to all stakeholders about the proposed level of evaluation and the costs and benefits of a full evaluation study. • If the need for evaluation data is significant, plan a more sophisticated study and budget appropriately. • In addition to questionnaires, such a study is likely to include pre- and post-tests of performance and/or interviews.
Step 2 **Select Tool**	• Achievement measure: • The most popular and accurate measures of learning outcomes are achievement measures. • Self-report measure: • When it is not possible to use the achievement measure, a self-report measure can be used. • Self-report measures do not have the objectivity of achievements measures but do provide valuable information about class progress.
Step 3 **Adapt Tool**	• Modify content: • Reword, delete, or add items to ensure goals are achieved and to measure accomplishment of learning objectives. • Modify the format: • Change rating scale items to open-ended questions or vice versa, depending on the kind of information desired.
Step 4 **Implement**	• Announce and explain the use of this evaluation and its purpose at the beginning of the course. • Allow enough time to complete the evaluation. • Participants turn in their evaluations when finished. • Those who are finished may leave the room so as not to distract others.
Step 5	• To analyze this self-assessment measure, subtract the "before" rating

Table 4. Six Steps for Developing a Level 2 Evaluation

Step	Procedure
Analyze	from the "after" rating. • Total the results and divide by the number of items to find the average perceived gain for each individual. • The higher the number, the greater the impact of the course. • Use this process to find the average for the whole class or the aggregate of classes.
Step 6 Report	• Write a short "Class Evaluation Report" summarizing the results (*see* page 39). • Always compute mathematical averages of students responses, even if it is necessary to convert them to numbers. • If someone gives a rating outside the limits (*e.g.*, 10 on a scale of 1 to 5), convert it to the scale (in this case to 5, since that is the highest score). • Always include all student comments, either by attaching copies of the forms themselves or, if time permits, by retyping them. • Send copies of the report to the trainer, the trainer's supervisor, and to the organization that funded the training.

Sample Level 2 Evaluation

Table 5 contains an example of a Level 2 Evaluation form, based on a document created

by the U. S. Department of Transportation (2005).

Table 5. Example Level 2 Evaluation Form

Your Name (*Optional*)	Course Name
Trainer's Name	Course Date

Your level of learning is important to us. Please respond to the questions and statements in the spaces below. Your individual score will remain anonymous, but the collective scores will be used to plan future training.

Course Objectives	Before Taking this Class	After Taking this Class
As a result of taking this course, I am able to:	Before taking (Course Name), my level of knowledge or competency for this objective was:	After taking (Course Name), my level of knowledge or competency for this objective is:

Course Objectives	Before Taking this Class	After Taking this Class
1. *State the course objective here.* *Example:* Demonstrate how to use the mouse to select an icon in Windows.	1 2 3 4 5 Low - Moderate - High	1 2 3 4 5 Low - Moderate - High
2. *State the course objective here.*	1 2 3 4 5 Low - Moderate - High	1 2 3 4 5 Low - Moderate - High
3. *State the course objective here.*	1 2 3 4 5 Low - Moderate - High	1 2 3 4 5 Low - Moderate - High
4. *State the course objective here.*	1 2 3 4 5 Low - Moderate - High	1 2 3 4 5 Low - Moderate - High
5. *State the course objective here.*	1 2 3 4 5 Low - Moderate - High	1 2 3 4 5 Low - Moderate - High

Level 3 (Transfer-to-the-Job) Evaluation

Background

A Level 3 Evaluation normally occurs in the workplace following training. The most-common form of Level 3 Evaluation is an Employee Performance Review. However, in an environment such as NCBT's, where students are in vocational training that includes formal externships, quasi-Level 3 Evaluations can be performed while training is taking place (Marshall, 2005).

The following guidelines were adapted by the U. S. Department of Transportation (2005) from Donald Kirkpatrick's, *Evaluating Training Programs: the Four Levels*:

- Use a control group, if practical.

- Allow time for behavior change to take place.

- Evaluate both before and after the program if practical.

- Survey and/or interview one or more of the following: trainees, their

    immediate supervisor, their subordinates, and others who observe their

    behavior.

- Get 100 percent response or a sampling.

- Repeat the evaluation at appropriate times.

- Consider costs versus benefits.

In a training analysis prepared for the U. S. Department of Transportation, Arthur

Andersen Co. recommended the following protocols for Level 3 Evaluations that are

universally applicable (DOT, 2005):

- 30% of courses should be evaluated at this level.

- Identify level of evaluation for each program.

- Establish schedule for routine follow-up of participants.

- Increase ownership of participants to follow-up survey efforts.

Creating a Level 3 Evaluation

Table 6 contains six steps for developing a Level 3 Evaluation adapted by the U. S.

Department of Transportation (2005) from *Diversity Training Evaluation Toolkit*, published in

1994 by the U. S. Federal Aviation Administration.

Table 6. Six Steps for Developing a Level 3 Evaluation	
Step	*Procedure*
Step 1 Plan	• Decide whether you will evaluate to confirm the effectiveness of the course or to improve it. • List specific values and skills that you will measure. • Talk to managers and other stakeholders about the proposed level of evaluation and the costs and benefits of a full evaluation study. • If your need for evaluation data is limited or your budget is not sufficient for a full study, use the sample forms in this guide.

Table 6. Six Steps for Developing a Level 3 Evaluation

Step	*Procedure*

	• This will permit collection of limited data on Level 3 for a very low cost.
	• Determine which employees will be included.
	   • Ideally, everyone who has taken the course should be included.
	   • If time and budget are a constraint, consider using a sample group.

Step 2
Select Tool
	• Post-training surveys:
	   • Questionnaires that are sent to employees and/or their managers one to three months after training.
	   • Their purpose is to determine which skills learned in the course are being used.
	• Direct observation:
	   • Consists of checklists to record actual observations of employee application of skills learned in training.

Step 3
Adapt Tool
	• Modify content:
	   • Reword, delete, or add items to make them fit your learning objectives.
	• Modify the format:
	   • Change rating scale items to open-ended questions, or *vice versa*, depending on what kind of information you wish to obtain.

Step 4 **Implement**	• Two dates should be chosen: one soon after the training, and one later. • Send the evaluation and other relevant information to the employee. • Send reminders to those who are slow to return the forms and keep records of the collected evaluations to ensure good return. • Consider measuring both student responses and the responses of their managers or supervisors.
Step 5 **Analyze**	• For each item, count the number of responses to each rating and find the average. • Review the comments, categorize them based on similarities, and write a brief summary. • Total the results and divide by the number of items to find the average perceived gain for each individual. • Use this process to find the average for the whole class or the aggregate of classes.
Step 6 **Report**	• Write a short "Class Evaluation Report" summarizing the results (*see* page 40). • Always compute mathematical averages of students responses, even if it is necessary to convert them to numbers. • If someone gives a rating outside the limits (*e.g.*, 10 on a scale of 1 to 5), convert it to the scale (in this case to 5, since that is the highest

Table 6. Six Steps for Developing a Level 3 Evaluation

Step	*Procedure*
	score). • Always include all student comments, either by attaching copies of the forms themselves or, if time permits, by retyping them. • Send copies of the report to the trainer, the trainer's supervisor, and to the organization that funded the training.

Sample Level 3 Evaluation

Table 7 contains an example of a Level 3 Evaluation form, based on a document created by the U. S. Department of Transportation (2005).

Table 7. Example Level 3 Evaluation Form

Your Name (*Optional*)	Course Name
Trainer's Name	Course Date

Your level of learning is important to us. Please respond to the questions and statements in the spaces below. Your individual score will remain anonymous, but the collective scores will be used to plan future training.

Specific Task	Prepared	Use	Importance
Example: Use MS Word to create tables in reports.	How well did the course prepare you to perform this task?	How often do you use this knowledge or skill on the job?	How important is this skill or knowledge to your job?
State specific task here.	o Poorly o Somewhat o Very well	o Seldom o Sometimes o Very often	o Not at all o Somewhat o Very much
State specific task here.	o Poorly o Somewhat o Very well	o Seldom o Sometimes o Very often	o Not at all o Somewhat o Very much
State specific task here.	o Poorly o Somewhat o Very well	o Seldom o Sometimes o Very often	o Not at all o Somewhat o Very much
State specific task here.	o Poorly o Somewhat o Very well	o Seldom o Sometimes o Very often	o Not at all o Somewhat o Very much
State specific task here.	o Poorly o Somewhat o Very well	o Seldom o Sometimes o Very often	o Not at all o Somewhat o Very much

Level 4 (Organizational Impact) Evaluation

Background

Perhaps the least-used evaluation method because of the inherent burden of data collection, analysis, and reporting, Level 4 Evaluations are critically important when training for purposes other than new-hire, emergency-response/damage-control, or regulatory compliance must be justified. The following guidelines were adapted by the U. S. Department of Transportation (2005) from Donald Kirkpatrick's, *Evaluating Training Programs: the Four Levels*:

- Use a control group, if practical.

- Allow time for results to be achieved.

- Measure both before and after the program if practical.

- Repeat the measurement at appropriate times.

- Consider costs *versus* benefits.

- Be satisfied with evidence if proof is not possible.

In a training analysis prepared for the U. S. Department of Transportation, Arthur Andersen Co. recommended the following protocols for Level 4 Evaluations that are universally applicable (DOT, 2005):

- 10% of courses should be evaluated at this level.

- Limit the goal of conducting a Level 4 evaluation.

- Adopt a Return on Expectations (R-O-E) philosophy in lieu of Return on Investment (R-O-I) philosophy.

- Determine secondary sources of data.

Creating a Level 4 Evaluation

Table 8 contains six steps for developing a Level 4 Evaluation adapted by the U. S. Department of Transportation (2005) from *Diversity Training Evaluation Toolkit*, published in 1994 by the U. S. Federal Aviation Administration.

| **Table 8. Six Steps for Developing a Level 4 Evaluation** ||
Step	*Procedure*
Step 1 **Plan**	• Level 4 evaluation is used to determine whether any organizational benefits have resulted from the training. • The goal is expressed in terms or organizational results which are measured and can be related to training. • Results can be in the form of perceptions, performance improvements, or financial benefit. • Talk to managers and other stakeholders about the proposed level of evaluation and the costs and benefits of a full evaluation study. • Determine which employees will be included. • Include everyone who has taken the course. • If time and budget are a constraint, consider using a sample group.
Step 2 **Select Tool**	• Post-training surveys: • Questionnaires that are sent to employees and/or their managers after training. • Their purpose is to determine which skills learned in the course are being used. • Direct observation: • Consists of checklists to record actual observations of employee application of skills learned in training.
Step 3 **Adapt Tool**	• Modify content: • Reword, delete, or add items to make them fit your learning objectives. • Modify the format: • Change rating scale items to open-ended questions, or vice versa, depending on what kind of information you wish to obtain.
Step 4 **Implement**	• Organizational results are usually observed 3 to 6 months after the training. • Send the evaluation to managers who can observe organizational results. • Send reminders to those who are slow to return forms and keep records of the collected evaluations to ensure good return.

Table 8. Six Steps for Developing a Level 4 Evaluation

Step	*Procedure*
Step 5 **Analyze**	• For each item, count the number of responses to each rating and find the average. • Review the comments, categorize them based on similarities, and write a brief summary. • Total the results and divide by the number of items to find the average perceived gain for each individual. • Use this process to find the average for the whole class or the aggregate of classes.
Step 6 **Report**	• Write a short "Class Evaluation Report" summarizing the results (*see* page 40). • Always compute mathematical averages of students responses, even if it is necessary to convert them to numbers. • If someone gives a rating outside the limits (*e.g.*, 10 on a scale of 1 to 5), convert it to the scale (in this case to 5, since that is the highest score). • Always include all student comments, either by attaching copies of the forms themselves or, if time permits, by retyping them. • Send copies of the report to the trainer, the trainer's supervisor, and to the organization that funded the training.

Sample Level 4 Evaluation

Level 4 Evaluation forms should be completed by all stakeholders, so that input and observations from every interested party can be collected, analyzed, and reported.

Table 9 contains an example of a very simple Level 4 Evaluation form, based on a document created by the U. S. Department of Transportation (2005).

Note that this form can also be modified slightly and used by supervisors to report on the results of employee observation. Primary modifications would be change "Your Name" and "Your Supervisor" to "Trainee's Name" and Supervisor's Name," respectively and change the instructional paragraph to read "Trainees' levels of learning are..."

Table 9. Example Level 4 Evaluation Form

Your Name (*Optional*)	Your Supervisor (*Optional*)
Course Name	Course Location
Trainer's Name	Course Date

Your observation of learning <u>application</u> is important to us. Please respond to the questions and statements in the spaces below. Use 1 for lowest, and 5 for highest. Your individual scores will remain anonymous, but the collective scores will be analyzed to evaluate this course and plan future training.

Organizational Result	**Rating**				
	Strongly Agree	Agree		Strongly Disagree	
1. State desired organizational result here. *Example*: Using computers is part of my everyday job.	1	2	3	4	5
2. State desired organizational result here.	1	2	3	4	5
3. State desired organizational result here.	1	2	3	4	5
4. State desired organizational result here.	1	2	3	4	5
5. State desired organizational result here.	1	2	3	4	5

Jack Phillips and "Level 5" Evaluations

> *"Pay attention to what works and do more of it and try to understand the principles involved. And also: pay attention to what doesn't work and stop doing it." ~ Nathaniel Brandon*

Building on the work of Donald Kirkpatrick, Jack Phillips began publishing his theories on a fifth level of evaluation in the 1980's. Phillips' work extended Kirkpatrick's Level 4 – Organizational Impact – and made it business-friendly by adding a detailed "Return on Investment" feature. Many modern authors mistakenly merge Kirkpatrick's Level 4 and Phillips' ROI tenets.

Just as Kirkpatrick's Level 4 Evaluations are not done at the campus level within NCBT's corporate structure, Phillips' ROI evaluations are generally not performed at the campus level. Therefore, NCBT-Knoxville instructors get no formal experience in preparing a "Level 5" Evaluation.

Some instructors may wish to undertake an informal "Level 5" evaluation, however, for personal knowledge and planning. Phillips, Phillips & Hodges (2004) discuss data collection and reporting in detail in their book, *Make Training Evaluation Work*. In small studies, such as those NCBT instructors might want to undertake, the authors recommend using readily available software, such as Microsoft Excel® for data compilation and statistical analysis.

Table 10 contains a sample basic ROI analysis of Web-based training (WBT) compared to traditional classroom instruction, prepared by Horton & Horton (2000) using Microsoft Excel®. The complete analysis contains seven spreadsheets entitled "Costs," "5-Year Costs," "Lost Opportunity," "Soft Costs," "Time Value," "Time Value 2," and "Shared Costs."

Horton & Horton (2000) introduce their spreadsheet package with a humorous

observation: "Estimating costs of training projects is a black art only slightly more precise

than reading tea leaves or bird entrails." These are familiar words to any Training Evaluator

who has tackled the process of analyzing costs and/or ROI of a training event.

Table 10. Example ROI Analysis

Money unit: US Dollars (USD)

	Classroom	WBT	
Per-course costs	Classroom	WBT	
Course length	8	8	hours
x Development time rate	50	200	hours devt/course hr
x Development cost rate	50	100	USD/hour devt
= Total per-course costs	20000	160000	USD
Per-class costs	Classroom	WBT	
Instructor/facilitator salary	800	800	USD
+ Instructor/facilitator travel	1500	0	USD
+ Facilities	500	50	USD
= Subtotal (per class)	2800	850	USD
Number of learners	200	200	learners
÷ Class size	20	20	learners
= Number of classes	10	10	classes
Total class-offering costs	28000	8500	USD
Per-learner costs	Classroom	WBT	
Learner's travel	1500	0	USD
+ Learner's salary	800	800	USD
+ Instructor/facilitator's salary	25	50	USD
= Subtotal (per learner)	2325	850	USD
x Number of learners	200	200	learners
= Total learner costs	465000	170000	USD
Total costs	Classroom	WBT	
Course costs	20000	160000	USD
+ Class costs	28000	8500	USD
+ Learner costs	465000	170000	USD
= Total project costs	513000	338500	USD
Cost savings for WBT	174500	USD	
Return on investment	1.246428571		

Evaluation Reports

> *"When you produce results you gain credibility. When you have credibility, you will have an easier time producing results." ~ Brian Koslow*

Evaluation reports can range from simple averages to extensive statistical analyses with supporting documentation, charts, graphs, and observations. Reports may contain the results of one training session, or they may track long-term training efforts for one course or multiple tracks.

Table 11 contains a simple Evaluation Report, published by the U. S. Department of Transportation (2005). Note that Evaluator's comments ("Special Issues") are included as a method for explaining any potential skewed analysis results.

Phillips, Phillips & Hodges (2004) developed a simplified form for tracking multiple courses as shown in Figure 13. Such a form allows Evaluators to maintain records in an organized manner, either manually or electronically, so the records are quickly available for reference.

Table 11. Example Simple Evaluation Report

Evaluator's Name	Course Name
Trainer's Name	Course Location
No. of Students	Course Date

Following is a compilation of scores and comments submitted by students in the training course described above.

Evaluation Averages				*Rating Scale*	
Relevance to your job	4.8	Course materials	4.6	1	Poor
Clarity of objectives	4.7	Instructor(s)	4.9	2	Satisfactory
Course length and pace	4.6	Facilities	4.2	3	Average
Overall quality	**4.8**			4	Very Good
				5	Excellent

Student Comments

"I would have liked more time on the case studies."

"This class has done a great deal to increase my overall knowledge and confidence as a supervisor."

"Instructors were very helpful given the wide disparity of levels of knowledge!"

"Terrific; this information will help me in my new job."

Special Issues

One student left class at lunch on the second day due to illness and did not complete the final test or fill out an evaluation form.

Figure 13. Learning & Development Scorecard (Phillips, Phillips & Hodges, 2004)

References

American Society for Training and Development (ASTD). (1975). *Evaluating training*

    programs. Madison, WI: Author.

Arthur Andersen & Co. (1996, February). *US Department of Transportation, training program*

    review, final report. Publication data unknown; cited in US DOT, below.

Clark, D. (2004). *Instructional system development - evaluation phase - chapter VI.* Retrieved

    October 24, 2005, from www.nwlink.com/~donclark/hrd/sat6.html

Clark, D. (2000). *Don Kirkpatrick and evaluating training*. Retrieved October 24, 2005, from

    www.nwlink.com/~donclark/hrd/history/kirkpatrick.html

Dabbagh, N. (2002). *Instructional design knowledge nase: Conduct evaluation of instructional*

    program. Retrieved October 24, 2005, from

    http://classweb.gmu.edu/ndabbagh/Resources/IDKB/evaluation.htm

Horton, W. & Horton, K. (2000, September). *WBT: Return on investment or dead on arrival*?

    ASTD TechKnowledge. Boulder, CO: William Horton Consulting, Inc.

Keller, J. M. & Young, A. (1995). *How to do training evaluation: A practical handbook: Produced*

    for the Federal Aviation Association, Office of Human Resource Development, Washington D. C.

    Washington: U. S. Government Printing Office.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco: Berrett-

    Koehler Publishers.

Marshall, V. (2005, October-November). Lectures and ancillary handouts, Education 524

    Class, Tusculum College, Knoxville, TN.

National Institute for Occupational Safety & Health. (n. d.). *NIOSH Mining: Training*

      Evaluation Tips. Retrieved October 24, 2005, from www.cdc.gov/Niosh/mining/topics/training/traini

Nickols, F. (2000). Evaluating training: There is no cookbook approach. Retrieved

      September 12, 2005, from http://home.att.net/~nickols/evaluate.htm

Phillips, J., Phillips, P. & Hodges, T. (2004). *Resources needed to make evaluation work*.

      Retrieved October 24, 2005, from www.learningcircuits.org/2004/nov2004/phillips.htm

Scriven, M., Stufflebeam, D., *et al.* (2005). *Evaluation checklist project.* Retrieved October 3,

      2005, from www.wmich.edu/evalctr/checklists/

United States Department of Transportation. (n. d.). *Learning and development framework:*

      Training evaluation guide: Table of contents. Retrieved October 24, 2005, from

      dothr.ost.dot.gov/HR_Programs/Learning___Development/LDFramework/evaltoc.htm

United States Office of Personnel Management, Human Resources Development Group,

      Office of Research and Information. (1994, March). *Establishing the value of training:*

      Resource guide. HRDG-023. Washington, D. C.: U. S. Government Printing Office.

Appendices

Because the scope of this Job Aid is limited, several excellent articles, checklists, and whitepapers could not be included. Therefore, they have been inserted in the following pages as Appendices.

Training evaluation is, by nature, ever-changing. Although Kirkpatrick's model has existed almost-unchanged for more than 45 years, new theories and models are published and/or discussed regularly. Therefore, Evaluators should constantly review and update their skills, procedures, forms, and tools. As new resources are discovered, place them in this "Appendix" and catalogue them in the table below.

ID	Title/URL	Author	Date	Pgs.
A	"The State of Learning Analytics" *Training & Development* (58:6)	Berk, J.	June, 2004	6
B	"Training Evaluations" *Training & Development* (58:9)	Berk, J.	Sep., 2004	6
C	*Checklist for Formatting Checklists* www.wmich.edu/evalctr/checklists/cfc.pdf	Bichelmeyer, B.	Oct., 2003	3
D	*Kirkpatrick's Training Evaluation Model* www.businessballs.com/kirkpatricklearningevaluation model.htm	Chapman, A.	2005	11
E	*Performance Appraisals* www.businessballs.com/performanceappraisals.htm	Chapman, A.	2005	22
F	*Making Evaluation Meaningful to all Education Stakeholders* www.wmich.edu/evalctr/checklists/makingevalmeaning ful.pdf	Gangopadhyay, P.	Sep., 2002	3
G	"Digital Beat: A Love-Hate Thing" *Training & Development* (58:6)	Hartley, D.	June, 2004	3
H	"Transferring Learning to Behavior" *Training & Development* (59:4)	Kirkpatrick, J.	Apr., 2005	2

ID	Title/URL	Author	Date	Pgs.
I	"Are You an Active Listener" *Facts & Findings*	National Association of Legal Assistants	Nov., 1993	2
J	"Trainer, Assess Thyself: Do It for You" *Training & Development* (56:7)	Newman, A.	July, 2002	3
K	*Qualitative Evaluation Checklist* www.wmich.edu/evalctr/checklists/qec.pdf	Patton, M.	Sep., 2003	13
L	"Concerning the Fallibility of Judgment from the Side, the Rear, and On-High: A Dialogue About Scriven's Critique" *Journal of Personal Evaluation in Education* (16:4)	Roth, W. & Tobin, K.	Dec., 2002	7
M	"Out of the Frying Pan, into the Fire: Comments on Roth/Tobin" *Journal of Personal Evaluation in Education* (16:4)	Scriven, M.	Dec., 2002	4
N	*Key Evaluation Checklist* www.wmich.edu/evalctr/checklists/kec_april05.pdf	Scriven, M.	Oct., 2004	7
O	*Evaluating On-Line Learning* www.fastrak-consulting.co.uk/tactix/Features/evaluate/evaluate.htm	Shepherd, C.	1999	8
P	"President's Page: The Bottom Line on Training" *Training & Development* (57:2)	Sung, T.	Feb., 2003	1