

Using Kirkpatrick's Four Levels of Evaluation in Real-World Organizations

Principles,
Tips, &
Tricks

April 6, 2008--ISPI NYC

Richard B. Pearlstein, PhD

1

Preamble: Adding value with evaluation

- Management often doesn't want real evaluation—it wants endorsement
 - Management likes “manipulable” data*
 - Evaluation data are not easy manipulated
- Contract with management before starting
 - Point out that evaluation can improve programs or be used to support eliminating ineffective ones
 - Remind management that if it doesn't like the data, it doesn't have to report them

*Mintzberg, H. (1973)

April 6, 2008--ISPI NYC

Richard B. Pearlstein, PhD

2

What is evaluation in HPT context?

- **Actionable** information
- **Two kinds:**
 1. Information to help us improve HPT interventions
 - Effectiveness of design & supporting tools and media
 - Effectiveness of delivery & results
 2. Information to help us get support for interventions
 - Impact of interventions on work outcomes
 - Return on investment (ROI) of intervention

The wrong way to evaluate

- Many people think evaluation has 2 phases:
 1. Ask a whole bunch of questions
 2. Figure out what to do with the answers



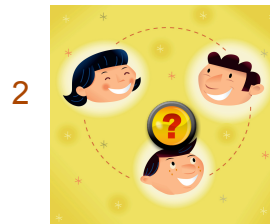
The right way to evaluate

- The actual process is pretty much the reverse:

1. Figure out what questions you need to answer



2. Figure out what information you need to answer those questions accurately



Example of the right way to evaluate

- Start with the questions you need to answer:
 - Example: Is **Intervention X** worth the investment?
- Then ask, "What do we need to know before we can answer?"
 - Example: We need to know how much **X** costs (-\$), and how much it saves/gains us (+\$)
- Next ask, "How can we find that info?"
 - Example: We can find intervention costs directly, but we will need a way to measure savings/gains

Example (continued)

- Next ask, “Okay, how do we measure that?”
 - Example: We will need to ask supervisors for data on savings and gains
- Then ask, “What data do they have available?”
 - Example: # of tasks completed/year, \$/task
- Then ask, “How do we collect that data?”
 - Example: Use survey, examine production records
- And so on. When you are finished, you will have an evaluation plan

April 6, 2008--ISPI NYC

Richard B. Pearlstein, PhD

7

Kirkpatrick's 4 levels in HPT context

- Level 1—Participant reactions
 - What worked, what didn't from participants' perspectives?
- Level 2—Participant changes at end of intervention
 - Can participants perform differently in measurable ways?
- Level 3—Is change maintained on the job?
 - Do participants reliably perform in new ways on the job?
- Level 4—Impact on organizational outcomes
 - What has improved as a result of intervention?
 - Was it worth the investment?

Jack Phillips
calls ROI
“Level 5”

April 6, 2008--ISPI NYC

Richard B. Pearlstein, PhD

8

Two views of level 5

- Jack Philips' level 5 can be seen as a “deep dive” into Kirkpatrick's level 4
 - Some say that this level 5 has always been a part of Kirkpatrick's level 4
- Roger Kaufman's level 5 is a new level
 - It measures impact on outcomes in the greater world in which the organization is embedded
 - Important for societal & global well-being

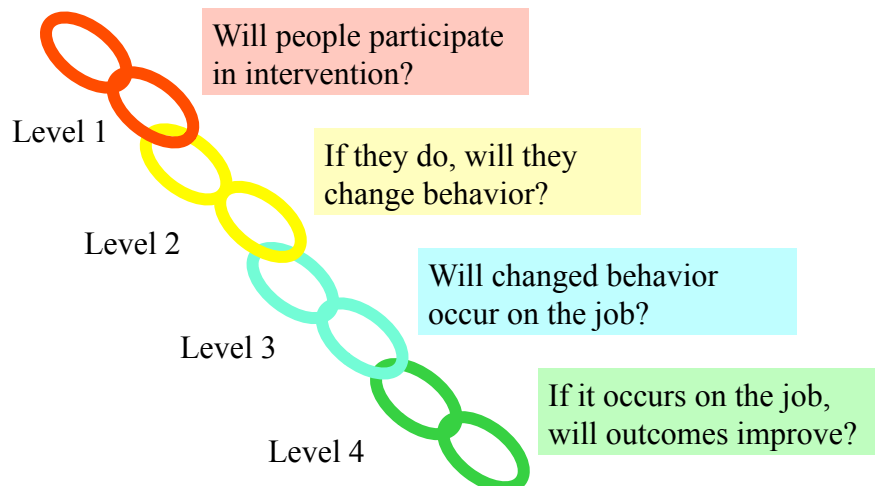
A note on “Level 0”

- Level 0 is my contribution to the taxonomy
 - “Hallway evaluation”— “How'd you like the training?”
 - The most common type of evaluation
- Don't ignore level 0
 - Although it is purely anecdotal, it is persuasive
- Exploit level 0
 - Include testimony from widely-respected figures
 - Also include at least level 2 data

Using data from Kirkpatrick's 4 levels

- The information you collect at levels 1 & 2 helps you improve design & delivery
 - 1) Participant reactions can surface problems
 - 2) Measurement shows whether intervention worked initially
- The information you collect at levels 3 & 4 helps you get organizational support
 - 3) On-the-job application results link to management attention & reinforcement
 - 4) Outcome data make case for intervention's value

Levels as links in a chain



Examples of primary & follow-up interventions

Primary “Acquisition”*	Follow-up “Maintenance”*
Participants attend training	Supervisors encourage use of new skills
Participants receive job aids	Management sends repeated reminders that it expects job aids to be used consistently
Participants come under a new reward system	Management shows that it uses new reward system fairly

*For the seminal article acquisition & maintenance, see Brethower, K. S. (1967).

Level 1 tip

- Ask questions that tap participants’ **direct experience**
 - Level 1 asks for reactions—useful reactions are:
 - Based on direct experience, NOT opinion,
 - Descriptive, not judgmental

Ask this	NOT this
“Did the instructor answer questions to your satisfaction?” “Were you able to hear the instructor easily?”	“How well did the instructor teach?”

Additional level 1 tips

- **Be careful what you ask for**
 - Limit issues and extent of comments
 - A useful guideline = about 4 – 7 questions
- **To collect more meaningful comments, consider holding a “staff meeting” at end of day’s training**
 - Use representatives from tables
 - Discuss what worked, what didn’t
 - Agree on changes for next day (or future classes)

Level 1 trick

- **Ask how comfortable participants would be using specific skills back on the job**
 - Their answers are more likely to be meaningful
- **Don’t ask how well they learned something**
 - Research shows participants’ self-ratings of their skills don’t correlate with actual proficiency

Ask this	NOT this
“How comfortable would you be on the job translating an article from French?”	“How skilled are you now at translating French to English?”

Level 1 trick

- Use this template for Likert* scale items:

The instructor moved activities at a comfortable pace:	1 Strongly disagree	2 Disagree	3 Neither agree nor disagree	4 Agree	5 Strongly Agree
(Check ONE box only)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Comments:					

*Pronounced "Lick-urt," he said professorially

Level 2 tip

- Measure performance at end of intervention
Performance is:

- A skill used under specified conditions to meet specified standards, or
- Knowledge applied under specified conditions to meet specified standards

Example Performance Test Question	Non-Performance Question
Review this article & list at least 3 areas requiring analysis for impact on U.S.	How have you deepened your appreciation of analysis?

Another level 2 tip

- **Measure pre-intervention performance first**
 - Pre- and post-measures give basis for showing that improvement resulted from intervention
 - Sometimes logistics prevent taking pre-measures
- **If pre-measures are not possible:**
 - Screen candidates for previous experience
 - Report that post measures show only that participants could perform at end of intervention
 - They may have been able to before intervention

Level 2 trick

- **Base end-of-intervention performance tests on intervention objectives**
 - Ask: “*What does Objective X state that employees will do differently on the job?*”
 - Then ask: “What can I measure at the end of the the intervention to see if employees can do that Objective X thing to standard on the job?”

Example: Given the following information, prepare a project breakdown structure meeting X standards.

Level 3 tips

- Use existing measures of on the job performance.
- Use simple quasi-experimental designs
 - Collect pre- and post-measures of those who have completed intervention
 - Compare with equivalent time series measures of those who have not yet begun intervention

Another level 3 tip

- Ask supervisors if there is a difference between what people who have already had Intervention **X** do on the job compared to people who have NOT yet had it.

Example: “Bob, two of your managers have been using job aid **X** during performance appraisals. Have you found any differences in how they document innovative solutions? If so, what are they?”

Level 3 trick

- If you can get supervisors* to provide information on about 30 participants in an intervention and 30 non-participants, you can test differences between the 2 groups
 - You can answer: “Are the differences due to chance or are they due to the intervention?”
 - Excel comes with a data analysis “toolpak” that allows for analyses like this.

*Use at least 5 supervisors to get the necessary number of participants

Level 4 tip

- Level 4 is the bottom line. Senior management wants to know if an intervention’s results benefited the organization.

- And, were the benefits worth the investment?

Example: You’ve found that participants in the **X** mentoring program are better at giving useful feedback to their staffs. Now ask “So what?”

- Are their staffs doing better on quality & quantity measures; if so, are the improvements worth the cost of the mentoring program?

Other level 4 tips

- Measure senior management's desired outcomes (don't guess about them)
 - Use their existing measures whenever possible
 - Suggest that they consider micro, macro, and mega outcomes
- Use managements' standard approaches to costs for ROI
- Again, use quasi-experimental design—similar principle to test marketing new products (stagger roll-out, and treat each as a separate test market)

Level 4 trick

- You are much more likely to be able to evaluate at Level 4 if you get upper management support before you start.
 - Make a strong business case for doing Level 4:
 - E.g., "We're spending a lot of money on leadership development. Let's find out if it's worth it. If it's not, we can either improve the program or eliminate it."
 - Tell upper management what you'll need:
 - E.g., "We'll need technical support on surveys; we'll need you to send an e-mail to senior management, etc."

Summary: The value of evaluation

Evaluation can prove the worth of your HPT interventions and show you how to make them even better.



Bibliography

- Brethower, K. S. (1967). "Maintenance Systems: The Neglected Half of Behavior Change," in *Managing the Instructional Programming Effort* (Rummler, Yaney, & Schrader, Eds.). University of Michigan.
- Farrington, J., & Clark, R. E. (2000, November/December). "Snake oil, science, and performance products." *Performance improvement*, 39 (10).
- Hale, J. (2002). *Performance-Based Evaluation: Tools and Techniques to Measure the Impact of Training*. San Francisco: Pfeiffer.
- Kirkpatrick, D. (1994). *Evaluating Training Programs: The Four Levels*. San Francisco: Berrett-Kohler
- Mintzberg, H. (1973). *The Nature of Managerial Work*. New York: Harper & Row.
- Pearlstein, R. B., Brethower, D., Lindsley, O., & Deterline, W. (1994). First annual NSPI research symposium: Commonly overlooked research findings that can make training more successful. San Francisco: Symposium held in conjunction with the National Society for Performance and Instruction Annual Conference, April.
- Phillips, J. (1997). *Return on Investment*. Houston: Gulf Publishing.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.
- Swezey, R. W., & Pearlstein, R. B. (2001). Selection, training, and development of personnel. In Salvendy, G. (Ed.). *Handbook of industrial engineering, third edition*. New York: John Wiley. Pp. 920-947.
- Wallace, G. (1999) *lean-ISD*. Naperville IL: CADDI Press.

Bio Sketch

Rich has more than 35 years' experience in researching, using, and evaluating ISD and other performance improvement processes. He helped develop the Air Force's original ISD manual, and co-authored the Army's Criterion-Reference Test Construction Manual. He managed training development at Federal agencies and private corporations. A long-term ISPI member and a past board member, he has presented at 22 ISPI Annual Conferences. He currently manages evaluation of a large certification program for a Federal agency.

Richard B. Pearlstein, PhD, Director of Organizational Effectiveness

The Center for Systems Management

1951 Kidwell Drive, Suite 750

Vienna, VA 22182

703-852-3336 (Direct Line)

703-852-3301 (Fax)

rpearlstein@csm.com

www.csm.com