

# Developing Valid Level 3 Evaluations

**Presented by:**  
Ken Phillips  
Phillips Associates  
November 15, 2010



# AGENDA

- ▶ Review Kirkpatrick/Phillips five level evaluation model
- ▶ Examine Level 3 evaluation facts
- ▶ Examine overall guidelines for conducting Level 3 evaluations
- ▶ Analyze survey creation errors in a sample Level 3 evaluation
- ▶ Apply 11 tips for creating valid Level 3 evaluations



# KIRKPATRICK/PHILLIPS EVALUATION MODEL

Levels of Evaluation	Measurement Focus	Time Frame
Level 1: Reaction	Participant favorable reaction to a learning program	Conclusion of learning program
Level 2: Knowledge	Degree to which participants acquired new knowledge, skills or attitudes	Conclusion of learning program or within 6 to 8 weeks after
Level 3: Behavior	Degree to which participants apply back-on-the-job what was learned in program	2 to 12 months
Level 4: Results	Degree which targeted business outcomes were achieved as a result of	9 to 18 months
Level 5: ROI	Degree to which monetary program benefits exceed program costs	9 to 18 months



# “CHAIN OF EVIDENCE” DETERMINANTS

- ▶ **Strategic importance of learning program**
- ▶ **Cost of learning program**
- ▶ **Number of participants attending program**
- ▶ **Stakeholder/business executive expectations**



# LEVEL 3 EVALUATION FACTS\*

- ▶ **55% of organizations evaluate at least some learning programs at Level 3**
- ▶ **75% of organizations view data collected as having high or very high value**

**\*ASTD Research Study, “The Value of Evaluation: Making Training Evaluations More Effective,” 2009**



# DATA COLLECTION METHODS

Methods	EVALUATION LEVELS			
	1 Reaction	2 Learning	3 Behavior	4 Results
Survey/Questionnaires	●	●	●	●
Interviews	●	●	●	●
Focus Groups	●	●	●	●
Knowledge Testing		●		
Work Review		●	●	
Structured Observation		●	●	
Action Planning		●	●	●
HR/Business Unit Reports				●
Key Performance Indicators				●

(Kirkpatrick, 2006)



# LEVEL 3 EVALUATION FACTS\*

- ▶ **Most common methods for evaluating behavior**
  - **Participant surveys (31%)**
  - **Action planning (27%)**
  - **Performance records monitoring (24%)**
  - **On job observation (24%)**

\*ASTD Research Study 2009



# OVERALL GUIDELINES\*

- ▶ **Use pre- post-test evaluation design**
- ▶ **Where possible, use control group**
- ▶ **Allow time for behavior change**
  - **How soon can participants apply what was learned?**
  - **How long before new behavior is recognized?**
- ▶ **Include some measures that can be independently verified**

\* Adapted from Donald & James Kirkpatrick, 2006.



# MEASUREMENT SOURCES

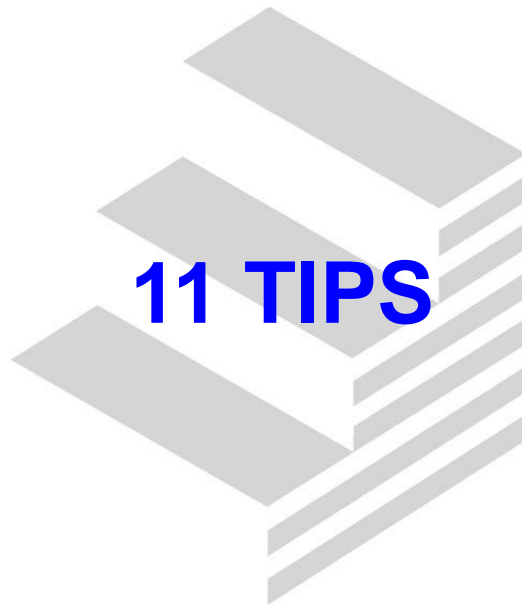
- ▶ **Learner**
- ▶ **Peers/colleagues**
- ▶ **Direct reports**
- ▶ **Manager**
- ▶ **Trainer**
- ▶ **Customers**





**SAMPLE  
LEVEL 3  
EVALUATION**





# 11 TIPS



# EXAMPLES

8. Before providing employees with feedback about their job performance, my manager considers whether or not he or she is knowledgeable about their job.
  
25. When giving feedback to an employee my manager considers whether it should be done privately or in the presence of others.



# TIP 1 (CONTENT)

- ▶ Focus on observable behavior not thoughts or motives\*



\* Palmer Morrel-Samuels, “Getting the Truth into Workplace Surveys”, *Harvard Business Review*, 2002.



# EXAMPLES

- 14. My manager gives his or her employees feedback just as soon as possible after an event has happened and avoids getting emotional or evaluative.**
- 18. My manager provides employees with regular ongoing feedback about their job performance and speaks in a normal conversational tone or manner when delivering the feedback.**



# TIP 2 (CONTENT)

- ▶ **Limit each item to a single description of behavior\***

\* Palmer Morrel-Samuels, 2002.



## TIP 3 (CONTENT)

- ▶ **Word about 1/3 of the survey items so that the desired answer is negative\***



# EXAMPLES

2. My manager doesn't get to know his or her employees as individuals before providing them with feedback about their job performance.
7. When giving employees feedback about their job performance, my manager doesn't distinguish between patterns of behavior and random one-time events.



# EXAMPLES

- ▶ **Building Trust**
- ▶ **Credibility**
- ▶ **Feedback Sign**
- ▶ **Feedback Timing**
- ▶ **Feedback Frequency**
- ▶ **Message Characteristics**



# TIP 4 (FORMAT)

▶ **Keep sections of the survey**

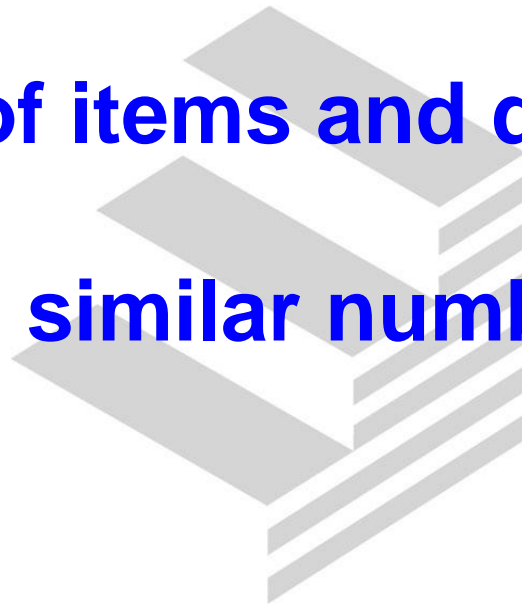
**unlabeled\***

\* Palmer Morrel-Samuels, 2002.



## TIP 5 (FORMAT)

- ▶ **Design sections to contain a similar number of items and questions to contain a similar number of words\***



\* Palmer Morrel-Samuels, 2002.

# TIP 6 (MEASUREMENT)

- ▶ **Collect data from multiple observers or a single observer multiple times**



# TIP 7 (MEASUREMENT)

- ▶ Create a response scale with numbers at regularly spaced intervals and words only at each end\*



\* Palmer Morrel-Samuels, 2002.

# EXAMPLES

**This:**

Never

Always

1

2

3

4

5

**Not This:**

Never

Almost  
Never

Sometimes

Almost  
Always

Always

1

2

3

4

5



# TIP 8 (MEASUREMENT)

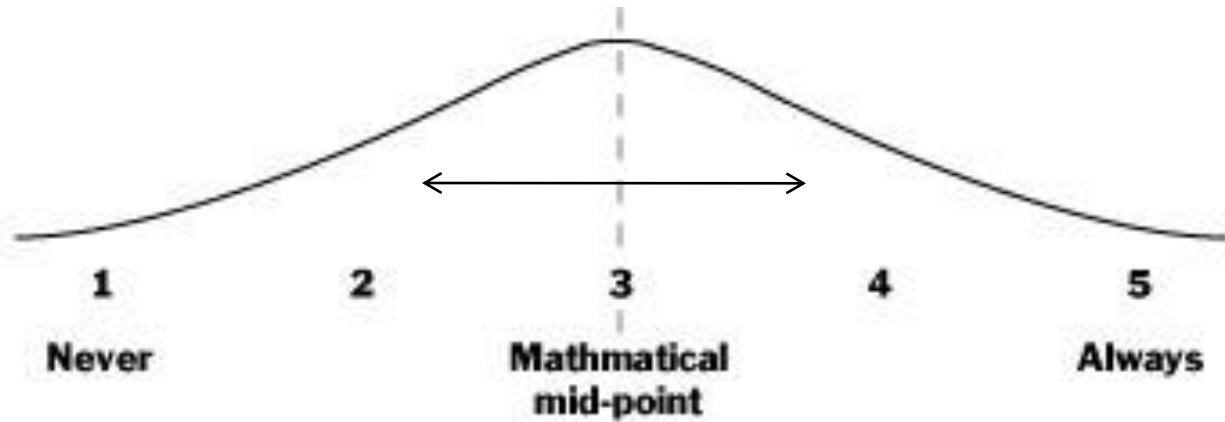
- ▶ Use only one response scale with an odd number of points (e.g. 5, 7, 9, 11)\*



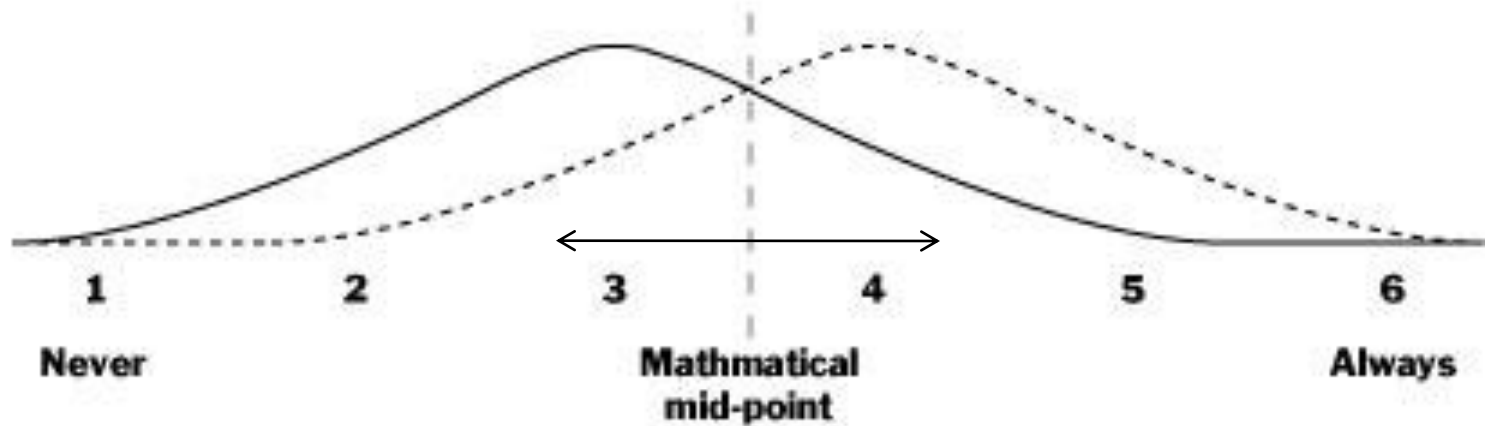
\* Palmer Morrel-Samuels, 2002.

# ODD vs EVEN SCALE

This:



Not this:



# TIP 9 (MEASUREMENT)

- ▶ **Use a response scale that measures frequency not agreement or effectiveness\***

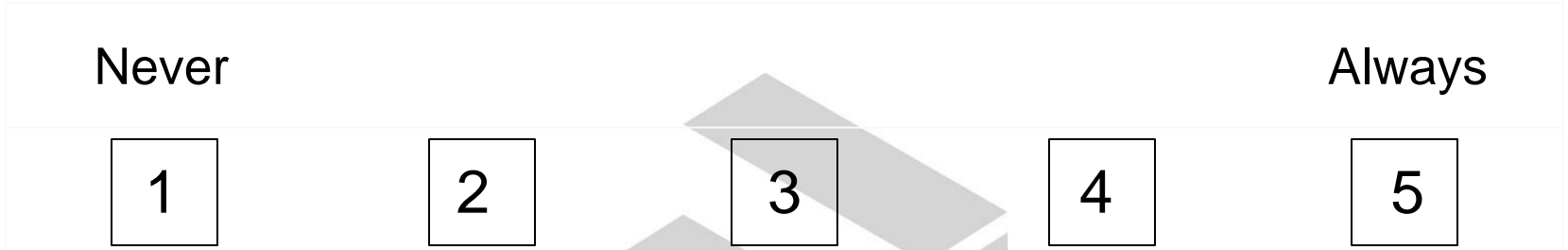


\* Palmer Morrel-Samuels, 2002.

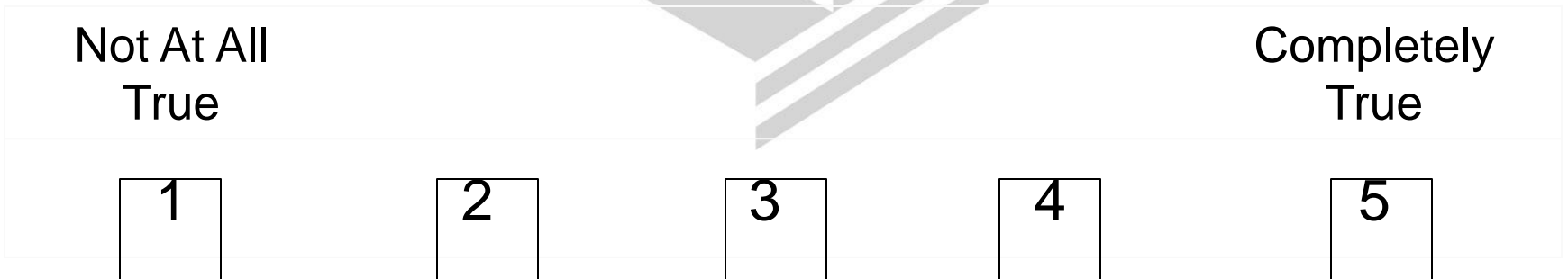


# EXAMPLES

**This:**



**Or this:**



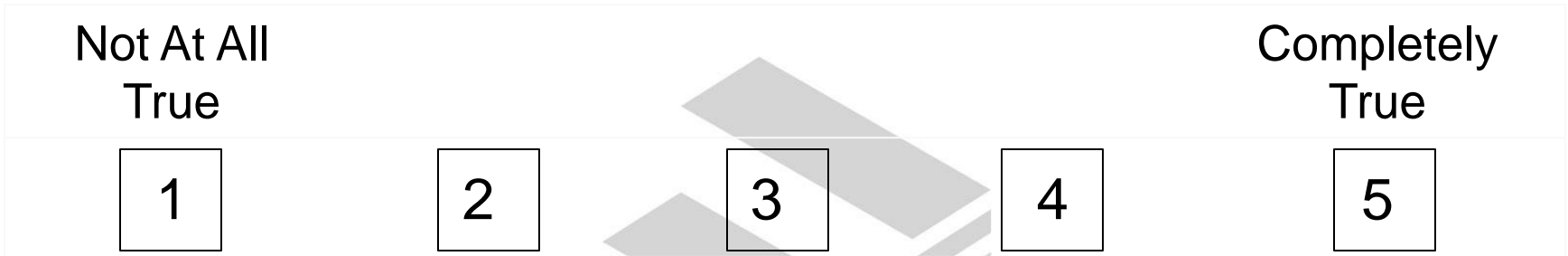
# TIP 10 (MEASUREMENT)

- ▶ Use small numbers at left or low end of scale and large numbers at right or high end of scale



# EXAMPLES

**This:**

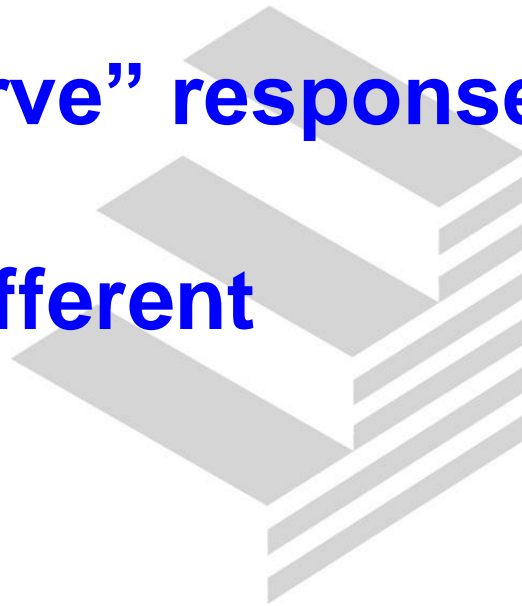


**Not this:**



# TIP 11 (MEASUREMENT)

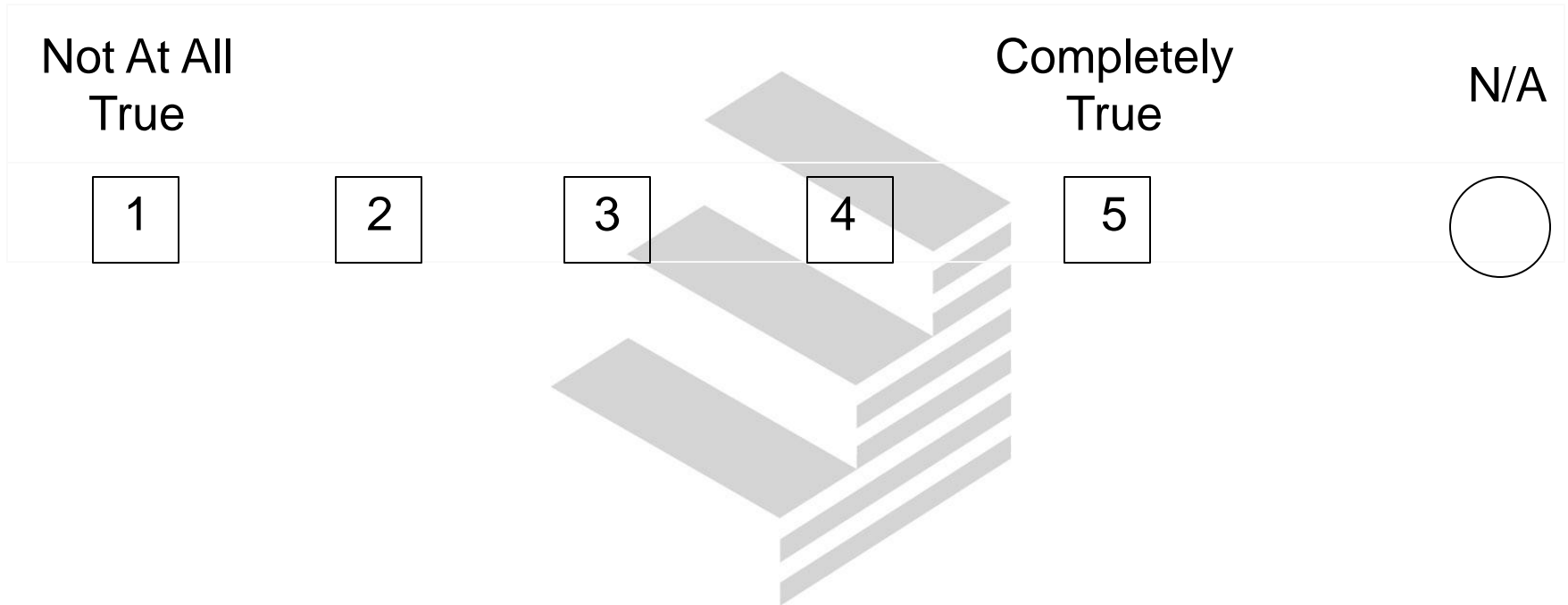
- ▶ **Include a “Not Applicable” or “Did Not Observe” response choice and make it different**



\* Palmer Morrel-Samuels, 2002.



# EXAMPLE



# SUMMARY

- ▶ Focus on observable behavior
- ▶ Limit ideas to a single description of behavior
- ▶ Word 1/3 of items as reverse score
- ▶ Keep survey sections unlabeled
- ▶ Design sections to contain similar number of items & questions similar number of words
- ▶ Collect data from multiple observers or multiple times



# SUMMARY

- ▶ **Create a response scale that:**
  - **Has words only at each end**
  - **Has an odd number of points**
  - **Measures frequency**
  - **Has small numbers at left and large numbers at right**
  - **Includes a “Not Applicable” that is different**



# REFERENCES

**Kirkpatrick, Donald & James, “Evaluating Training Programs: The Four Levels,” 2006.**

**Morrel-Samuels, Palmer, “Getting the Truth into Workplace Surveys,” *Harvard Business Review*, February 2002, pps. 111-118.**

**Phillips, Ken, “Eight Tips on Developing Valid Level 1 Evaluation Forms,” *Training Today*, Fall 2007, pps. 8 & 14.**

**Phillips, Ken, “Developing Valid Level 2 Evaluations,” *Training Today*, Fall 2009.**



**Ken Phillips**

**Phillips Associates**

**34137 N. Wooded Glen Drive**

**Grayslake, Illinois 60030**

**(847) 231-6068**

**[www.phillipsassociates.com](http://www.phillipsassociates.com)**

**[ken@phillipsassociates.com](mailto:ken@phillipsassociates.com)**

