Eight Tips On Developing Valid Level 1 Evaluation Forms*

by: Ken Phillips

Phillips Associates

*Published in *Training Today*, Fall 2007 (A quarterly magazine published by the Chicagoland Chapter of ASTD)

Eight Tips On Developing Valid Level 1 Evaluation Forms

By Kenneth R. Phillips

President, Phillips Associates

Level 1 evaluations, for better or worse, are a ubiquitous part of the workplace learning and performance landscape. In fact, according to research conducted by ASTD (ASTD 2005 State of the Industry Report), 91 percent of all learning events are evaluated at least at Level 1 of Kirkpatrick's four level evaluation model (Reaction, Learning, Behavior Change and Business Results). However, despite their widespread use, do you ever wonder about the validity of the results you obtain? You should. According to research conducted by Richard Clark and Fred Estes and published in a book titled *Turning Research into Results: A Guide to Selecting the Right Performance Solutions*, Level 1 evaluation results generally demonstrate a negative or inverse correlation with actual on-the-job behavior (Level 3). In other words, Level 1 evaluations often indicate the opposite of what actually happened in a learning program – either rating an effective program poorly or an ineffective program highly. This then raises a key question: "How do you design valid Level 1 evaluation forms?" The following eight tips are offered as suggestions.

Asking questions about things you can't do anything about to improve the effectiveness of a learning program wastes participant time and may eventually lead to participant frustration.

1. Only ask questions that lead to actionable data.

Asking questions about things you can't do anything about to improve the effectiveness of a learning program wastes participant time and may eventually lead to participant frustration. For example, asking a question about how effectively a training room contributes to participant learning when it has a post in the middle that obstructs participant view, but is the only training room available, is a waste of time. If the room can't be modified to eliminate the post or you're not collecting data to build a business case for getting a new training room, stop asking the question. Answering the same question over and over and not seeing anything change leads to participant frustration and a lack of interest in completing the evaluation form.

2. Write learner-centered evaluation items not trainer-centered.

This:
I found the room comfortable and conducive to learning.

Not This:
The room lighting and temperature were conducive to learning.

In a 2008 article titled "The New World Level 1 Reaction Sheets," Jim Kirkpatrick points out that most Level 1 evaluation items are written from a 'trainer-centered' rather than a 'learner-centered' perspective (see the examples above). Jim's point is that instead of asking participants for their thoughts about us and how well we clarified the learning objectives, organized the program material, kept the program moving, responded to their questions, etc., we should be asking participants questions about them and how they experienced the learning program relative to their own needs. He makes a good point. After all, we go to great lengths to make sure our learning programs are participant-centered so why shouldn't we follow the same model when developing our Level 1 evaluation forms?

3. Where appropriate, match up qualitative questions with quantitative measures.

Example:
In a word, how would you describe this session? _____

Using a number, how would you describe this session?

No Value				Great Value
1	2	3	4	5

Using qualitative questions and quantitative measures to assess the same dimension is an effective way to cross validate item results. Specifically, asking participants to rate a learning event in a word and then to rate it using a number enables you to see if the word descriptions match the numeric value. For example, if participants use words like "Outstanding", "Excellent", and "Great" to describe a learning event and then numerically rate it a 3.2 on a five-point scale, something is wrong – either the participants misunderstood one of the measures or the learning event wasn't as good as the words suggest. On the other hand, if the numeric rating associated with these same words was 4.5, you could feel very confident that the learning event was a success because of the high positive correlation between the words and the numeric rating. However, when using qualitative questions, keep the following caution in mind: Level 1 evaluations are typically administered at the conclusion of a learning event when participants have psychologically "checked out" and are physically ready to leave. Therefore, administering a Level 1 evaluation with lots of open-ended questions is a sure fire way either to get no response or thoughtless responses. The solution: either keep the number of qualitative questions to a minimum or limit the required response to a word or two such as in the example above.

Using qualitative questions and quantitative measures to assess the same dimension is an effective way to cross validate item results.

4. When collecting quantitative data using a Likert scale, create a response scale with numbers at regularly spaced intervals and words only at each end.

Example:
I felt engaged during the session because the facilitator kept the program moving.

Not at all True				Completely True
1	2	3	4	5

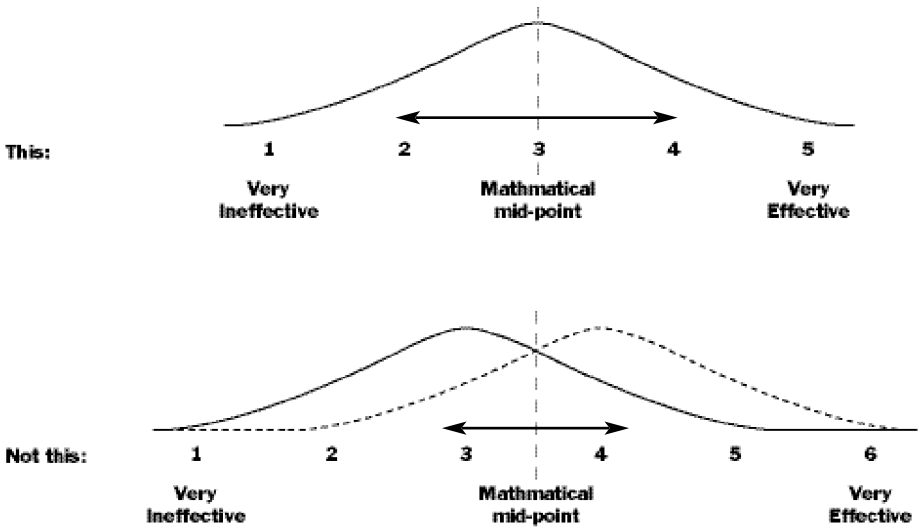
Many Level 1 evaluation forms use words to describe all the points along a scale. For example, in the scale above, words like "A Little True," Somewhat True" and "Mostly True" might be used to describe points 2, 3 and 4. However, as Palmer Morrel-Samuels points out in a 2002 *Harvard Business Review* article titled "Getting the Truth into Workplace Surveys,"

"The results from this type of evaluation [scale] are notoriously unreliable." He goes on to point out that because different words are used to describe each of the scale points, even though they may be in a plausible order, the distance between each pair of descriptors is not necessarily the same. For example, for many people, the distance between "Not At All True" and "A Little True" (points 1 and 2) may be closer to each other than "Mostly True" and "Completely True" (points 4 and 5) are to each other. Because of this, the response choices are no longer spread across an evenly spaced mathematical continuum thus making it difficult to conduct informative statistical tests on the results obtained. Another potential problem with labeling all the points on a scale identified by Morrel-Samuels is that often the descriptors overlap ("Mostly True" and "Completely True") and they may mean different things to different people thus making it difficult to compare results across groups. However, both these problems, as well as others created by word labels, can be eliminated by creating scales with only two word labels anchoring either end and a continuum of numbers in between.

However, both these problems, as well as others created by word labels, can be eliminated by creating scales with only two word labels anchoring either end and a continuum of numbers in between.

5. Use only one response scale with an odd number of points (e.g. 3, 5,7,9).

Again, according to Morrel-Samuels, single-scale evaluation forms, where the same two word labels are used to anchor either end of every evaluation item, are better than multiple-scale evaluation forms. Single scale evaluation forms take less time for participants to complete, provide more reliable data and make the comparison of results between different items easier. However, it may not always be possible to create a single scale Level 1 evaluation form and in these instances you should keep the number of different scales to a minimum as well as try to cluster the same scale items together.



Using an odd numbered scale with 5 to 9 response options is preferred over an even numbered scale of a similar length.

Using an odd numbered scale with 5 to 9 response options is preferred over an even numbered scale of a similar length. Odd numbered scales allow participants the option of choosing a neutral response, which is a perfectly valid response. Odd numbered scales also more readily allow for the possibility of obtaining a normal bell shaped curve distribution of responses across the scale because it has an actual mid-point. Even numbered scales, on the other hand, increase the possibility of obtaining a skewed distribution of responses above or below the actual mathematical mid-point, such as in the example above, because participants aren't allowed to register a neutral response. The net result is that something that scored highly or poorly may not be as good or bad as the scores suggest.

6. Use small numbers at the low or left end of the scale and larger numbers at the right or high end of the scale.

Example:
The learning activities used in this session helped me to achieve proficiency with the program material.

Not at all True				Completely True
1	2	3	4	5

Sometimes you'll see evaluation forms where the scale used runs in descending order or from high to low (e.g. 5, 4, 3, 2, 1) instead of low to high. However, this runs counter to the way we count and can create problems when participants are in a hurry to complete the evaluation form and mistakenly mark their responses at the right end of the scale thinking these are the better responses. The result is that good things about your learning event may come out looking bad and bad things may come out looking good. Although not as common, another mistake occasionally made on evaluation forms is to create a scale where low numbers represent positive responses and high numbers represent negative responses (e.g. 1 = Completely True and 5 = Not at all True). Here again the scale is counter intuitive because we generally associate higher numbers with better and may create the same kind of problem described above where good things get rated low and bad things get rated high.

7. Write items either as a continuum or as a statement.

Examples:
This: How effectively did the AV materials used during the session help reinforce your understanding of the program material?

Not Effectively				Very Effectively
1	2	3	4	5

Or This: The AV materials used during the session helped reinforce my understanding of the program material.

Not at all True				Completely True
1	2	3	4	5

Not This: Did the AV materials used during the session help reinforce your understanding of the program material?

Strongly Disagree				Strongly Agree
1	2	3	4	5

Another mistake some people make when creating Level 1 evaluation items is to write an item that asks for a "yes/no" answer and then use a Likert scale for recording responses. For example, the question "Did the AV materials used during the session help reinforce your understanding of the program material?" asks for a "yes" or "no" answer, not an answer that falls along a continuum such as the "This" and the "Or This" dexamples above. While this may not have an adverse effect on item results, at a minimum it defies logic.

8. Include at least one item asking participants how relevant the learning event/material was to them and their job.

Example:
How would you rate the overall relevance of this session to you and your job?

Not at all Relevant				Very Relevant
1	2	3	4	5

Does this mean that every Level 1 evaluation form should include at least one question asking participants how relevant the training was? According to research conducted by Neil Rackham, author of *SPIN Selling* and *Major Account Sales Strategy*, and reported in *Training* magazine, the answer is a resounding "Yes!" In fact, Rackham's research suggests that not only does a relevance scale have a high positive correlation with learning (Level 2), it also has a higher correlation with learning than an item evaluating participant learning.

In summary, Level 1 evaluations, while ubiquitous, often miss the mark because they are poorly designed and result in the capturing of misleading or invalid data. However, by following the tips described above you'll be able to improve the validity of the data you collect and make better decisions regarding improvements needed in your learning programs, which after all is the real purpose of Level 1 evaluations anyway.

However, by following the tips described above you'll be able to improve the validity of the data you collect and make better decisions regarding improvements needed in your learning programs...

For more information, contact:
Ken Phillips
Phillips Associates
34137 N. Wooded Glen Drive
Grayslake, IL 60030
(847) 231-6068
www.phillipsassociates.com
ken@phillipsassociates.com