Kirkpatrick's Evaluation Model

*The four levels of learning evaluation*

Donald L. Kirkpatrick, Professor Emeritus, University of Wisconsin, first published his ideas in 1959, in a series of articles in the Journal of American Society of Training Directors. The articles were subsequently included in Kirkpatrick's book *Evaluating Training Programs*. Kirkpatrick was president of the American Society for Training and Development (ASTD) in 1975.

Kirkpatrick's 1994 book *Evaluating Training Programs* defined his originally published ideas of 1959, thereby further increasing awareness of them, so that his theory has become arguably the most widely used and popular model for the evaluation of training and learning. Kirkpatrick's four-level model is considered an industry standard across the HR and training communities.

The four levels of Kirkpatrick's evaluation model essentially measure:
- **reaction of learner** - what they thought and felt about the training
- **learning** - the resulting increase in knowledge or capability
- **behaviour** - extent of behaviour and capability improvement and implementation/application
- **results** - the effects on the business or environment resulting from the trainee's performance

All these measures are recommended for full and meaningful evaluation of learning in organisations, although their application broadly increases in complexity, and usually cost, through the levels from level 1-4.

Kirkpatrick's four levels of training evaluation

This table illustrates the basic Kirkpatrick structure at a glance. The second table, beneath this one, is the same thing with more detail.

<table>
<thead>
<tr>
<th>Level</th>
<th>What is Measured</th>
<th>Evaluation description and characteristics</th>
<th>Examples of Evaluation Tools and Methods</th>
<th>Relevance, Practicality and Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reaction</td>
<td>Reaction evaluation is how the learners felt about the training or learning experience.</td>
<td>'Happy sheets', feedback forms. Verbal reaction, post-training surveys or questionnaires.</td>
<td>Quick and very easy to obtain. Not expensive to gather or to analyse.</td>
</tr>
<tr>
<td>2</td>
<td>Learning</td>
<td>Learning evaluation is the measurement of the increase in knowledge - before and after.</td>
<td>Typically, assessments or tests before and after the training. Interview or observation can also be used.</td>
<td>Relatively simple to set up; clear-cut for quantifiable skills. Less easy for complex learning.</td>
</tr>
<tr>
<td>3</td>
<td>Behaviour</td>
<td>Behaviour evaluation is the extent of applied learning back on the job - implementation.</td>
<td>Observation and interview over time are required to assess change, relevance of change, and sustainability of change.</td>
<td>Measurement of behaviour change typically requires cooperation and skill of line-managers.</td>
</tr>
<tr>
<td>4</td>
<td>Results</td>
<td>Results evaluation is the effect on the business or environment by the trainee.</td>
<td>Measures are already in place via normal management systems and reporting - the challenge is to relate to the trainee.</td>
<td>Individually not difficult; unlike whole organisation. Process must attribute clear accountabilities.</td>
</tr>
</tbody>
</table>
**Kirkpatrick’s four levels of training evaluation in detail**

This table illustrates Kirkpatrick's structure in detail, and particularly the modern-day interpretation of the Kirkpatrick learning evaluation model, usage, implications, and examples of tools and methods.

<table>
<thead>
<tr>
<th>Level</th>
<th>What is Measured</th>
<th>Evaluation description and characteristics</th>
<th>Examples of Evaluation Tools and Methods</th>
<th>Relevance, Practicality and Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reaction</td>
<td>Reaction evaluation is how the learners felt, and their personal reactions to the training or learning experience, for example: Did the learners like and enjoy the training? Did they consider the training relevant? Was it a good use of their time? Did they like the venue, the style, timing, domestics, etc? Was the level of participation sufficient? Ease and comfort of experience. Level of effort required to make the most of the learning. Perceived practicality and potential for applying the learning.</td>
<td>Feedback forms based on subjective personal reaction to the training experience. Verbal reaction which can be noted and analysed. Post-training surveys or questionnaires. Online evaluation or grading by learners. Subsequent verbal or written reports given by learners to managers back at their jobs.</td>
<td>Can be done immediately the training ends. Very easy to obtain reaction feedback. Feedback is not expensive to gather or to analyse for groups. Important to know that people were not upset or disappointed. Important that people give a positive impression when relating their experience to others who might be deciding whether to experience same.</td>
</tr>
<tr>
<td>2</td>
<td>Learning</td>
<td>Learning evaluation is the measurement of the increase in knowledge or intellectual capability from before to after the learning experience: Did the learners learn what was intended to be taught? Did the learners experience what was intended for them to experience? What is the extent of advancement or change in the learners after the training, in the direction or area that was intended?</td>
<td>Assessments or tests before and after the training. Interview or observation can be used before and after although this is time-consuming and can be inconsistent. Methods of assessment need to be closely related to the aims of the learning. Measurement and analysis are possible and easy on a group scale. Reliable, clear scoring and measurements need to be established, so as to limit the risk of inconsistent assessment. Hard-copy, electronic, online or interview style assessments are all possible.</td>
<td>Relatively simple to set up, but more investment and thought required than reaction evaluation. Highly relevant and clear-cut for certain training such as quantifiable or technical skills. Less easy for more complex learning such as attitudinal development, which is difficult to assess. Cost escalates if systems are poorly designed, which increases work required to measure and analyse.</td>
</tr>
</tbody>
</table>
| 3 | Behaviour | Behaviour evaluation is the extent to which the learners applied the learning and changed their behaviour, and this can be immediately and several months after the training, depending on the situation:

Did the learners put their learning into effect when back on the job?
Were the relevant skills and knowledge used?

Was there noticeable and measurable change in the activity and performance of the learners when back in their roles?

Was the change in behaviour and new level of knowledge sustained?
Would the learner be able to transfer their learning to another person?
Is the learner aware of their change in behaviour, knowledge, skill level?

|  |  | Observation and interview over time are required to assess change, relevance of change, and sustainability of change.

Arbitrary snapshot assessments are not reliable because people change in different ways at different times.

Assessments need to be subtle and ongoing, and then transferred to a suitable analysis tool.

Assessments need to be designed to reduce subjective judgement of the observer or interviewer, which is a variable factor that can affect reliability and consistency of measurements.

The opinion of the learner, which is a relevant indicator, is also subjective and unreliable, and so needs to be measured in a consistent defined way.

360-degree feedback is useful method and need not be used before training, because respondents can make a judgement as to change after training, and this can be analysed for groups of respondents and learners.

Assessments can be designed around relevant performance scenarios, and specific key performance indicators or criteria.

Online and electronic assessments are more difficult to incorporate - assessments tend to be more successful when integrated within existing management and coaching protocols.

Self-assessment can be useful, using carefully designed criteria and measurements.

|  |  | Measurement of behaviour change is less easy to quantify and interpret than reaction and learning evaluation. Simple quick response systems unlikely to be adequate.

Cooperation and skill of observers, typically line-managers, are important factors, and difficult to control.
Management and analysis of ongoing subtle assessments are difficult, and virtually impossible without a well-designed system from the beginning.

Evaluation of implementation and application is an extremely important assessment - there is little point in a good reaction and good increase in capability if nothing changes back in the job, therefore evaluation in this area is vital.

Behaviour change evaluation is possible given good support and involvement from line managers or trainees, so it is helpful to involve them from the start, and to identify benefits for them, which links to the level 4 evaluation below. |
Results evaluation is the effect on the business or environment resulting from the improved performance of the learner - it is the acid test. Measures would typically be business or organisational key performance indicators, such as: Volumes, values, percentages, timescales, return on investment, and other quantifiable aspects of organisational performance, for instance; numbers of complaints, staff turnover, attrition, failures, wastage, non-compliance, quality ratings, achievement of standards and accreditations, growth, retention, etc.

It is possible that many of these measures are already in place via normal management systems and reporting. The challenge is to identify which and how they relate to the learner's input and influence. Therefore, it is important to identify and agree accountability and relevance with the learner at the start of the training, so they understand what is to be measured. This process overlays normal good management practice - it simply needs linking to the training input. Failure to link to training input type and timing will greatly reduce the ease by which results can be attributed to the training. For senior people particularly, annual appraisals and ongoing agreement of key business objectives are integral to measuring business results derived from training.

Individually, results evaluation is not particularly difficult; across an entire organisation it becomes very much more challenging, not least because of the reliance on, line-management and the frequency and scale of changing structures, responsibilities and roles, which complicates the process of attributing clear accountability. Also, external factors greatly affect organisational and business performance, which cloud the true cause of good or poor results.

The Kirkpatrick model is recognised as a robust and effective method of evaluation. It lends itself to evaluating the effectiveness of a coaching approach to learning and development, particularly as the learner can play an active part in the design of the evaluation process.