Evaluation of learning – a family affair?
A white paper by Imogen Casebourne & Andrew Downes
Most workplace training is commissioned with a specific organisational goal in mind, be that to prevent workplace accidents, improve efficiency or furnish workers with new skills. However, the required outcomes aren’t always clearly defined at the outset and it can be difficult to be sure how much difference training has made.

Evaluation allows the organisation to find out to what extent a particular training initiative has met its target. As a process, evaluation needs to be planned and carried out in tandem with training design, but requires a slightly different set of skills; those of a researcher rather than a designer.

In this white paper, we outline the industry standard Kirkpatrick and Phillips evaluation models, introduce the newer Brinkerhoff Success Case methodology and cover various methods for data collection.

We also consider why evaluation isn’t carried out more often, and address some of the difficulties and challenges involved in carrying out effective evaluation. Finally, we offer some case studies which may give inspiration as you plan your own evaluation initiative.

First though, an observation: the two key figures involved in bringing rigour to the evaluation of training, Donald Kirkpatrick and Jack Phillips, both still operate US consultancies offering evaluation. And both work with family members: Kirkpatrick with his son James and daughter-in-law Wendy, and Phillips with his wife Patti. Evaluation, it would seem, is often a family affair.

Why evaluate?

Training is all too often seen as a costly overhead rather than an essential contributor to business success. This suggests that, while they may pay to commission it, deep down many CEOs don’t truly believe that training can make a positive impact to the bottom line, and it is certainly the case that they frequently opt to cut back on it when times get tough; a danger that then-ASTD president, Donald Kirkpatrick, identified back in the fifties. His solution to this was evaluation.

Rigorous evaluation of the effectiveness and impact of training offers training departments and the organisations which employ them the ability to see where learning interventions have made a difference and moved things in the required direction, and also where they haven’t.

In fact, it offers the opportunity to be viewed as strategic business partners, rather than an expendable cost-base.
What is the Kirkpatrick evaluation model?

The Kirkpatrick model was first outlined in a series of articles published from 1959 collected as Techniques for Evaluating Programs and then in Evaluating training programmes: The four levels (1994). Donald Kirkpatrick proposed a standard approach to the evaluation of training that has become a de facto standard in the industry.

Kirkpatrick identified four levels of evaluation, which are described in the following table.

Building on Kirkpatrick’s work, Phillips later added a fifth level, Return on Investment (ROI), which compares the cost of developing and delivering the training compared to the cost benefits achieved by the training. The costs, benefits and payback period are fully evaluated in relation to the training deliverables.

Below we briefly outline Kirkpatrick’s four levels.

<table>
<thead>
<tr>
<th>Level</th>
<th>Target</th>
<th>Evaluation goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reaction</td>
<td>Training</td>
<td>Initial endorsement by participants of the training</td>
</tr>
<tr>
<td>2. Learning</td>
<td>Learner on course</td>
<td>That learning occurred as a result of the training</td>
</tr>
<tr>
<td>3. Behaviour</td>
<td>Learner on job</td>
<td>That learning affected behaviour, or performance on the job</td>
</tr>
<tr>
<td>4. Results</td>
<td>Organisation</td>
<td>That the training had the desired results in the organisation</td>
</tr>
</tbody>
</table>

**Level one (reaction)**

At reaction level, we ask learners to comment anonymously on the adequacy of the training, the approach and perceived relevance. The goal at this stage is simply to identify glaring problems – not to determine whether the training has worked.

**Level two (learning)**

The learning level is more formal, requiring a pre- and post-test. This allows us to identify those who had existing knowledge, as well as those at the end who missed key learning points. It is designed to determine whether the learners actually acquired the identified knowledge and skills as a result of the training.

**Level three (behaviour)**

At the behavioural level, we attempt to measure the transfer of the learning to the job. Data is gathered from questionnaires or interviews with the learner, their peers and managers, and might even have involved observations. In future, technologies like XAPI hold out the promise that we can start to track some job tasks automatically, particularly when these tasks are carried out using a computer or internet-enabled machine.

Behavioural evaluation can include an immediate evaluation after the training and a follow-up after a couple of months.

**Level four (results)**

The results level looks at improvement in the organisation.

The organisation generally already has data and reports a variety of key performance indicators (KPIs) that may be relevant to the e-learning. The challenge here as learning professionals is to identify which KPIs are most relevant and then establish a link to the uptake of our learning.

More about Kirkpatrick

You can learn more about Kirkpatrick’s work by visiting his website, which offers a wealth of insight and resources: http://www.kirkpatrickpartners.com/
What is the Phillips evaluation model?

Jack Phillips took Kirkpatrick’s work and sought to extend it, by also seeking to evaluate return on investment or ROI. He has elaborated his thinking in numerous books published over the past decades.

In brief, Phillips advocates five levels of evaluation: Kirkpatrick’s initial four, which we met in the previous section, plus the fifth, added by Phillips, ROI. Phillip’s five levels are outlined in the following table.

<table>
<thead>
<tr>
<th>Level</th>
<th>Target</th>
<th>Evaluation goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reaction</td>
<td>Training</td>
<td>Initial endorsement by participants of the training</td>
</tr>
<tr>
<td>2. Learning</td>
<td>Learner on course</td>
<td>That learning occurred as a result of the training</td>
</tr>
<tr>
<td>3. Behaviour</td>
<td>Learner on job</td>
<td>That learning affected behaviour, or performance on the job</td>
</tr>
<tr>
<td>4. Results</td>
<td>Organisation</td>
<td>That the training had the desired results in the organisation</td>
</tr>
<tr>
<td>5. Evaluation</td>
<td>Value for money</td>
<td>That given the costs related to the desired impact on the organisation, and the costs of creating the training, the training represented good value for money.</td>
</tr>
</tbody>
</table>

You can learn more about Phillips and his work by visiting his website and accessing his ‘At a glance’ series of publications: [http://www.roiinstitute.net/roiataglance](http://www.roiinstitute.net/roiataglance)
Kirkpatrick argues, and we’ll come back to this later, that all levels are essential for a thorough evaluation. However, it is often only the first two that are addressed, if evaluation is undertaken at all, and this is unlikely to make a convincing picture.

In fact, Van Buren’s 2001 survey for the ATD (Association for Talent Development) showed that organisations measure training outcomes in reverse order of importance.

**Common evaluation mistakes and how to avoid them**

Kirkpatrick argues, and we’ll come back to this later, that all levels are essential for a thorough evaluation. However, it is often only the first two that are addressed, if evaluation is undertaken at all, and this is unlikely to make a convincing picture.

In fact, Van Buren’s 2001 survey for the ATD (Association for Talent Development) showed that organisations measure training outcomes in reverse order of importance.

**Level one evaluation** is the perhaps easiest to undertake and therefore the most often done – it involves presenting learners with a questionnaire, often referred to as a happy sheet, asking them about the learning experience.

The problem with a happy sheet is that it tells us little more than the immediate reaction of learners to a course, often measuring emotional and social satisfaction, rather than any really useful lessons about learning and there is evidence from numerous studies that satisfaction can’t be linked directly to learning.¹

On the contrary, we know intuitively that memorable and effective training is not necessarily a pleasant experience. In his book, The Social Neuroscience of Education, Louis Cozolino talks of needing to support students to cope with ‘the stress of new learning’.² Training that challenges our current understanding of the world or shakes existing self image or beliefs may be upsetting and may consequently not score well on a ‘happy sheet’ as a result.

However, by asking the right questions, it is possible to gather useful evaluative data at this level and that data can form part of the trail of evidence that eventually indicates whether training has been successful.

In evaluating the learners’ reactions to an e-learning course, you should be concerned with:

- pre-course preparation
- the learner’s reaction to the course
- how much the learner anticipates being able to apply the content of the course
- any housekeeping difficulties encountered by the learner.

Level two evaluation should yield more data about whether learning has taken place, but where it is carried out, it isn’t always conducted as well as it might be.

Too often assessments in e-learning test whether learners have memorised relatively unimportant facts and information, rather than whether they can produce required behaviours under the appropriate circumstances.

Additionally, where assessments are only taken minutes after the course, they may function as a test of short-term memory rather than deeper learning. If the learning is a ‘one-off’ intervention, level two evidence of learning gathered right after the course is quite unlikely to translate to ability to produce the behaviour...
in the workplace at a later date when, even if it was truly 'learned' in the first place, it is quite likely to have been forgotten again.

Research into memory and retention suggests that in order to deeply embed learning so that it is not forgotten, it may be necessary to repeat practice and test opportunities for some time.³

Deciding exactly how to assess a training initiative at level three is therefore not a straightforward question and one which Kirkpatrick rightly argues should be closely linked to the original learning requirements and overall course design.

Level three evaluation seeks to find out whether new learning is actually resulting in new behaviours in the workplace. This is a test of whether learning is being applied, rather than whether it has been learnt – a subtly different thing. Of course, if learning isn’t being applied, it may not be the fault of the training but a result of other factors in the workplace which are acting to prevent new learning being adopted.

Level three evaluation is considerably more difficult, time consuming and expensive to carry out effectively than the first two levels, which is probably why it is conducted less often.

A common assumption is that the impact of training can only be measured by looking at the average or overall changes that can be attributed to a training programme. In other words, the assessor asks, “By what percentage have corporate sales increased since sales staff undertook this training programme?” But Brinkerhoff’s Success Case Method is based on a very different view.⁴

The Success Case Method first follows up training with a short questionnaire to all trainees. Trainees assess the extent to which the training has actually improved their performance and, if not, what obstacles to success they have experienced. The method then concentrates on the two extremes: those who report the greatest benefits, and those who report the greatest obstacles. This step involves in-depth interviews to quantify the benefits in terms that will ‘stand up in court’ ⁵; in other words, evidence of increased profitability, better customer care or return on investment (ROI).

At the same time, the interviews find out the obstacles to success. For example, they may identify that a course in presentation skills has been very successful in changing the behaviour of client-facing staff but that the learning was also undertaken by technical staff with no client-facing role. This insight would lead to a change in recruitment to the course, so that it was only delivered to staff who actually needed the skills, thus improving the ROI by cutting the overall cost of course administration.

The Success Case Method has proven value in linking training to business objectives by tying learning to results (the fourth and highest of Kirkpatrick’s four levels of learning). It also demonstrates that it’s beneficial to concentrate on the extremes of success and failure instead of looking for the average benefit, and then presenting the results in qualitative terms: success stories which can motivate, and barriers that can be overcome. This canny approach also saves money!

Level four evaluation seeks to identify the impact to the bottom line that results from trainees learning and then practicing new behaviours during and after the training session.

Although this is commonly considered to be the most difficult level of evaluation to carry out, in fact the information needed is usually easier to gather than level three data because it is generally already there, in the form of data already collected by other departments, such as HR and sales.

Things will be much easier, however, if you follow Jack Phillips advice and gather level four data before you launch your training initiative, as well as afterwards. It may be possible to identify ‘before and after’ data after the fact, but it will be more straightforward to gather the ‘before’ data before the initiative is launched.

Of course, it isn’t always easy to show that it is training, rather than other factors, which has positively impacted the bottom line – more of which later.

Recent research undertaken by Phillips found that whilst 96% of Fortune 500 CEOs wanted to know the business impact of their learning and development, only 8% saw this happening now ⁶.

While the previous section addressed common failings in the implementation of the Kirkpatrick model, it is too often the case that no evaluation is attempted. There

Why isn’t training evaluated as a matter of course?

may be several reasons for this:

> Lack of interest
> Cost
> Disruption
> Difficulty
> Statistical doubts
> Fear

Lack of interest

The first reason is perhaps the most common. Many organisations have, historically, never received such data, so don’t expect any. They assume that training is effective and regard it as an ‘act of faith’. It’s pointless creating evaluation reports if the organisation’s management is uninterested in reviewing them.

This attitude is changing, as all ‘cost centres’ come under ever greater scrutiny. As training department budgets continue to be cut, there will be increased emphasis on predicting and demonstrating a return on investment for training before it is commissioned.

Cost

Serious evaluation can be costly. It has to be budgeted to include design, planning, implementation, analysis and reporting. Full Kirkpatrick analysis is expensive. The cost can therefore be considerable, especially if objectivity is expected by buying external expertise.

Ask the simple question “Who will pay for evaluation?” Is it the business sponsor? The training department? The decision to evaluate often stops at this point.

Technological developments such as XAPI API mean that the costs of evaluation may soon start to come down. Consider an organisation that has evaluation data for an information security course and also evaluates information security in the organisation, but has no way of this linking the two data sets. XAPI allows us to bring all this data together in the same format, reusing work that’s already being done. A reporting tool can then automatically compare this data, producing charts and tables. We can even trigger events as a result of the data, for example automatically sending a learner performance support materials or alerting a manager.

In some contexts, this will make evaluation relatively cheap in the long run, although this will involve investment to build the system in the first place. As XAPI becomes more common though, expect generic off the shelf reporting tools to be developed that will help to bring your costs down.

Disruption

Evaluations can be experienced as disruptive and can eat up valuable resource. This includes both trainers and employees, who are all under pressure to be productive. Statistically relevant evaluation can need significant resource, which is rarely available during and after the course has been delivered.

Again, XAPI may help in future. By tracking what’s already happening, any disruption is minimal. If XAPI were to be integrated into work-based computer systems such as a sales database, it could be used to track learners seamlessly as they carry out their day-to-day work. Our definition of ‘computer’ can be quite broad here: any device that can connect to the internet can send record tracking data, even storing data up until a connection becomes available. Machines, vehicles, till systems, handheld devices and tools could all potentially be used to track job performance.

Where work is not computer based, tracking may be more disruptive. Even then though, there may be creative and innovative ways to evaluate learning with minimal disruption. A 360 review tool, for example, can be used to evaluate employees. Such a system might be no more disruptive than an email to the employee’s manager or peers asking them to click a five star rating scale for the employee’s competency in a given area.

Difficulty

It’s easy to feel that the task is just too difficult. Not only must appropriate data be gathered but it must then be analysed and presented. There may be low evaluation skills in the training department. Evaluation is a skills-
set that, because it is so rarely applied, is usually in
short supply. The ability to handle the implementation,
analysis and reporting often requires a knowledge of
statistics beyond what a given department possesses.

**Statistical weakness**

Unless the number of learners is high, the data may be
statistically weak. Evaluative studies, including return
on investment calculations in learning, have been
criticised for lack of rigour and therefore significance.

For cost reasons, not enough data may be collected at
level three, which is less easy to automate, and overall
conclusions may therefore be weaker.

**Fear**

‘As long as trainees are happy and do not complain,
trainers feel comfortable, relaxed, and secure.’ These
words were written by Donald Kirkpatrick himself.

A key reason why trainees may opt not to evaluate is
the fear that evaluation will show up training as less
effective than expected.

This could lead to a reduction in budget or status. This
takes us back to the first point. If senior management is
not actively asking for this data, why put your neck on
the line? If it ain’t broke, don’t fix it.

The answer, of course, is twofold. Firstly, as responsible
learning developers, we should surely always be
seeking to find ways of making training ever better.
Secondly, senior management may not be asking
today, and they may not ask tomorrow, but the day
after tomorrow they could decide to axe the training
department as an unneeded overhead. It pays to be
proactive rather than reactive.

**Should you evaluate everything?**

Probably not. Full-scale evaluation, however efficiently
conducted, does add to overall costs. This may not
make sense for very small, low-cost initiatives (although
this is no reason why you shouldn’t design them to make
any later evaluation straightforward – clearly identifying
business objectives and relating these to measurable
learning objectives).

It makes sense to focus evaluation efforts on large scale
and high-profile training initiatives.

**How to evaluate – some high-level decisions**

So you’ve decided you will evaluate a training initiative,
and you want to do it thoroughly. How do you set about
it? This section gives some pointers.
When should you evaluate?
Once you have determined why and what you want to evaluate, ask yourself when you want to evaluate.

Gather data before designing the training
The more data you can gather on the current situation, as well as the change you and the business hope to see following delivery of training, the easier it will be to evaluate the success of that training after it has been rolled out (and the more likely it is that the training will be effective).

There is a strong argument, put forward by Dan Tobin and others, that puts emphasis on making the business case for training upfront and aligning that case with the business objective and the business measures. Tobin points out that ‘in the world of corporate financing, ROI is a forward looking calculation, not an after-the-fact justification of an investment’.

Tobin also makes the point that the success of a training initiative in improving performance by changing behaviour is strongly contingent on factors unrelated to the training itself, particularly the willingness of managers to support adoption of new behaviours in the workplace. Failure at level three (changed workplace behaviour) may not be a failing of training at all, but a wider corporate failure to support the change. This is a valid point. The implications for training departments are two-fold: firstly, to thoroughly plan for success, training designers must prepare the managers who will oversee workplace implementation as well as plan for training delivery, and they should also design appropriate performance support aids, and secondly, level three evaluations will be key in throwing up non-training related implementation issues.

For the above reasons, Tobin says he won’t carry out ROI studies. However, Tobin’s two very valid points do not preclude carrying out evaluation after a training initiative has been launched to see whether the initiative has met projected ROI predictions and has had the expected impact.

Following Tobin’s advice and aligning initiatives to a clearly defined business need upfront, and planning beyond the training department, will make subsequent evaluation much more straightforward. You will have a clear picture of the situation before the training and the change you hope to see as a result of the training. It will then be much more straightforward to collect data after the training and compare that to the other two sets of figures.

Hard data on business performance can be found in a number of places, including:

- profit
- sales
- increased market share
- shareholder value
- reduced complaints
- reduced transaction costs
- the number of calls handled per hour
- fewer defects
- delivery on time.

Cathy Moore refers to the process of gathering upfront data as action mapping and offers valuable advice on how to do this9. Business goals, required actions to achieve the goals and any other success criteria are defined upfront before the solution is designed. After the project, the impact can be measured against those success criteria. This approach not only makes evaluation easier but helps to filter out problems where e-learning is not necessarily the best solution.

Consider pre-launch evaluation of training to weed out errors
Pre-launch evaluation is the evaluation of training interventions prior to launch through user-centred design and focus groups. This can be done quickly and cheaply, and will filter out poor quality training. Your organisation has a lot to gain by reducing unnecessary training time and increasing the effectiveness of the training. When a large confectionery company launches a new chocolate bar, it does not launch it and then ask everyone who ate the bar what they thought of the experience. They do extensive pre-launch and market tests prior to launch.

Consider post-learning comparative studies
Comparative studies will only be possible if it is feasible and ethical to have one group of individuals act as a control group and not receive any training until the

evaluation period is complete.


This approach is particularly valuable if no data was gathered before the training was commissioned, developed and rolled out, as they enable a clear comparison between people who have received training and people who have not.

Comparative studies involve taking data from comparative groups; for example, a comparison study might seek to assess the difference between learners who have gone through:

1. no learning (control group)
2. e-learning.

An alternative is to add a third learning group using a different learning approach:

3. No learning (control group)
4. Classroom learning

<table>
<thead>
<tr>
<th>GROUP</th>
<th>NUMBER OF FAILURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>No learning at all (control group)</td>
<td>28</td>
</tr>
<tr>
<td>Classroom training</td>
<td>14</td>
</tr>
<tr>
<td>E-learning</td>
<td>7</td>
</tr>
</tbody>
</table>

A more sophisticated method is to use the data to create standard deviation curves, and to compare these.

<table>
<thead>
<tr>
<th>VS.</th>
<th>NO LEARNING</th>
<th>CLASSROOM TRAINING</th>
<th>E-LEARNING</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO LEARNING</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>CLASSROOM TRAINING</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>E-LEARNING</td>
<td>0.25</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

If this approach is followed, it is essential that the sampling of individuals for the comparison groups is rigorous. The group who don’t receive the training should consist of a range of people with similar characteristics (demographics) as the people who do receive the training. To make this clear consider the following proposals for comparison groups:

5. E-learning

Positive transfer is shown to have occurred if one group outperforms the other. Negative transfer is shown to have occurred if either is worse than the control group.

It is possible to quantify this using measurable criteria. The simplest approach is to identify a fault, skills gap or competence that can be clearly identified, then count the failure frequency. If the failures add up to 28 over a set period for the ‘no-learning’ group, 14 for the ‘conventional learning’ group and only 7 in the ‘e-learning’ group, then the transfer of training ratio would be calculated as follows: We can then compare the effectiveness of different types of learning:

We can then compare the effectiveness of different types of learning:

Based on our fictional data, the table illustrates that no learning is half as effective as classroom training, and the e-learning is four times as effective as no learning.
Proposal one
> control group: employees with ten years service – no training
> e-learning group: employees with less than ten years standing

Proposal two
> control group: half employees with ten years standing, half employees with less then ten years standing – no training
> e-learning group: half employees with ten years standing, half employees with less then ten years standing

In the above rather over-simplified example, the sampling in proposal one will lead to unusable data because any noticeable differences between the two groups could be influenced by the difference in experience as well as by having received or not received the training.

On the other hand, proposal two involves two groups containing a similar demographic mix, so it will be much more straightforward to identify any impact caused by training.

These comparisons can also be made to relate directly to the business by taking two demographically similar retail outlets or offices and comparing the results of one against the other. This gives direct access to measurable business outcomes such as margin or sales.

How can you link training results to business objectives?
Categorically linking training to positive business outcomes is difficult because there are usually many other factors in play at the same time as the training initiative and these could also have influenced business outcomes.

It is hard or impossible to separate out the impact of training from the impacts of other unrelated factors. In a recent book Donald Kirkpatrick’s son and business partner, James Kirkpatrick, advocates providing a chain of evidence linking training success to positive business outcomes, rather than seeking definitive proof.10

His argument is that if the training department can prove that a training initiative was well received, that trainees had learnt something at the end of it, that some time after the training positive behavioural changes could be observed in the workplace and that all this was followed by an improvement in related business metrics, then the chain of evidence linking the success of the training programme to the improved bottom line is very strong, even if there is no definitive proof that the training caused the positive change.

He points out that this type of chain of evidence can be enough to get you convicted in court (citing a case involving Martha Stewart), and is also likely to be enough to persuade the CEO that training is really having a positive business impact.

Should you carry out ROI calculations?
Return On Investment (ROI) compares the investment in a training deliverable with the eventual cost benefit over a specified period of time.

Costs, benefits and payback period can be fully evaluated in relation to productivity increases, cost savings, etc. There are two major components in ROI:

> benefits
> costs.

A basic ROI gives an overall figure (or can be expressed as a percentage) based on calculating benefits and costs.

<table>
<thead>
<tr>
<th>EXPRESSED AS</th>
<th>FORMULA</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A figure</td>
<td>Benefits – Costs = £</td>
<td>£ (£180,000 - £80,000) = £100,000</td>
</tr>
<tr>
<td>A percentage</td>
<td>(Benefits / Costs) x 100 = ROI%</td>
<td>(£180,000 - £80,000) / £80,000 x 100 = 125%</td>
</tr>
</tbody>
</table>

It may be necessary, in some cases, to calculate a payback period to give a more accurate figure. The addition of a payback period tells you how long it takes for the training to break even and therefore pay for itself.

How to gather evaluation data

This section gives an overview of data gathering techniques.

To evaluate a programme with typical members of your target audience, you can use paper questionnaires, telephone interviews, voiced user experiences or electronic gathering of data and analysis.

End-point paper questionnaires are cheap and easy to distribute. The problem is in gathering and analysing the data. The data has to be input by hand, which is time consuming and costly.

**Conduct interviews**

Telephone interviews are a quick and relatively cheap way of evaluating. This is a good tactic for a target audience that is geographically widespread and where time is tight. Data from users is more useful after learners have returned to work, relying less on short-term reactions.

Interviews during use allow the user to ‘voice’ concerns and problems as they happen. This ‘voicing’ technique is commonly employed in usability testing and is a powerful technique when evaluating reactions to the learning experience. It does not rely on the learner remembering what happened throughout the course, and avoids the ‘recency’ effect, where the learner makes judgements on what they’ve last seen or done. It is time consuming but useful in uncovering design flaws.

**Consider online questionnaires**

In e-learning, a software evaluation questionnaire is common. You will have to design and build this tool but, if well designed, it can be used time and time again. Tools such as Survey Monkey or Google spreadsheets can make creating a quick questionnaire very easy indeed.

Questionnaires can be made to appear before the final assessment, before leaving the e-learning course or as a standalone tool, perhaps automatically emailed a few days after the course is completed. Questionnaires provide an automated, low-cost method of gathering data.

A range of question types can be presented electronically including:

- binary choices
- multiple choice
- scales
- ranking
- open input.

In framing questions it is important to avoid ambiguity, complexity, leading statements and assumptions.

**Consider automating some elements of evaluation**

Tracking data, showing who did what, where, when and how will also give you a wealth of information about the users in relation to the learning content. How many times did the user go back? Where did he or she fail questions? What support functions did they access?

With XAPI, this data can now include tracking of the learner in their day-to-day job role. Activity and achievement in the e-learning can be compared with behaviour and success in the real world.

Other advantages of technology based evaluation include:

- lower costs
- speed
- objectivity
- better presentation
- automatic analysis
- automatic reporting
- automatic archiving.

Consider using case studies to evaluate at level three

The Brinkerhoff success case method is a naturalistic approach to evaluation which avoids the challenges associated with gathering enough data to generate statistically meaningful results.11
When following the success case method, evaluators seek out the individuals for whom the programme has proved a resounding success and try to find out what it is about these individuals and their circumstances that have made the training such a success for them.

This involves three key steps. Step one (searching for success cases) usually involves conducting a survey and step two (finding out what has made the programme successful for these individuals) generally involves structured or semi-structured interviews. Step three involves writing up the results.

Interviewing just those individuals for whom a programme has worked well might sound rather one-sided but, at the same time, evaluators will generally also seek out those individuals at the other end of the spectrum for whom the training has had no positive impact and try to find out why this is the case for them.

At the end of the process, having generally spoken to a much smaller group of people than would be needed with a more quantitatively driven approach, evaluators have information about the conditions in place where individuals have been in a position to apply the training in the workplace, and conditions in place where individuals have been unable to do so.

Comparing the two sets of information gives useful pointers as to what may be facilitating transfer of training to the workplace and what may be preventing it.

What about evaluating informal learning?

In his Informal Learning Basics, Saul Carliner outlines an approach for evaluating informal learning, beginning with a survey of the challenges in applying Kirkpatrick’s model to informal learning. Carliner’s main criticism is that the Kirkpatrick model relies on evaluating learning experiences against specific objectives, whereas informal learning, Carliner suggests, often does not have formal objectives. In place of Kirkpatrick, Carliner suggests a two-part evaluation model evaluating both individual learning and the use of informal learning resources.

Individual learning

According to Carliner, evaluating individual learning involves collecting data about what learners learnt and how they learnt it. This learning can then be recognised through learning records, certifications, badges or other rewards.

Tracking and rewarding informal learning is made much easier with new technologies such as the XAPI API. For example, using XAPI, a learner can record their experiences and achievements in a simple learning journal tool. These can then be mapped to particular competencies and rewarded as appropriate.

One common criticism of using learning journals or other tools to record learning is that it is difficult to motivate employees to record their learning. In some situations we are able to avoid this issue by using technology to automatically track an experience by embedding tracking code within the systems and tools the learner uses to complete a task.

We shouldn’t write off learning journals too quickly though. Motivation is a challenge, but we’ve successfully motivated employees to complete online multiple choice quizzes for years. The same carrot and/or stick approaches can be applied to learning journals, with a little extra effort at first to change the culture. When considering your motivational strategy, think about how the data will be used in a meaningful way for everybody involved. How will you get buy in?

Informal learning resources

Carliner also encourages us to evaluate the effectiveness of our informal learning resources. Here, Carliner seems to apply a version of Kirkpatrick, asking us to consider what resources were most used, how satisfied learners were with the resources (Kirkpatrick level one) and the impact of those resources. Carliner suggests a number of approaches to evaluate impact which map to various levels of Kirkpatrick from one to four.

Again, XAPI has the potential to help to evaluate learning resources, from tracking how often they are used and by whom, through evaluating resources to evaluating impact. We can use XAPI tracking data, for example, to identify which resources are most commonly used by the employees who make the least errors or who are most rapidly promoted.

Let’s look now at some case studies which illustrate approaches to carrying out evaluation on all four levels. In both cases, the evaluation showed that the e-learning had been effective.
Evaluation case studies

British Airways

Every year, British Airways carries 35m passengers and records 30,000 medical incidents. The airline has very high standards in aviation medicine, over and above those required by the CAA, the aviation regulator. This ensures staff deliver the highest standard of medical care to passengers.

Aviation Medicine is critical to the overall cabin crew induction – passing the final assessment provides a CAA licence to fly. An intensive six-week classroom induction was proving highly effective at providing the skills required to deliver a quality service to BA’s customers.

In early 2009, following a corporate initiative to increase speed to market and increase agility, BA decided to review the induction, including the AvMed elements. Reducing the six-week course by introducing e-learning would ensure crew were competent to fly in a shorter timeframe and reduce the overall cost of induction.

The e-learning (developed by LEO) has now been evaluated at all four levels of Kirkpatrick’s evaluation model for workplace education:

Level one (reaction)
Feedback in online surveys from delegates rated the programme highly – not a single respondent scored it below ‘good to extremely well received’

Level two (learning)
Before AvMed, learner knowledge was rated at an average of 3.5 out of 10, rising to 9 out of 10 after the training.

Level three (behaviour)
76% of respondents have put their knowledge and skills into practice and not just within the cabin crew environment – in schools, shops, at bus stops, even at home. And the quality of care has increased, confirmed by continual monitoring of how medical incidents are handled, and with numerous checkpoints to confirm BA’s high standards have been maintained.

Level four (results)
The results from the revised programme are staggering:
Cost savings: BA’s cost calculator showed they would reap £1.1m in cost savings over the three years compared to delivering the previous classroom-only programme.
Speed to competence: The e-learning ensures all new recruits obtain their licence to fly a day quicker than previously, so BA is gaining 12,000 days productivity over three years from the initiative – 12,000 days when crew are in the air rather than being trained. And, of course, the 2,000 volunteers were also trained a day quicker than they would have been previously.

These proven results were a major factor in it winning a number of awards, including gold in the Excellence in the production of learning content – private sector category at the 2010 e-learning awards.
University of Birmingham

UoB’s Access to Birmingham (A2B) programme is specifically geared to attracting students from under-represented backgrounds to apply to university. Attracting these students is the first challenge – supporting their transition comes next. This isn’t easy, since they generally experience a more profound culture shock than their peers, as their family and friends have little or no experience of higher education.

As a response, UoB commissioned LEO to work with them to develop ‘skills4uni’, a highly interactive online programme, which gives prospective students an idea of what to expect from university life. It includes tools to help students identify what kind of learner they are, advice on critical and objective thinking, how to use the library, and practical tips on taking notes/asking questions in lectures.

It was crucial to fully analyse the programme’s effectiveness, as it would offer important lessons for the university sector as a whole. LEO worked in partnership with UoB to evaluate the learning using the Kirkpatrick model:

<table>
<thead>
<tr>
<th>Level one (reaction)</th>
<th>Level two (learning)</th>
<th>Level three (increased applications up to UoB)</th>
<th>Level four (improve preparedness for HE Study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>An online survey asked users to rate how useful they found each module in relation to the programme’s six key learning objectives. User reactions were highly positive, particularly in giving users the skills to plan ahead.</td>
<td>The post-course assessment results involved students answering randomised questions from a large database. Despite a challenging pass mark, only three % of A2B students failed the assessment. That was 2008. A year later, after students had completed their first year, further evaluation was conducted. So what impact did the course have?</td>
<td>Applications to UoB overall were up 10.5 % in 2009 and up 14 % from A2B students, compared to a national increase of just 8 %, evidence that skills4uni had a highly positive impact.</td>
<td>A year after starting, skills4uni students and their parents were sent online surveys relating to how they/their children coped and performed in their first year. The results were then compared with those from identical questionnaires sent to a control group of non-users and their parents. The results were highly impressive, e.g: 77 % of students who completed skills4Uni felt confident learning independently, compared to just 52 % of those who didn’t complete the e-learning. 60 % of users felt their first year met their expectations of university life compared to only 40 % in the control group.</td>
</tr>
</tbody>
</table>

“It gave an invaluable insight into what uni life was like, and what was expected, in a virtual manner, even before we started uni, so I felt I kind of had the upper hand compared with students who hadn’t done anything as such!”

A2B students about skills4Uni
Level five (reduce dropout rates)

The overall aim was to ensure that students were better prepared for higher level academic study, thus minimising the number of students dropping out of university before completing their studies. This has major cost, resource and skills shortage implications for the UK economy. Having measured the dropout rates over two years, UoB is delighted that the number of students failing to complete the A2B scheme in preparation for university has halved - falling from 12% to just six%. Once at the University, retention for A2B students from first to second year is 96%. The e-learning programme is making a major contribution to this performance.

The robust evaluation and clear business benefits enabled LEO and UoB to win a 2010 National Training Award for ‘Partnership and Collaboration’.

Summary

Most organisations are aware that evaluation is important, but few act on this knowledge and carry out any evaluation. When they do evaluate, the evaluation often follows a badly applied version of the Kirkpatrick model, focusing on the first and second levels and ignoring the third and fourth.

In this white paper we’ve argued that effort should be focused on evaluating the higher levels: behaviour and results. This kind of evaluation should soon be made easier and cheaper by technologies such as XAPI API, particularly when we define our success criteria (aligned to business objectives) at the start of the project.

Evaluation of informal learning requires a different approach; finding ways to track informal learning and measure the use of informal learning resources.

The good news is that investment in evaluation is likely to pay dividends in validating and informing your learning strategy, and there are new tools and methodologies which can take the strain as well as experts who can advice you on the best approach.