Evaluation Practice Toolkit

Evaluation practice toolkit developed by King’s College London Clinical Education and Patient Safety Research Group in partnership with Health Education England

Dr Thomas Simpson MA(Cantab) MBBS MRCP FHEA thomas.simpson@kcl.ac.uk
Ms Siân Kitchen BA(Hons) MA ACIS MAPM
Dr Mary Lavelle PhD CPsychol

Dr Janet E Anderson BBehSc (Hons), MPsych (La Trobe), PhD (Qld.), PGCertAP (Sur.)
Reader in Healthcare Improvement
Dr Gabriel Reedy CPsychol CSci AFBPsS FAcadMed SFHEA gabriel.reedy@kcl.ac.uk
## Contents

<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreword</td>
<td>01</td>
</tr>
<tr>
<td>Authors’ introduction</td>
<td>02</td>
</tr>
<tr>
<td>1. Defining your project:</td>
<td>06</td>
</tr>
<tr>
<td>Start here if you haven’t yet defined your project…</td>
<td></td>
</tr>
<tr>
<td>• Learning aims and objectives</td>
<td>08</td>
</tr>
<tr>
<td>• Aims and objectives activity</td>
<td>10</td>
</tr>
<tr>
<td>• SMART model</td>
<td>11</td>
</tr>
<tr>
<td>• The cognitive domain</td>
<td>12</td>
</tr>
<tr>
<td>• The affective domain</td>
<td>13</td>
</tr>
<tr>
<td>• The psychomotor domain</td>
<td>14</td>
</tr>
<tr>
<td>2. Using the toolkit:</td>
<td>15</td>
</tr>
<tr>
<td>Start here if this is the first time you have used the toolkit…</td>
<td></td>
</tr>
<tr>
<td>• Three boxes evaluation model</td>
<td>17</td>
</tr>
<tr>
<td>3. Evaluation practice decision guide:</td>
<td>19</td>
</tr>
<tr>
<td>Start here to select the right type of evaluation for your project…</td>
<td></td>
</tr>
<tr>
<td>• Introduction</td>
<td>20</td>
</tr>
<tr>
<td>• Black box</td>
<td>22</td>
</tr>
<tr>
<td>• Grey box</td>
<td>24</td>
</tr>
<tr>
<td>• Clear box</td>
<td>26</td>
</tr>
<tr>
<td>• Capacity (primary)</td>
<td>28</td>
</tr>
<tr>
<td>• Theory (secondary)</td>
<td>30</td>
</tr>
<tr>
<td>• Purpose (secondary)</td>
<td>32</td>
</tr>
<tr>
<td>• Project Features (tertiary)</td>
<td>33</td>
</tr>
<tr>
<td>• Evaluation strategy</td>
<td>36</td>
</tr>
<tr>
<td>4. Your proposed evaluation strategy:</td>
<td>37</td>
</tr>
<tr>
<td>Start your own evaluation strategy here…</td>
<td></td>
</tr>
<tr>
<td>• Collecting the right data</td>
<td>38</td>
</tr>
<tr>
<td>• Questions</td>
<td>40</td>
</tr>
<tr>
<td>• Does ‘it’ work?</td>
<td>41</td>
</tr>
<tr>
<td>• Data collection</td>
<td>44</td>
</tr>
<tr>
<td>• Real world impacts</td>
<td>45</td>
</tr>
<tr>
<td>Evaluation models: further reading</td>
<td>46</td>
</tr>
<tr>
<td>• Stufflebeam</td>
<td>47</td>
</tr>
<tr>
<td>• Kirkpatrick’s Model</td>
<td>49</td>
</tr>
<tr>
<td>Examples</td>
<td>50</td>
</tr>
<tr>
<td>• Specialty Trainees in General Practice (STiP)</td>
<td>51</td>
</tr>
<tr>
<td>• Training of teaching assistants in child occupational therapy</td>
<td>52</td>
</tr>
<tr>
<td>• Youth Empowerment Skills</td>
<td>54</td>
</tr>
<tr>
<td>• Tier 1 dementia training</td>
<td>55</td>
</tr>
<tr>
<td>• Incident reporting frequency</td>
<td>56</td>
</tr>
<tr>
<td>• Adult Advanced Life Support</td>
<td>57</td>
</tr>
<tr>
<td>References</td>
<td>58</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>60</td>
</tr>
<tr>
<td>Glossary</td>
<td>62</td>
</tr>
<tr>
<td>Activity 1: Aims and objectives</td>
<td>63</td>
</tr>
<tr>
<td>Activity 2: Evaluation strategy</td>
<td>71</td>
</tr>
</tbody>
</table>
Foreword

Since 2013, Health Education England’s south London local team has been proud to support a range of innovative and exciting training and education projects that support the local delivery of HEE’s mandate and strategic priorities for our workforce, as well as enabling development of new learning initiatives.

Understanding the effectiveness and impact of training initiatives is a critical aspect in capturing and sharing best practice and championing the widest possible spread and adoption of effective training and development projects.

As part of our on-going engagement with training and education in South London, we are delighted to have commissioned a programme of work that promotes continual awareness of and improvement in the evaluation of training programmes.

The evaluation toolkit has been commissioned in response to a request from our stakeholders to provide some dedicated assistance to increase capacity and capability to undertake effective evaluation of their projects and to promote shared learning. It is part of an on-going programme to improve the quality of evaluation and reporting on training and development.

I would like to thank all those who have provided input into the development of the toolkit, especially Dr Thomas Simpson, Dr Mary Lavelle, Dr Janet Anderson and Dr Gabriel Reedy at King’s College London and Siân Kitchen, Senior Programme Manager for HEE’s south London team. Their enthusiasm and dedication to this project, along with their focus on co-design with stakeholders, has resulted in development of an exciting and valuable resource that I hope will continue to contribute to the impact and effectiveness of the fantastic work our stakeholders and partners undertake.

Aurea Jones, Local Director, South London
Health Education England

It is through education and sharing knowledge, research evidence and best practice, that we inspire and inform innovation and improvements in patient care and safety, experience and outcomes.

Professor Ian Cumming OBE
Chief Executive, Health Education England
Knowledge for healthcare: development framework 2015-2020
Authors’ introduction

The education of healthcare professionals is a high priority for an effective and safe healthcare system.

Ensuring that the education and training is achieving its aims and objectives, whilst remaining both relevant and cost-effective, should be the goal of all involved in healthcare education and training in the NHS.

The only way to know whether educational projects are achieving these goals is through appropriate evaluation.

As well as informing the wider world about the outcomes of an educational project, effective evaluation should also inform the ongoing, cyclical improvement and refinement of these interventions.

The goal of evaluation should be not just to prove, but to improve.

In order to bring this to fruition, we are building a toolkit of resources to support evaluation practice. The toolkit will comprise 3 major components:

- **Toolkit 1**: Introduction to evaluation models and strategies (*this publication*)
- **Toolkit 2**: Evaluation methods and guidance
- **Toolkit 3**: Evaluation reporting and dissemination

This toolkit is designed to support those in the healthcare education community with less experience of evaluation practices.

It will provide support and guidance in establishing a pragmatic evaluation strategy, but will not make those decisions for you. We hope that it may provide new perspectives for those with experience of performing evaluation and although this toolkit does not provide guidance about specific methods of data collection and analysis, significant support in this area will be provided as the project develops.

We hope that you use his toolkit, that you enjoy it and that it enables you to design more effective evaluations.

Training and education is a passion for many in the healthcare community. Accurately reporting our work, both triumphs and disasters, should similarly be a moment for enthusiasm and pride.

Effective evaluation, pragmatically designed and delivered, enables us as healthcare professionals and educators to present our work with confidence and allows the communities within which we operate to learn from what we have done.
Authors’ introduction (continued)

Evaluation is often defined as:
The systematic acquisition and assessment of information to provide useful feedback about some:
• Course/project
• Intervention
• Action

Effective evaluation of education in healthcare is increasingly important in an era of rising pressure on clinical services and tighter funding constraints for the NHS. New technologies and innovations such as e-learning are proposed to replace traditional methods of education.

Whilst intuitive benefits such as easier access and efficiency of delivery may be assumed with new technologies versus traditional methods, this may not actually lead to a greater impact as an educational intervention due to unexpected mechanisms. These include short-cut behaviours such as scanning through an e-learning module and completing the assessment until passing it, without appropriately engaging with the content.

The choice of one over another may therefore be based on multiple factors. This means that evaluation needs to be performed to an appropriate standard – allowing us to understand which interventions have been successful or where we encounter challenges. We need effective evaluations to learn what does and does not work, ensuring that frontline staff are able to use what training time they have for maximum benefit. This also means funding bodies will achieve the greatest impact for the support they provide.

We also exist in a time of great scrutiny of the NHS, with emphasis on the delivery of effective and compassionate care by appropriately skilled staff. Healthcare policy and funding decisions over the last two years, along with an increased demand for NHS services have had a significant impact on the delivery of training to staff. For example, the GMC report on ‘The State of Medical Education in Practice’ highlights the challenges on doctors in training, emphasising that when services are under pressure, time and resources for education may be the first to be sacrificed.

In this environment, robust evaluation is vital to inform choices and strategies at all levels. Much like the rise of evidence-based clinical care, the delivery of healthcare education generally is moving towards an evidence-informed model, with the Best Evidence Medical Education Collaboration leading the way in medical and health professions. However, in the evaluation of education in healthcare there remains significant issues with quality. The Health Education England (HEE) report ‘Improving Patient Safety Through Education and Training’ highlighted that:

‘Widely differing approaches to evaluation and often unreliable data or methods are used. Evaluations rarely include comparative analysis and do not robustly assess impact on patient outcomes. A review of the grey literature reinforced our finding that whilst training and education interventions can improve skills and knowledge, there is no conclusive evidence to show which types improve health outcomes or safety. It also underscored our discovery that little has been researched on whether one type of training or education is better than another. We need robust evaluation and measurement, using proven methodologies, so the focus can be on effective education and training.’
Authors’ introduction (continued)

As both teachers and learners we often experience evaluations of educational interventions which are limited in scope and method, represented most commonly by the widespread use of simplistic post-intervention satisfaction scores, also known as ‘happy sheets’. A measure of satisfaction and engagement is an important component of any evaluation, particularly in situations where engagement has been problematic before, however, this is not sufficient to gain a deeper understanding of the outcomes and impacts of a project.

One of the main challenges is that educational evaluation is very much a social science, with its own methods of collecting, analysing and reporting data, which can be more appropriate and powerful in developing our understanding of what we are doing when we teach. However, having been trained as clinical scientists, many educators may try to approach the evaluation of education as if it were a clinical science. This can lead to the overuse of methods such as randomised controlled trials, attempting to demonstrate objective quantitative outcomes, when this may not be the appropriate approach.

This is known as a positivist approach and indeed positivist methods of data collection and analysis are highly prevalent in evaluation of medical education. Positivist approaches can provide powerful evidence of educational benefit when correctly applied, but many fail to show significant results, which limits our understanding of education. Other approaches often have merit and are worth considering for many types and scales of educational intervention. These include methodologies associated with the social sciences, such as qualitative research, where objectivity is not assumed and instead a subjective viewpoint is welcomed and incorporated within the methods used.

With this toolkit we hope to make the case that:
- effective evaluation is within the scope of educators in healthcare;
- a wide variety of methods are available to enable effective evaluation to take place;
- effective evaluation can inform a deeper understanding of the impact of educational interventions leading to a focus on achieving meaningful aims and objectives.

What is evaluation for?
Effective evaluation can provide insight into many aspects of a project. In demonstrating the impact of an intervention, the reporting of meaningful outcomes and degree to which learning objectives have been met is clearly a primary goal. However, if done well, evaluation can also provide valuable insights into how a programme has achieved its goals and importantly, highlight areas where goals were not achieved. This allows a deeper inspection of the impact of a project.

Effective evaluation, as well as telling us ‘what works,’ can also inform us how and why a programme works or does not work for different people. It can be a powerful tool to feedback into the ongoing development of a programme, supporting a progressive cycle of improvement and further evaluation. A well-developed evaluation strategy also supports presentation and publication of work. This can greatly assist the spread and adoption of best practice while preventing uptake of interventions that do not work, a vital contribution from effective evaluation which allows better use of resources.
From the point of view of funding organisations or regulatory bodies, evaluation provides a means to understand impact across a portfolio of projects as well as informing the development of higher level strategies of workforce education and training.

Educators are frequently enthusiastic about the work that they do, often in their own time and from their own initiative. They should rightly be proud of what they do and effective evaluation is a vital part of this, confirming successes and providing enlightenment where things have not gone to plan.

Problems and pitfalls with evaluation
A principle issue with evaluation is that many educators lack knowledge, experience and confidence in performing evaluations. In the same way that excellence in clinical medicine does not guarantee excellence in clinical research, excellence in teaching does not guarantee excellence in evaluation. Evaluations can fail for a number of reasons. Educators may not know what to evaluate and how to perform evaluations using methods appropriate to generating significant insights.

Effective evaluation strategies require consideration and planning from the very start of programme development allowing appropriate data to be collected from the outset. Also, the creation of appropriate learning objectives for a project is a vital step in enabling rational decisions to be made about the evaluation strategy. However learning objectives may be missing or poorly constructed and project managers and evaluators are also busy managing the daily pressures of work in the NHS alongside their teaching responsibilities, reducing their capacity to design an appropriate evaluation.

Bias
Another potential pitfall is bias. This is usually unconscious rather than deliberate, but those conducting evaluations should be mindful of the potential for bias.

Relevant types of bias include:
- Publication bias, whereby evaluations which produce negative results, or indicate no significant differences from an intervention, are not publicised.
- Outcome reporting bias, whereby the results reported do not fairly reflect the real impact of an intervention. For example, emphasis on outcomes which do not represent the primary part of an evaluation.

Evaluations may also use inappropriate methods, or produce data on outcomes which have little meaningful translation into significant change. There is also danger of evaluation fatigue; learners bored of answering the same questions, educators swamped with data they don’t need and significant volumes of data left unpublished.

There is no such thing as a perfect evaluation strategy and problems and pitfalls will occur in all projects. Awareness of these and actively addressing them will increase the chances that evaluations will be completed and will be effective. We must remember that without effective evaluation, we cannot know if we have made a significant impact with our interventions and we may struggle to identify what works, how it works and whether we should continue to support it.

It is through education and sharing knowledge, research evidence and best practice, that we inspire and inform innovation and improvements in training with the goal of better patient care, safety, experience and outcomes.
1. Defining your project
In order to evaluate your project, you must first define it.

It is useful to consider the context in which your project sits and which has led to its creation.

Elements of the CIPP model as proposed by Stufflebeam in the section on evaluation models, also provide a useful guide to defining your project.

Questions you need to consider are:

- What are your goals?
- What are your plans to achieve those goals?
- How do you intend to put those plans into action?

It is also helpful to think about the differences between the aims and objectives of your project for you as project lead, versus the aims and objectives you have for the participants in your project as learners.

Funding bodies may also have their own aims and objectives for supporting your project.

Are there significant differences between these aims and objectives? If so, you will need to define and consider them separately when choosing your evaluation strategy. This will help you to explore the degree to which you achieve these separate aims and objectives. Without clarification of these separate and potentially competing objectives, evaluation may be compromised.

With answers to these questions in mind you should now proceed to consider your learning aims and objectives in more detail.
Learning aims and objectives

Writing learning aims and objectives is a key step to developing a suitable evaluation strategy. It enables you to clarify what it is your learners will be able to do differently at the end of your project.

Having defined this, it is easier to consider how to evaluate the success of your project in achieving those objectives.

Most objectives relate to changes in learners and these changes usually form either a change in reasoning or resources.

Aims

These should relate to the longer term and perhaps how the benefits of your project might translate into the wider context within which it is situated.

See examples > 50

Objectives

These should be compatible with the SMART model where possible. Using this model helps to generate measurable learning objectives, which in turn can support a more focussed evaluation. Another useful way of thinking about your learning objectives is through Bloom’s Taxonomy a well-recognised way to categorise different types of learning objectives.

It classifies educational learning objectives into three broad domains:
1. Cognitive > 12
2. Affective > 13
3. Psychomotor > 14

Given the wide variety of projects in healthcare education it is worthwhile exploring these three domains as many projects involve change in more than one domain. Historically, many healthcare education projects have been focussed on changes in knowledge amongst learners and so we will focus initially on the cognitive domain.

The cognitive domains provide useful language with which to create learning objectives which are feasible and measurable and therefore more likely to be consistent with the SMART model.
If you can create at least three learning objectives covering different levels of complexity, then you will be able to start thinking of ways of gathering some useful data for your evaluation. As an example you could state:

- The knowledge you expect your learners to learn.
- The skill that that knowledge will enable them to perform.
- The ability to apply objectives 1 and 2 in a specific context.

The affective domain covers changes which learners may experience in their attitudes, feelings and emotions as a result of a project. The psychomotor domain similarly has levels available to guide consideration of appropriate learning objectives, relevant to practical skill learning, which we have not chosen to expand upon here, but which can be easily found.

### Instructional design

Another less prescriptive way of considering your learning objectives is based on the theory of instructional design. This is an approach to creating efficient, effective and appealing ‘instructional’ experiences, through a process of detailed analysis of learning needs and the desired outcome for learners; design of an intervention to assist in the transition of learners; development of that intervention followed by implementation and evaluation stages. Simply put, answering the following three questions should give you a better grasp of the outcomes of your project which are amenable to evaluation:

1. What do you want the learners to be able to do?
2. In what conditions do you want your learners to perform this action?
3. To what standard do you wish them to perform this action?

While Bloom’s taxonomy and instructional design may be very helpful in creating SMART objectives, not all learning will be so easily captured by learning objectives developed in this way. Increasingly healthcare education is multi-professional and significantly aims to break down barriers and silos, to improve multi-professional working and create collaboration.

Aims such as these may not translate into neatly-defined SMART learning objectives and may require you to consider less quantitative and measurement-based approaches from the very outset of your project.

See examples > 50
Aims and objectives activity
Now try to write the all the relevant learning objectives for your project for yourself, your learners and your funding bodies or supporters if different.

Go to the form  

Further reading
Learning aims and objectives  >  08
CIPP model  >  47
Evaluation models  >  20
Examples  >  50

Further reference
SMART model  >  11
The cognitive domain  >  12
The affective domain  >  13
The psychomotor domain  >  14

Activity sheets
Aims and objectives activity  >  63
SMART model of educational objectives

S  Specific  says exactly what the learner will be able to do.
M  Measurable  can be observed by the end of the training session or within an otherwise definable timeframe.
A  Attainable  for the learners within scheduled time and specified conditions.
R  Relevant  to the needs of the learner and the organisation.
T  Time-framed  achievable by the end of the training session or otherwise definable timeframe.
1. Defining your project  the cognitive domain

Cognitive
The cognitive sub-categories are, in increasing order of complexity:
1. Remembering
2. Understanding
3. Applying
4. Analysing
5. Synthesising
6. Evaluating

Using language which matches these six steps of cognitive performance can be a useful structure for building your learning objectives.

Examples of appropriate words corresponding to these levels of Bloom’s cognitive domain include:
1. Knowledge/remembering: define, list, recognise.
2. Comprehension/understanding: characterise, describe, explain, identify, locate, recognise, sort.
5. Synthesis/creating: construct, design, formulate, organise, synthesise.

Examples of words to avoid
Try to avoid the following words in your learning objectives, as they are difficult to define and measure. They may be more suited to the language of your project aims.
• Understand
• Appreciate
• Know about
• Become familiar with
• Learn about
• Become aware of

Aims and objectives activity > 63
Affective
The affective sub-categories are, in increasing order of complexity:
1. Receiving: passively paying attention to a stimulus.
2. Responding: actively responding.
3. Valuing: assigning values to knowledge which is acquired.
4. Organising: organising new knowledge into a mental picture or schema.
5. Characterising: the creation of abstract knowledge.

Using language which correlates to behaviours which represent these five steps of affective performance can be a useful structure for building your learning objectives:

1. Receiving: acknowledge, asks, follows, listens, understanding.
2. Responding: answers, assists, aids, discusses, helps, labels, performs, presents, tells.
3. Valuing: appreciates, demonstrates, initiates, invites, joins, justifies, proposes, respect, shares.
5. Characterising: acts, discriminates, displays, influences, modifies, performs, qualifies, questions, revises, serves, solves, verifies.

As with the cognitive domain it is better to craft learning objectives which describe specific behaviours you want to see in your learners, rather than overall changes in their affective approach to a subject.
1. Defining your project: the psychomotor domain

Psychomotor
The psychomotor sub-categories are, in increasing order of complexity:

1. Perception: the ability to use sensory cues to guide motor activity.
2. Set: A learner’s readiness to act, from a mental, physical and emotional disposition.
4. Mechanism: the basic confidence and proficiency to perform a complex skill.
5. Complex overt skill: the skilful performance of complex motor acts, often smoothly, automatically and without hesitation.
6. Adaptation: skills are well-developed and individual movement patterns can be modified to fit special requirements.
7. Origination: creating new movement patterns to fit a particular situation or specific problem.

Using language which correlates to behaviours which represent these seven steps of psycho-motor performance can be a useful structure for building your learning objectives:

1. Perception: describes, detects, distinguishes, identifies, isolates, notices, recognises, relates, selects, separates, touches.
2. Set: arranges, displays, explains, moves, prepares, proceeds, reacts, shows, states, volunteers, responds, starts.
3. Guided response: constructs, copies, displays, dissects, fixes, follows, imitates, manipulates, reproduces, responds, tries.
4. Mechanism: completes, makes, organises, performs. Many of the words used for level three can also apply here, albeit it a higher standard may be expected.
5. Complex overt response: assembles, builds, calibres, constructs, coordinates, demonstrates, dismantles, displays, dissects, fastens, fixes, manipulates, measures, mends, mixes, organises, sketches.
6. Adaptability: adapts, adjusts, alters, changes, integrates, rearranges, reorganises, revises, solves, varies.
7. Origination: arranges, builds, combines, composes, constructs, creates, designs, formulates, initiates, makes, modifies, originates, redesigns, trouble-shoots.

NB: The key words are very similar across levels 3, 4 and 5 but will have adverbs or adjectives that indicate that the performance is quicker, better, more accurate.
2.

Using the toolkit
This toolkit will help you go through some of the decisions you need to make about the evaluation of your project.

It is not meant to be prescriptive and your final strategy will have to be something you have the capacity to complete.

The toolkit may identify that your current evaluation approach is appropriate, hopefully offering you further confidence in that approach.

However, where there are areas where you may need support, we anticipate that early identification will enable you to seek appropriate support for your evaluation, maximising the chance of successfully completing work of high quality.

In order to achieve this, we have provided guidance on the writing of learning objectives which will underpin many of the other decisions you will make.

Working through the stages of this toolkit will help you towards creating an overall evaluation plan, as well as stimulating you to think of elements of your project beyond simple outcome reporting.

There will be guidance on considering appropriate evaluation questions, before an introduction to some of the evaluation techniques you may wish to explore and employ.

At the end of the toolkit we aim for you to have a plan for how you will approach your evaluation, with ideas about the types of methods you will use and a clear rationale for these decisions.

This will be your evaluation strategy.
What is evaluation practice?
The systematic acquisition and assessment of information to provide useful feedback about some:
• Course/project
• Intervention
• Action

There are many models of evaluation. The model most commonly quoted in healthcare education literature is that proposed by Kirkpatrick.

This model also forms the basis of the Best Evidence Medical Education Collaboration approach to assessing research quality.

This can be a useful way to understand the different levels of outcome which can be measured in an evaluation, it does not necessarily help to understand the wider potential scope of an evaluation. Nor does it provide guidance on appropriate methods.

In this toolkit we use three models:
1. Scriven’s three boxes evaluation model
2. Stufflebeam’s CIPP model
3. Kirkpatrick’s model of outcomes

The three boxes evaluation model
One model to understand evaluation at a deeper level is represented by three boxes and was first developed by Scriven.

This model of evaluation provides a visual representation of different approaches to evaluation.

These can be useful to understand the ways in which different considerations can impact your evaluation decisions.

They can also be used to select evaluation strategies and as such provide a helpful way to link the educational project, the educator/evaluator and the evaluation methods in a way which should enable more appropriate decisions to be made about the purpose and performance of evaluation.

There are three ‘boxes’ within the model, representing different approaches and these will be discussed in turn.

They are not presented as a hierarchy with one approach being superior to the others as they each have vital roles to play in different evaluation approaches.
The three boxes evaluation model: overview
Evaluation strategies can be thought of as adopting three main approaches.

Scriven and later Kazi, have adapted a model from software analytics to make these three approaches easier to visualise and understand.

The three boxes evaluation model describes three ‘boxes’ of evaluation, black, grey and clear.

The colour of the box indicates the degree to which the internal processes of the programme under analysis are examined by the analytic process.

It also indicates whether the focus of the analysis or evaluation is primarily on the effect of the programme or on the processes by which that programme achieves its outcomes.

We will expand on this model as it has been adapted for educational evaluation throughout this toolkit.
3. Evaluation practice decision guide
The decision guide
From outlining your programme and defining your learning aims and objectives you should hopefully have a clearer idea of what it is you hope to do and how you hope to create change in your participants.

Working through the decision guide with these considerations in mind you will find some guidance on developing your evaluation strategy.

For each of these questions you will see that there is an associated graded scale, a visual representation of the spectrum of evaluation approaches according to the three boxes evaluation model.

This is not meant to be prescriptive but simply to afford you the opportunity to consider your evaluation from many different angles, to see where these angles may produce complementary or conflicting outcomes, and how you might work through these to come to a final decision about your strategy. Considerations focus on four key areas:

Capacity
The primary consideration in any evaluation strategy is what capacity you have to perform evaluation.

Theory
Your approach to theory will significantly inform the evaluation of your intervention.

Purpose
The next important consideration in the design of your evaluation strategy is to consider the purpose of the evaluation and who you are doing it for. Stufflebeam’s CIPP model of evaluation may help in analysing purpose.

Features
Having considered your capacity to perform an evaluation, and who you are doing it for, it is now reasonable to consider the features of your evaluation and the most appropriate strategy to adopt in light of these.

We have set the considerations out in this order to direct your attention to the real-world decisions which impact most on evaluation, before you commit to an ambitious approach which you may have difficulty completing.
3. Evaluation practice decision guide

**Black box**
Can measure inputs and outputs but internal processes are not evaluated. *Expanded definition > 22*

**Grey box**
Measures inputs and outputs but also attempts to discern how the actual processes of the project are perceived. *Expanded definition > 24*

**Clear box**
Measures inputs and outputs but also identifies contextual factors which impact upon learners and projects. Identifies and explores unexpected outcomes and both positive and inhibitory mechanisms which impact on learning. The priority is less on the inputs and outputs and more on understanding how the inputs affect the mechanisms by which learning occurs and how those mechanism lead to different outcomes. *Expanded definition > 26*

---

**Further reading**
- The three boxes evaluation model > 20

**Further reference**
- Capacity (primary consideration) > 28
- Theory (secondary consideration) > 30
- Purpose (secondary consideration) > 32
- Features (tertiary consideration) > 33

**Activity sheets**
- Evaluation strategy activity > 71
Black box evaluation
Does project X work as intended?

In this form of evaluation, we are not interested in the inner workings of a project or the mechanisms by which it achieves its outcomes.

Data will be collected on inputs and outputs only, eg pre-course and post-course knowledge and confidence scores.

Black box evaluation can be performed without any knowledge or understanding of the internal structures and processes involved in a project.

We are also not exploring the ways in which the project interacts with its context, as such, the project is context-independent.

Also by focussing on inputs and outputs only, we are able to evaluate an educational project which takes place on a large scale.

An example would be the evaluation of learning from an e-learning module created for national use.

The content of the module itself is fixed and applied in so many different contexts that it is not of value to attempt to evaluate the impact of context on the learning.

Another way of looking at this is that the project is transferable in its entirety and can be expected to function in essentially the same way wherever it is applied.

Black box evaluation techniques sit within the positivist paradigm and are quantitative in nature.

They include surveys, scales, questionnaires and assessments of knowledge. Any approach involving quasi-experimental methods or randomised-controlled trials will also rely on the principles of black box evaluation.

Results are generated that can answer questions such as: Does project X work as intended? The results may also be appropriate for statistical analysis, depending on sample size.

Black box evaluations have to be designed based on ‘theories’ about how a project will achieve its results. However, on their own they are unable to give deeper insights into how or why a project produces its effects. It is also very unlikely to identify unexpected outcomes (positive or negative), as tools must be designed in advance of any project, thus the questions are fixed and the range of possible answers limited.

Further reading
The three boxes evaluation model  
20

Further reference
Capacity (primary consideration)  
28
Theory (secondary consideration)  
30
Purpose (secondary consideration)  
32
Features (tertiary consideration)  
33

Evaluation strategy activity  
71
Black box evaluation summary

- measures inputs and outputs only;
- uses quantitative methods;
- does not take context into account;
- does not address the internal components of a project, but assess the impact of the project as a whole;
- answers the question: does the project work as intended?

Examples > 50
Grey box evaluation
Does project X work and how do people think it works?

Whilst black box evaluation provides valuable information about educational impacts, particularly on a large scale, it provides limited information on how or why a project has achieved the impacts it has. Without exploring the internal mechanisms of a project this level of insight is not possible.

Grey box evaluation combines the usual processes of black box evaluation with some attempt to ascertain the ways in which outcomes have been achieved, through combining black box approaches with qualitative approaches to explore perspectives on how a project works.

As such it may combine the components of a black box evaluation with methods such as free-text questionnaire items, interviews and focus groups.

Much routine evaluation which takes place in healthcare education currently sits within the grey box, combining some quantitative and qualitative elements to gain insight into the impact of an intervention, but also some deeper feedback on how the intervention has been perceived.

This is usually seen as a mix of tick box and free-text responses on evaluation sheets.

Whilst this may often be the correct approach, it may also have arisen organically, or been chosen due to expediency or prior experience without careful consideration of the benefits and challenges of this approach.

In particular, whilst a combination of quantitative and qualitative methods (mixed-methods) can support meaningful evaluation, the ways in which mixed-methods evaluation is carried out and reported may not lead to effective data collection and analysis or appropriate conclusions.
Grey box evaluation summary

- combines measurement of inputs and outputs with exploration of how the project has achieved these outputs;
- uses quantitative and qualitative methods;
- does not significantly take context into account;
- explores the perception of internal mechanisms of a project;
- answers the question: does the project work and how do people think it works?

Examples 50
Clear box evaluation

Clear box evaluation is focussed on the internal mechanisms and processes of the project itself and explores the ways the project and its context interact.

The evaluation uses mainly qualitative data from learners, educators and commissioners to gain insight into these elements in order to gain a deeper understanding of:
- how outcomes are generated
- the role of context in the outcomes generated
- the mechanisms by which the project generates those outcomes.

Outcomes form only a part of the evaluation, with the majority of attention on the processes and workings of the project. This can be very powerful in evaluating smaller projects, especially those which are complex in nature and in which the context plays an important role.

This approach is open to identifying unforeseen consequences and outcomes which can be very beneficial in a complex context, where projects may have conflicting and interfering patterns of outcome.

By carefully exploring all aspects of the project, as opposed to only inputs and outputs, clear box evaluations can enable you to find new theories about why and how your project is achieving outcomes – both positive and negative. Clear box techniques generally sit within a realist paradigm, exploring individual perceptions and behaviours in the context of an independent reality which can be approximated but not directly observed.

Positivism states that there is an objective truth and that we can find it and prove it through experimentation, such as randomised controlled trials. Interpretivism states that there is no objective truth, only subjective experiences, ie how the world is perceived by individuals and that we can explore these to understand the world, however this does not equate to a truth which is objective and independent of individuals’ perceptions.

Realism asserts that both an objective truth and subjective experience are real, but that we cannot directly observe the objective truth. We can explore both using various methods, which may allow us to suggest theories about the objective world and the ways in which it is perceived.
We recognise that this discussion may be conceptually complex for healthcare professionals and educators embarking on evaluation, perhaps for the first time. However, simply being aware of these approaches and their associated techniques can be the first step to developing new skills as an evaluator.

Clear box evaluations are increasingly widely used and may be particularly relevant as healthcare delivery and education continue to move from clinical settings into community settings, where contexts will have significant impacts on outcomes.

- seeks a deeper understanding of the internal processes and mechanisms of a project, the ways in which context affects those processes and the whole range of outcomes, positive and negative, which may result;
- uses predominantly qualitative methods;
- is much less focussed on inputs and outputs as a measure of impact;
- provides detailed evaluation of a project to inform programme development;
- answers the question: what works, for whom, in what respects, to what extent, in what contexts and how?

Examples > 50
3. Evaluation practice decision guide capacity

Capacity
What capacity do you have to perform evaluation?

What do we mean by capacity?
• Capability/experience in performing evaluation
• Time available for educators to perform evaluation
• Project timeline/deadlines
• Money
• Staff
• Academic support

Very limited capacity: black box

Extensive capacity: clear box

Further reading
The three boxes evaluation model > 20

Further reference
Capacity (primary consideration) > 28
Theory (secondary consideration) > 30
Purpose (secondary consideration) > 32
Features (tertiary consideration) > 33

Activity sheets
Evaluation strategy activity > 71
Capacity can include a number of different resources, some personal, such as your available time.

Your experience in performing evaluation and capability to use different methodologies and tools are also important considerations.

Capacity can also refer to your access to funds, academic or logistical support, or external evaluators.

A mismatch between the appropriate strategy and your capacity to perform it, will almost certainly derail any evaluation strategy and affect your ability to complete, report on and disseminate your evaluation.

Further questions will enable you to consider your appropriate evaluation strategy. If this highlights a mismatch between what is suggested and your capacity, then it is advisable that you raise this issue early.

It is good practice to include an allowance for evaluation within any formal project proposal, either setting aside funding or time for yourself, or seeking support from others. This will help you to clarify your available capacity.

This part of the guide is designed to highlight your potential needs in this area.

If you have limited capacity, then a black box approach may be easier and more pragmatic. If you have available capacity to conduct a more complex evaluation then a grey or clear box approach will likely be far more informative.
3. Evaluation practice decision guide

Theory
What is your approach to theory in your evaluation?

What do we mean by theory?
• Any model or belief that informs your educational project
• Significant models or beliefs which your learners might hold about your project
• New ideas and models which may emerge as a result of your project

Proof of your pre-existing theory: black box

Exploration of new and emergent theories: clear box

Further reading
The three boxes evaluation model > 20

Further reference
Capacity (primary consideration) > 28
Theory (secondary consideration) > 30
Purpose (secondary consideration) > 32
Features (tertiary consideration) > 33

Activity sheets
Evaluation strategy activity > 71
Theory
If you have a particular theory about the way(s) in which your project will create change in the learners, or impact on the organisations involved, then you may wish to simply see if those changes or impacts have occurred in the ways in which you intended or believed that they would.

As such you will design your evaluation to do this along a black box model measuring whether the changes you would expect to see, based on your theory, have happened.

However, if you wish to explore how your learners view your project and perhaps discover some of their theories about how your project may or may not have led to changes for them or their organisation, then you will need to take a more grey box approach.

Finally, if you wish to discover new theories and perhaps explore unintended consequences of your project then it will likely require you to adopt a more clear box approach.
Purpose
The purpose of your evaluation has a huge influence on the evaluation strategy you use.

Different agendas can be satisfied through different means and acknowledging this early can ensure that complementary or competing interests can be satisfied by the evaluation.

For example, reports for some funding bodies may potentially be satisfied by a more focussed approach, such as a black box evaluation to ensure delivery of agreed objectives for the project have been met. This is particularly true in ongoing, established projects.

An evaluation approach aimed at publication or a detailed exploration of the mechanisms by which a complex project has achieved its outcomes would require a deeper exploration through more involved methodologies.

If the project in question is ongoing there may be benefit in a really detailed exploration, informing an iterative improvement of the project.

This will require an evaluation that includes consideration of the context, planning, process and outcomes of the project, ie all four stages of the CIPP model. This would suggest a clear box approach as any other level of evaluation may not provide sufficient information to inform project change.

A grey box evaluation will provide information on outcomes and perceptions of the project from learners and may perhaps inform project changes in response to learner opinions; however these changes may be limited in their scope.

If you are primarily interested in evaluating outcomes alongside a simple description of your project but without evaluation of context, planning and process, then a black box evaluation should suffice.
3. Evaluation practice decision guide

Features
What are the features of your project that will inform your evaluation choice?

Large scale project: black box (large scale intervention)*
Simple/repeatable/context-independent: black box
Transferable in totality: black box

Small scale project: clear box
Complex/unique/context-dependent: clear box (large scale, but small sample)*
Transferable in component form: clear box

Further reading
The three boxes evaluation model > 20

Further reference
Capacity (primary consideration) > 28
Theory (secondary consideration) > 30
Purpose (secondary consideration) > 32
Features (tertiary consideration) > 33

Activity sheets
Evaluation strategy activity > 71
Features
These tertiary considerations are not the only issues to consider but constitute the main features which should guide your strategy decisions.

It is important to note that the results of these considerations may be different to the results from the primary and secondary considerations.

This is likely when your capacity to perform evaluation doesn’t match the suggested evaluation strategy.

Further, when the evaluation is being performed according to the separate requirements of people or organisations outside of the project, recognising these conflicts in priorities early can enable pragmatic decisions to be made.

This will support an effective evaluation strategy for your project which is achievable and satisfies all stakeholders. The main features to consider at this point are:

Scale of project:
• A large scale project involving lots of learners makes deeper evaluation of contexts and process significantly more challenging and points towards a more black box evaluation. This can produce useful metrics and statistical measures of change in a larger number of learners. While a large scale project evaluated as a whole may require a black box approach use of sampling techniques to select a smaller number of learners for evaluation may allow a more grey box or even clear box approach to evaluation.

• A smaller scale project with fewer learners is less likely to generate change in a way which can be measured quantitatively to produce meaningful results. However, it does allow for a deep exploration of the changes seen through more qualitative approaches.

• A project involving just a few learners allows for a deep exploration of the role of context and individual experience on the outcomes observed. Though unlikely to translate into widely generalisable conclusions, this may significantly inform the development of other projects or the iterative improvement of the specific project itself.
When an educational project is relatively simple, repeatable or fixed in form, such as an e-learning module or a standardised package (e.g., advanced life support training), the context in which that project is delivered may be far less important to include in the evaluation. Evaluation which does not consider context sits at the black box end of the spectrum.

When projects are complex, flexible or responsive to context then an ideal approach to evaluation should include consideration of context. Exploration of the relationship between the context and the project is likely to be required to fully explain the ways in which the project has led to impacts. Therefore, a more clear box approach would be informative.

Much like context, opportunities for complete transferability (i.e., use by other educators in other sites) are increased in projects that are simple, repeatable and fixed in form. If the intention is to transfer the project in its totality from one setting to another, it may be helpful to treat it as a black box for the purposes of evaluation.

When a project is not transferable in totality either due to significant contributions from the context or the complexity of the project, it is instead the component parts of the project that become more interesting to outsiders as elements which they may take away to adapt for their own work. As such, a deeper exploration of the project will be needed. This will benefit from a more grey or clear box approach to support this deeper exploration.
Design your evaluation strategy

At this point, you should have some idea of the most appropriate evaluation strategy for your project.

It is useful here to stop and consider the conclusions you have reached at each stage and the degree to which they are consistent.

Where the outcome of tertiary considerations (features) clash with primary considerations (capacity)... Do you need to seek further capacity? .. or reduce the scope of your evaluation?

There may be other ways to achieve your aims on a limited capacity:
• Existing data collected? Can this be used to inform elements of your evaluation, avoiding duplication?
• Existing tools/instruments? Can you find useful tools that have already been created and validated in order to perform your evaluation?
• Small representative sample in depth rather than total population. Sampling can be a complex process however there are some common methods which can be easy to apply, particularly in qualitative research. There will be more guidance on this in Toolkit 2.

Where the outcome of tertiary considerations clash with secondary considerations (theory and purpose)... Do you need to discuss with funders / adjust your strategy?

How to maximise your chances of finding important unexpected or negative outcomes and what to do with them... Why doesn't it work in the way it was intended?’

Go to the form 71
4. Your proposed evaluation strategy
4. Your proposed evaluation strategy

Evaluation strategy
Having explored the options and validated your strategy, you should hopefully be thinking clearly in terms of one of the three approaches, black, grey or clear box evaluation.

These are the kind of questions you will be asking:
- Black box: Does ‘it’ work?
- Grey box: Does ‘it’ work and why do people think ‘it’ works?
- Clear box: More complex but along the lines of: what works, for whom, in what respects, to what extent, in what contexts and how?

Collecting the right data
One of the main reasons for establishing your evaluation strategy early and ideally at the same time as you develop the programme, is to ensure that you are collecting the right data from the start.

Evaluation reporting is often compromised because data is collected retrospectively and does not provide information that is relevant to the learning objectives.

Choosing an evaluation strategy and deciding at what level of outcome you are expecting to see change can enable the appropriate data collection to be integrated into the programme from the start.

It is also worth considering what information is either already available, or would be routinely collected during the course of the intervention and what use can be made of this.
Collecting the right data (continued)

Data collected in evaluations may duplicate data already gathered, or eventually be unused. Having too much data can be just as unhelpful as having too little. Equally, an extensive data collection process can be fatiguing for participants and project teams in its administration.

A further advantage to using existing forms of data collection is that these may be already established as robust in terms of methodology and method.

Use of a previously validated tool or instrument for data collection provides a degree of academic rigour to the process, as opposed to creating your own, which may be flawed in ways which are not apparent to you.

Where possible, collect only the data that you need to answer the questions you have, use data that exists already and if new data needs to be collected, try to use existing tools and instruments which have been validated rather than creating new ones.

The final advantage to establishing your data collection and analysis methods at the outset of your project, is that you will be reducing the possibility of bias in your work.

Bias may occur when we adopt data collection and analysis techniques that are more likely to give us the answer we want.

Adopting data collection and analysis methods after a project has begun often leads to bias as we have some insight into how things are going and how best to show the impact of the project in a positive light.

By agreeing these elements in advance, this potential bias is reduced and a more balanced evaluation is likely to result. Addressing these issues in the evaluation design with the project leaders also allows an early discussion about the possibility that findings may be unexpected, or even negative and agreement about how this outcome might be best addressed. This can help to reduce anxiety about the evaluation process itself.
Questions
Having a question in mind then helps to suggest methodologies

- **Black box:** ‘Does ‘it’ work?’
  Quantitative methodologies: surveys, scores, statistical analyses.

- **Grey box:** ‘Does ‘it’ work and why do people think ‘it’ works?’
  Mix of quantitative and qualitative methodologies (as black box), including interviews, focus groups, interpretivist approaches.

- **Clear box:** What works, for whom, in what respects, to what extent, in what contexts and how?
  Predominantly qualitative approaches including interviews, focus groups, observational data collection, with pragmatic approach. Specific methodologies include realist and illuminative evaluations.

Further reading

- Questions > 40
- Does ‘it’ work > 41
- Real world impacts > 45

Further reference

- Black box > 22
- Grey box > 24
- Clear box > 26

Activity sheets

- Evaluation strategy activity > 71
What do we mean by does ‘it’ work?

It is important to look at Kirkpatrick’s model of evaluation when we consider what we mean by work.

We need to think about the learning objectives and decide at what level we would expect or hope to see changes.

Level 1: Reaction:
Learner reaction to and engagement with a learning encounter is a commonly recorded part of many evaluations. It is obviously important to ensure that you are creating a positive learning experience for your learners. In some cases, particularly novel projects involving new or notably hard-to-reach groups, engagement is in fact a vitally important outcome to demonstrate and discuss.

However, too much evaluation may provide information only at this level, so-called ‘happy sheets’ and does not significantly inform decisions about effectiveness or invite meaningful comparisons with other teaching. Measuring level 1 can be as simple as recording registers of attendance but can include more detailed satisfaction scores.

Level 2a: modification of attitudes/perceptions:
Level 2b: modification of knowledge/skills:
These are changes within the learners and can also be relatively easy to measure through qualitative and quantitative means.

They can be meaningful changes depending on the objectives of your project. However, a change in attitude, perception, knowledge or skill recorded following training does not necessarily translate into either long-term change, or changes in behaviour in actual practice.

Measuring level 2 outcomes ideally includes actual tests of knowledge or skill, but often relies on learners’ self-reported changes. Asking for specific examples where they can highlight when they used new knowledge or skills can be helpful in strengthening this type of data.
What do we mean by does ‘it’ work? (continued)

Level 3: behavioural change:
Changes in actual behaviour are harder to measure as they require effort to engage with learners at some point following an educational project and finding a way of measuring changes in the behaviour in question. This leads to a further challenge to identify the educational project as the cause of any behavioural changes seen.

Whilst challenging, changes seen here are considered to be high level and significant and it is well worth considering whether and how measurement of change at this level might be approached as part of your evaluation.

Measuring level 3 outcomes usually requires some degree of external observation of the learner for the highest validity, however, self-reported changes in behaviour with specific examples can be valuable in an evaluation.

Work-place based assessments may provide a validated, pre-existing method of assessing changes in behaviour in the actual clinical environment following an educational experience.

Level 4a: change in organisational practice:
Level 4b: benefits to patients/clients:
The most challenging outcomes to demonstrate from a learning project are changes in real world impacts for those with whom your learner interacts, either at individual or an organisational level. The variables involved and the difficulty in demonstrating causality, as seen with level 3 outcomes, are greater here.

Outcomes at this level can be shown from various sources of data, including patient and organisational level data. Increased usage of detailed electronic records of patient outcome and clinical practice may enable level 4 data to be collected in novel ways. However, in a complex and adaptive open system such as healthcare, the challenge will always be demonstrating that changes are due to a specific educational project. Further, a single educational intervention, as many training events are, make these outcomes even more difficult to lay claim to. Imagining a single educational programme or intervention as having this sort of impact equates education with drug therapies.

In clinical research, when you give a drug to a patient, you can usually be confident that you know the mechanism by which the drug will work. You also know the effects that drug should have on the majority of patients, and therefore, provided they take the drug as prescribed, you can make predictions about the impact of that drug on a group of patients, and over time, demonstrate significant changes to their health outcomes (equivalent to a level 4 outcome).
Education, however, is not like a drug. We may not know the mechanisms by which a particular education intervention creates learning or changes practice in learners. These mechanisms may be different in different learners and in different contexts.

Further, educational projects are rarely introduced in isolation. They may form part of a package of measures in response to a particular need or challenge, particularly in relation to quality improvement. These packages as a whole may demonstrate Level 4 outcomes.

Exploring the ways in which the educational component supported other measures, informed the development of components, or helped staff to engage with interventions may demonstrate that, although the educational project itself did not directly lead to a Level 4 outcome, it played an important role.

An example might be a course introduced as part of a trust-wide approach to improve sepsis outcomes. This overall approach may have many different components relating to staffing, resources, IT systems, clinical pathways and reporting.

The overall package may lead to an improvement in Level 4 outcomes for sepsis. This can be demonstrated using qualitative and quantitative methods, which explore how the course helps learners to engage with other components of the package, and the ways in which feedback from the course informs other elements of the package.
Data collection

This part of the toolkit briefly addresses some issues around data collection which may be important to bear in mind before settling on a specific evaluation strategy.

Please also review the section on collecting the right data.

We will deal with specific methodologies and methods of data collection and analysis in Toolkit 2.
As well as considering at what level we would expect or hope to see change, there has to be a realistic approach to what level we can actually demonstrate change.

As such, our evaluation strategy must align with both your capacity to perform evaluation and your purpose in performing the evaluation. Choosing a strategy of data collection and analysis which does not align with these considerations is likely to lead to difficulties.

If you have not already worked through the evaluation strategy form we suggest you do so now to ensure that these different considerations are correctly addressed and aligned in your strategy.
Evaluation models: further reading
**Evaluation models: further reading**

**Stufflebeam’s CIPP model of evaluation**

As well as criticisms of Kirkpatrick’s model of evaluation based on its narrow concept of outcomes, it is also deficient in guiding a full assessment of a project which might lead to changes and improvements.

A model of project evaluation which does support this is Stufflebeam’s CIPP evaluation model developed in the 1960s.

It is designed to inform decisions about project design through the evaluation of four different aspects:

1. Context
2. Input
3. Process
4. Product

This allows decisions to be made about the degree to which a project has been successful.

Outcomes (or ‘products’ in this model) can be considered in healthcare in terms of Kirkpatrick’s model of evaluation.
**Evaluation models: further reading**

Stufflebeam’s CIPP model of evaluation (continued)

This model bears superficial resemblance to the Plan, Do, Study, Act (PDSA) cycle, a model for improvement widely used in healthcare to implement changes and develop tests.

Whilst the two models do not directly map onto each other, they share the concept that information from a project should be gathered and used to improve that project in a cyclical fashion.

Too often evaluation is focussed solely on the outcome of a project: *Does X work as intended?* rather than asking *Does X work as intended? If not, why not and how can we collect and use evidence to improve it in the future?*

Considering each of these elements provides insights into how a project is developed and how it operates which can inform its further development.

Again, taking a CIPP approach to evaluation from the outset allows you to pose the following questions during the project:

**What should we do? (context/goals)**
Establish goals, objectives and priorities through evaluation of the context in which the project will sit, exploring what is currently available and baseline perceptions, needs and attitudes.

**How should we do it? (input/plans)**
Evaluation of other successful projects or a pilot can inform the design of a project.

**Are we doing it as planned? (process/actions)**
Evaluation of the implementation of the project, monitoring adherence to plans and guidelines and identifying issues as they arise. This is extremely valuable information to inform refinements of the project and allows others to learn from your approach.

**Did the project work? (product/outcomes)**
Measurement of outcomes and the degree to which they match what was intended, is a common part of evaluation as it is usually done in healthcare education. This allows decisions to be made about the degree to which a project has been successful.

Outcomes are often considered in healthcare in terms of Kirkpatrick’s model of evaluation.
Kirkpatrick’s Model

The Kirkpatrick model, originally designed for evaluating industrial and commercial training programmes, can provide another way of thinking about outcomes from a project. Healthcare and education are more complex and adaptive systems than the industrial contexts for which the model was initially designed. However, it has been partially adapted – and widely adopted – in medical education.

Although not necessarily directing evaluators to consider the contexts and processes of an educational project, it may be valuable when we talk about outcomes or when we ask questions such as Does X work? Kirkpatrick’s model helps us to consider what we might mean by outcome or work. The model describes four levels of outcome which can be measured as a result of the project. These are:

1. **Reactions**: How do people engage with the project? How relevant do they think it is? How satisfied are they with it?
2. **Learning**: What knowledge or skills have they learned from the project? Have there been changes in attitude, confidence or commitment as a result of the project?
3. **Behaviour**: What changes in behaviour have occurred as a result of the project?
4. **Results**: What impacts does the project have on the wider world? In healthcare we often talk about patient outcomes (improved patient safety, better prescribing) and organisational outcomes (better staff retention, reduction in critical incident reports).

This allows decisions to be made about the degree to which an project has been successful.

Outcomes are often considered in healthcare in terms of Kirkpatrick’s model of evaluation. While the four levels represent outcomes which are increasingly ambitious from a training programme, it is not necessary to see them as mutually exclusive or hierarchical. All levels have value.

What is required is to consider which levels are most appropriate given the objectives of your project and which are feasible to be evaluated. It may be appropriate to focus on engagement (Level 1) as an outcome if the target group for a project have historically been poorly engaged with healthcare or education and that improving this is an objective of the project. This can be relatively more straightforward to demonstrate. However, it is always challenging to demonstrate changes to outcomes for patients or healthcare organisations (Level 4) from training projects.

With increasing use of electronic records, novel techniques may exist to support higher level outcomes to be evaluated. However, using Kirkpatrick’s model to direct attention towards measuring changes in learning (Level 2) and behaviour (Level 3), where possible, is an important part of making evaluation more effective.
Examples
Project overview
To promote and develop collaborative patient-centred musculoskeletal practice with better outcomes for patients, carers and families through a model of interdisciplinary learning in the community. Four specialty trainees in rheumatology were scheduled to spend 6-8 days in selected general practice surgeries over the course of six months.

Objectives
Specialty trainees should develop a better understanding and demonstration of patient-centred holistic interdisciplinary working in the community;

Specialty trainees should develop a better understanding of the scope of musculoskeletal disease in the community;

Enhancing and understanding of current specialist training and improving relationships between primary and secondary/tertiary care.

Considerations
Context is a huge factor in this project, with only four trainees going into three practices. As a pilot project the main focus has to be understanding the ways in which the project works, how it has been implemented and how it can be improved. As such a clear box evaluative strategy is most appropriate.

Evaluation strategy
Interviews of trainees, GPs and selected staff at the start of the project, followed by observations of STs on days at the general practice surgeries, as well as interviews at the end of the process. This allows a deep exploration of the ways in which the project is enacted at each site and for each individual trainee, helping to develop a better understanding of how the project might be developed for future iterations.

Outcomes
Specific changes were suggested to the ways in which the trainees were recruited. Personal development plans introduced and greater clarity of training opportunities in the GP surgery encouraged. A complex interplay between individuals with different but complementary areas of expertise was observed and needs further exploration.
Example TA training in child occupational therapy

Training of teaching assistants in child occupational therapy

Project overview
A pilot programme of training for teaching assistants (TAs) in two schools in occupational therapy (OT).

The aim is for children to be better supported where concerns have been identified requiring OT to support recovery/development.

A training course of 11 workshops covering six modules, supported by workbooks.

Objectives
- To provide structured training to 12 TAs in OT to increase their knowledge and understanding of OT.
- For TAs to be able to provide OT support to children who require this and would not otherwise receive it – application and successful implementation of the training.

Considerations
The pilot training programme has been delivered in a busy NHS Trust and two large primary schools in south London.

As such a number of practical considerations have been taken in to consideration in conducting the evaluation, particularly limits on the time commitment available for all of those involved.

This has required:
- A focus on a small number of key areas of evaluation and responses/outcomes to be assessed. This will be particularly important in the design of the longitudinal evaluation strategy and plan.
- Noting the context and practical circumstances in which course delivery has taken place, working within the requirements and time available for the schools/TAs.
- Adapting to changing learners on the course from those who initially signed up.
- Taking into consideration and working within the limited time and opportunities for participants, trainers and others to provide feedback and being mindful of ‘evaluation overload’.
**Evaluation strategy**

As a pilot programme involving small numbers of learners, quantitative evaluation will provide some evidence of improvements and knowledge amongst participants.

The priority for the short term evaluation is in evaluating the programme itself, obtaining qualitative feedback to inform refinement and improvement in the course and its delivery ahead of further roll-out.

The longitudinal evaluation strategy is in the process of being developed with input from occupational therapists and others.

This will seek to identify impact on knowledge, behaviour and hopefully outcomes, in line with Levels 2, 3 and 4 indicated in the Kirkpatrick model.

Within the pilot, it has been identified that TA engagement with the programme is a key element to its success and potential roll-out, and therefore this aspect (Kirkpatrick Level 1) has been included in the assessment.

In order to meet this evaluation strategy, a mix of qualitative and some quantitative data (comprising self-efficacy surveys) have been gathered with the qualitative responsive likely to provide the most insightful feedback for programme development at this time.

Some initial quantitative data alongside the qualitative have provided some short term evidence of impact on participants.

**Outcome**

Initial evaluation of the course has been very positive, with great engagement with and enthusiasm for the workshops indicated by TAs who responded to surveys and provided informal feedback.

Some learners have indicated that they are already putting into practice what they’ve learnt with children in their school.

Valuable feedback has also been provided on timescales for delivery of the training and the assignment elements, which will inform development of these elements of the course as it is repeated with new cohorts, and endorsement for the course is sought.
Youth Empowerment Skills (YES):
An innovation in simulation-based health education for adolescents with diabetes

Project overview
The Youth Empowering Skills (YES) programme works with young people with diabetes who have low levels of engagement with diabetes care from socially deprived areas.

Three days in a youth centre delivered by MDT: Interactive and social forms of learning; simulations (mixed-methods); peer-led sessions; naturalistic activities (eg rock-climbing); Peer support and interaction maintained through social media facilitated by the youth worker.

Objectives
- Enhance engagement
- Develop diabetes life-skills
- Promote positive coping and resilience
- Provide social resources and support
- Improve health outcomes

Considerations
Intention to improve health outcomes (black box measure). Also targeting individuals perceptions of themselves, therefore will need some quantitative data. Pilot so some insight into the programme itself required. Complex programme but limited capacity to perform a detailed programme evaluation (which would be clear box).

Evaluation strategy
Mixed methods evaluation (grey box) addressed:
- Specific health outcomes (HbA1c) – black box measure
- Experience of the programme
- Impact participants’ lives

Overall this combines some measurement of specific outcomes of the project alongside an exploration of participants perceptions of their learning and their experience. This is a classic ‘grey-box’ evaluation.

Outcomes
- 2/3 participants had clinically significant reduction in their HbA1c (≥0.5%).
- All would recommend the programme to others.
- Highly rated the usefulness of the programme for their day-to-day life.
- Improvements in their engagement with diabetes care.
- Enhancement in their adaptation to a life with diabetes.
Tier 1 dementia training to hospice staff and volunteers

Project overview
Trinity Hospice in Wandsworth received a small grant to provide Tier 1 dementia training to its entire workforce and volunteers, 400 staff and volunteers in total.

Objectives
The hospice planned a whole organisation approach to raising awareness of dementia, to improve the skills and knowledge of staff and volunteers, with the aim of reducing variations and improving quality of care for patients living with dementia, their families and carers.

Training is to be given to all staff and volunteers through a series of 14 training sessions over one year.

Considerations
The hospice does not have any additional resources to conduct an evaluation other than within the working time of the director of patient services. The hospice has very limited capacity to undertake detailed, longitudinal evaluation.

Evaluation strategy
A pre and post course self-efficacy survey, limited to five key questions, is completed by all staff and volunteers undertaking the training, which is analysed to identify changes in awareness and knowledge following training. Surveys and scales collected in this way represent a common form of ‘black-box’ evaluation.

In order to provide some information on changes to attitudes and behaviour of staff and volunteers, and care of patients with dementia five members of staff have been identified for follow up interview three and six months following completion of the training.

Outcomes
• Of the 50 staff who have completed the training so far, almost all have reported increased awareness and understanding of dementia and skills that can support effective and compassionate communication with people with dementia.
• Highly rated the usefulness of the programme for their day-to-day life.
• Improvements in their engagement with dementia care.
• Enhancement in their adaptation to a life with dementia.

Taking a sample of trainees in this way and exploring their experience at a much deeper level is useful way of achieving some grey box evaluation with a large sample of trainees, and may get towards a clear box evaluation depending on the ways in which the interviews and focus groups are carried out.
Incident reporting frequency following a simulation training programme to improve management of medical deterioration in psychiatric settings

**Project overview**
Delivery of six, weekly half-day in-situ simulation courses to improve management of medical deterioration delivered in an inpatient psychiatric triage unit, aiming to train the whole team.

**Objectives**
To improve the technical and human factors skills of a psychiatric triage team in management of medical deterioration, and therefore the care received by patients.

**Considerations**
Attempting to identify longitudinal impact of training on patient outcomes, using a clinical system (incident reporting – Datix) that is already in place collecting data, therefore this information simply needs to be accessed. This outcome system directly relates to the aims of the training and provides a ‘black box’ approach to collecting data and outcomes (which can be supplemented with further qualitative data to reach grey box level).

**Evaluation strategy**
Collect the Datix (incident reporting rates) frequency data for 7 months prior to and 7 months after the delivery of the 6, weekly, training sessions. Datix data was limited to categories relating specifically to the course – medication issues, injury/illness to patient, self-harm/suicide. Further data quantitative and qualitative collection should help to provide context and explanation to these results.

**Outcomes**
Frequency of incident reporting increased by 33% in the 7 months following the training programme compared to the 7 months prior to training. This was specifically in the categories of medication issues, injury/illness to patient, self-harm/suicide, which were all addressed directly by the scenarios in the training programme.

This demonstrates tangible changes to behaviour of participants in the workplace following training, although there may be other confounding variables impacting this change also.
**Example Adult Advanced Life Support Course**

**Adult Advanced Life Support Course evaluation**

**Project overview**
The delivery of ALS courses takes place nationwide, at multiple centres and by a wide range of staff, to many thousands of learners every year.

**Objectives**
To evaluate the effectiveness of the ALS course.

**Considerations**
The context of each course is irrelevant. The course is delivered identically at each site, according to a strict handbook and timetable. Specific knowledge and skills are taught and required upon completion.

A pre- and post- assessment of knowledge forms part of the completion criteria for participants, alongside a demonstration of satisfactory skill level in a ‘simulation’ test of life support skills.

**Evaluation strategy**
The majority of evaluation data comes from participant performance in the assessments which form part of the programme.

This is purely measuring the output of the course and is therefore a ‘black box’ measure.

Surveys using Likert scales are used to collect participant feedback about the centre and faculty at which the course was delivered, which forms further ‘black box’ data.

This is an appropriate strategy due to the sheer numbers of participants in this programme.

**Outcomes**
Frequency of incident reporting increased by 33% in the 7 months following the training programme compared to the 7 months prior to training. This was specifically in the categories of medication issues, injury/illness to patient, self-harm/suicide, which were all addressed directly by the scenarios in the training programme.

This demonstrates tangible changes to behaviour of participants in the workplace following training, although there may be other confounding variables impacting this change also.
References
References

General
Knowledge for Healthcare: A development framework


Outcome Reporting Bias in Government-Sponsored Policy Evaluations: A Qualitative Content Analysis of 13 Studies Arnaud Vagana http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0163702

Learning objectives
SMART https://careacttarget.org/sites/default/files/file-upload/resources/T1G%20Objectives%202010.pdf

Bloom’s taxonomy: A taxonomy for learning, teaching, and assessing: a revision of Bloom’s taxonomy of educational objectives Lorin W. Anderson, David R. Krathwohl, Benjamin Samuel Bloom

Instructional design: Education and Technology: An Encyclopedia Ann Kovalchick, Kara Dawson

Evaluation (general)

Evaluation models

Realist Evaluation in Practice Mansoor A F Kazi.

CIPP Stufflebeam DL Stufflebeam

The CIPP Model for program evaluation GF Madaus, M Scriven, DL Stufflebeam

Evaluation Models: Viewpoints on Educational and Human Services Evaluation Kluwer Nijhof

Evaluating training programs: The four levels DL Kirkpatrick

Kirkpatrick’s levels and education ‘evidence’ Sarah Yardley, Tim Dorman Medical Education 2012. 46 97-106

Acknowledgments
The toolkit is more successful for being co-produced with our stakeholders. We would therefore like to extend our immense thanks to the following people, who have contributed to the development of the toolkit, by participating in focus groups and workshops; participating in a read through of the draft document; providing examples of evaluations; or otherwise providing valuable support and feedback:

- Mazin Al-Saffar
- Chris Attoe
- Annabel Beasley
- Chris da Silva
- Phil Carver
- Amy Chapman
- Alison Dear
- Elgonia Dreyer
- Laura Emson
- Robel Ghaim
- Catherine Holohan
- Antonina Ingrassia
- Sarah James
- Graeme Jeffs
- Natasha Knight
- Nicole Lallaway
- Zoe Lelliott
- Claire Mallinson
- Sharon Markless
- Ann Ooms
- James Pathan
- Emma Pedlar
- Marilyn Plant
- Martin Powell
- Alison Smith
- Jackie Tavabie
- Alex Thomas
- Beth Thomas
- Josie Turner
- Kathy Tyler
- Charlotte Welsh
- Liz West

...and the Simulation and Interactive Learning Centres at Guy’s & St Thomas’ NHS Foundation Trusts and Maudsley Simulation, South London and Maudsley NHS Foundation Trust.
Glossary
Affective
Affective, especially in relation to Bloom’s Taxonomy, refers to the ways our learners may feel, or deal with emotion.

Bloom’s Taxonomy
A theoretical approach to understanding different levels of learning and thinking. We may wish to consider the learning objectives for the projects we are developing in terms of the appropriate levels of Bloom’s taxonomy to ensure that our objectives are set at the appropriate level for our learners. Bloom’s taxonomy describes learning and thinking in 3 different domains: Cognitive, Affective and psychomotor, which roughly transcribes to ‘thinking, feeling and doing’. Considering the appropriate domain for our project can help guide the language of our learning objectives, which can support a more appropriate and effective evaluation.

Causality
The relationship between two things when one can be shown to cause the other. In educational research and evaluation we hope to prove that changes in knowledge, behaviour or outcomes are due to the education we have delivered. This is not as easy to prove as it sounds. Many things which appear to cause other things may simply be correlated with them. For example, London has an extensive public transport network, and yet has high levels of obesity. These things are correlated, but there is no suggestion that one causes the other. A causative phenomenon might instead be something like the amount of distance that Londoners have to walk on a daily basis. The extensive public transport network may limit the amount of walking that people do on a daily basis, hence leading to the correlation, but it is not directly causative.

Cognitive
Cognitive, especially in relation to Bloom’s Taxonomy, refers to the ways our learners may use knowledge and the mental skills required for this.

Cyclical
Anything which occurs in a cycle, or is recurrent. In education we often talk of cyclical development of a programme, whereby the evaluation of a project is used to develop the next version and so on repetitively until such time as programme is felt to be optimised.

CIPP model
A model of evaluation which is designed to inform programme development, through examining the development, delivery and outcomes of a programme. Originally developed by Stufflebeam, it is a useful way to consider the many purposes which your evaluation may serve.

Domains
Refers to the three types of learning identified within Bloom’s taxonomy. These are commonly referred to as cognitive, affective and psychomotor.

Experimental designs
Experimental designs in education are those which involve any degree of deliberate differentiation or splitting between parts or participants of the project being studied. This if you deliberately split learners into groups to assess learning or impact of a project then this is an experimental design. A randomised-controlled trial is a classic example of a fully experimental design.
Illuminative evaluation
A form of evaluation which explores the conditions and impacts of an educational project, mainly through qualitative means but which can also include quantitative approaches. It is an interpretivist approach and so focuses primarily on the ways in which an educational project is perceived and experienced by those involved. As opposed to a realist evaluation which seeks to create and prove hypotheses about an educational programme and how it works.

Instructional design
A theory which describes a particular approach to the creation of education experiences, through establishing the baseline knowledge of the learners and the desired aims and objectives of the learning experience, to finally create more efficient, effective and appealing learning content and methods.

Interpretivism
A philosophical approach, or methodology, to the acquisition and recognition of knowledge and evidence which contrasts with positivism. Interpretivism states that there is no objective reality or truth which can be measured. The only true knowledge which is worth collecting and analysis is the perceptions, beliefs and experiences of individuals. A significant amount of educational research comes from this philosophical position.

Kirkpatrick model
A commonly used model of evaluation which focuses on different levels of outcome which may be measured and commented on. Widely used in healthcare.

Mechanism
In educational evaluation, and particularly in realist evaluations, mechanisms are the things which lead to change in a learner. They may be positive or negative. They are usually thought of as things which either lead to a change in the reasoning or thinking of a learner, or as things which provide the learner with new resources. Examples might include ‘the stress experienced in a simulation scenario’ which could be either a positive or negative learning experience depending on the learner.

Methodologies
These are the collection of rules and procedures associated with a particular type of educational research. They are usually associated with a particular philosophical belief about knowledge. Examples might include positivism or interpretivism. Methodology refers to overarching approach to knowledge and the effects this will have on the collection and analysis of data. However, in social sciences methodology does not describe the actual methods themselves. In healthcare education, the terms methodology and method may sometimes be used interchangeably. This is often due to a misunderstanding of the difference between the terms by those coming from a background dominated by clinical science research.
Metrics
A system of measurement, or the standards to which measurement is done

SMART model
A useful model to consider when writing learning objectives. It assists in the writing of measurable objectives which are more amenable to an effective evaluation. However, not all learning objectives may fit into this model, especially those for projects which are more innovative or exploratory in their aims.

Paradigm
A particular model of beliefs or thoughts. It may also refer to a specific example of something which represents a particular paradigm. Eg positivism, interpretivism.

Positivist/positivism
A philosophical approach, or methodology, to the acquisition and recognition of knowledge and evidence which contrasts with interpretivism. Positivism states that only knowledge which can be verified through scientific experimentation, mathematical or logical proof can be considered as true. Thus positivists disregard the experience or perception of individuals as a form of knowledge, except where it can be measured and objectively verified. Clinical science is almost exclusively a positivist exercise, with randomised controlled trials a major form of positivist research. Some educational research is also done with a positivist approach. It can be recognised by the use of experimental designs and trials which are designed to prove or disprove a hypothesis.

Pragmatic
This refers to the collection and analysis of appropriate data, which is usually a mix of qualitative and quantitative and which is relevant to the question at hand, regardless of the separate paradigms (positivist or interpretivist) and their specific methodological constraints. Pragmatic approaches may be part of the realist paradigm, or may simply be guided by the availability and accessibility of data in the real world. Using a pragmatic approach can be beneficial in completing an evaluation but may need to be justified if seeking to publish your work.

Prescriptive
When a particular model, theory or belief is enforced upon a person or practice

Psychomotor
Psychomotor, especially in relation to Bloom’s Taxonomy, refers to the ways our learners may demonstrate manual and physical skills.

PDSA cycle
This is a model for introducing change, promoted for use in the NHS. PDSA is shorthand for a cyclical approach to testing a change by developing a plan to test the change (Plan - P), carrying out the test (Do - D), observing and learning from the consequences (Study - S), and determining what modifications should be made to the test (Act - A).

Qualitative
Relating to data which represents the quality of something. Qualitative data general involves description and text in many forms. The key feature is the use of original text to highlight conclusions drawn from the data.
Quantitative
Relating to data which represents the quantity of something. Data which is either directly measured in numbers, or which is translated into numerical form is quantitative in nature.

Quasi-experimental methods
In educational research, quasi-experimental designs share many features of an experimental design, such as deliberately splitting the participants of the study to assess the effects of the project on those who either do or do not receive the key parts. However, quasi-experimental design does not randomise participants to different groups. Instead there is deliberate assignment of participants to different groups. This is used in educational studies as a pragmatic approach, as a fully experimental design may be very difficult to achieve.

Randomised-controlled trials
A form of experimental design. Participants are randomly assigned to different groups before undergoing different conditions to assess the impact of those conditions on groups.

It assumes that randomisation will create groups which are roughly similar at the start of the trial, so that any changes seen can be assumed to be due to the conditions of the trial. This is an approach much used in clinical science and has also been applied in healthcare education research.

There are criticisms of this approach in education, as it implies that all variables can be controlled across different groups.

This assumption is not without risk as the variables inherent in delivery of education are significant in number and potential impact and it is extremely difficult to create a truly controlled environment which makes a randomised-controlled trial valid, but which still represents a realistic and relevant teaching environment.

Rationale
The reason for something being the way it is, or done the way it is done.

Realism
In educational research terms, realism refers to the belief that there are facts and truths about the world which we can attempt to objectively measure (as seen in positivism) but that all of these facts are only measurable through the perception of individuals (a feature of the interpretivist paradigm).

Realist evaluation
An evaluation which uses the philosophical paradigm of realism to inform its approach. Realist evaluations encourage the generation of hypotheses for testing but recognises that these hypotheses come from and must acknowledge the subjective and individual experiences of different participants, within a complex context. It incorporates features of both positivism and interpretivism.
Glossary (continued)

**Schema**
In education, schema refers to patterns of thought or behaviour which are due to the underlying organisation and relationships between different categories of information.

**Scriven’s model**
A model for understanding different types of evaluation by considering the project as a box and describing the degree to which the contents of the box are included within the evaluation, alongside the inputs and outputs from the box (project).

**Statistics/statistical analysis**
Any data which can be collected in numerical form, or translated into numerical form, can be analysed using statistics. Statistical analysis can provide very helpful information, such as the likelihood of results being a real result or simply a product of chance variation.

**Stimulus**
Any thing or event which triggers any form of response.

**Synthesis/synthesising**
The combination of different elements or ideas to create a whole concept with appropriate parts of each original idea and element included.

**Systematic acquisition**
The collection of something, in this case data, in a way which is deliberately designed and organised in a particular way to achieve specific goals.

**Variables**
Any element of a project or object of study which is liable to vary or change. In educational projects almost everything is variable, including the learners, the teachers, the context, the content and the ways in which they all interact.
“For many programme developers, this imperative of proof has led to summative models of programme evaluation and to an almost exclusive focus on the question of whether the intervention ‘worked’ (i.e. whether those who received the intervention improved on a predetermined outcome measure). One implication of this summative approach to programme evaluation is that a typical process of education innovation looks something like this:

1. identify a content area that needs to be taught;
2. develop a teaching module to match the content and implement the module;
3. test to see if it ‘works’;
4. try to figure out what went wrong;
5. tweak the design and delivery;
6. test to see if it ‘works now’ (if it does not, go back to step 4 or, eventually, give up), and
7. publish the success as demonstrating that the content area can be taught in this way.

The result of such an approach to programme development is that any iterative modifications made to the programme are based on post hoc guesses rather than on systematically collected data. Thus, programmes are, at best, improved in a suboptimal way and, more often, are abandoned before they have a chance to mature effectively through systematic and sustained innovation. More importantly for the science of education, regardless of whether the programme ultimately ‘works’ or is abandoned, we don’t learn anything meaningful from these efforts because we are more focused on whether a programme works than on why it does or doesn’t and the implications of its ‘success’ or ‘failure’ both for our understanding of learning and, through this understanding, for future education practices”.

Extract from:
It’s NOT rocket science: rethinking our metaphors for research in health professions education: G Regehr
Please contact: Dr Thomas Simpson MA(Cantab) MBBS MRCP FHEA thomas.simpson@kcl.ac.uk