Evaluation models and design

Evaluation models
The evaluation model that you choose to apply will:

- drive how the evaluation is structured
- the specific content of the data you collect
- the data gathering mechanism you employ
- how it is interpreted and reported on

There is no one single model that is best for any single intervention. The majority of tools or instruments (e.g. questionnaire, interview) or metrics (e.g. exam data, confidence levels) can be used with any of the models presented here.

The most commonly applied theory of evaluation that helps inform design of evaluation for any change or intervention, is Kirkpatrick’s Training Evaluation Model (2009). Kirkpatrick’s model allows for the measurement of potential effects from an intervention at four different levels (immediate to future-focused):

- individuals’ reaction to the training
- individuals’ learning from the training
- individuals’ change in behaviour from the training
- the result/return on investment of training on the organisation as an outcome of the individuals’ behavioural change.

Evaluation design
An evaluation design is a structure created to produce an unbiased appraisal of an intervention’s impact. There are many different types of evaluation design. Here is an overview of the three main types:

- Experimental
Quasi-experimental

Non-experimental.

When evaluating an intervention focused on differential attainment, in the majority of cases, it’s likely that a non-experimental design approach will be most suitable, feasible, and sufficient for the purposes of your evaluation. Within the parameters of applied social science research, this is an acceptable design for evaluation studies.

**Experimental design**

- This involves randomly assigning participants to a treatment or control group.
- This type of design is often considered to be the gold standard against which other research designs are judged, as it offers a powerful technique for evaluating cause and effect.
- Within a medical training context, there may be ethical issues with providing a training intervention to one group in need of support, and not to another.

**Quasi-experimental design**

- This does not have a random assignment component, but may involve comparing a treatment group to a similar group that is not participating in the intervention.
- Quasi-experimental methods are used to estimate the effect of a treatment, policy, or intervention when controlled experiments are not feasible.
- This might be appropriate to evaluate an intervention where participants cannot be randomly allocated to the intervention and non-interventions groups. For example, if those in the intervention group are selected on the basis of a specific criteria, such as having demonstrated lower attainment levels in an assessment. However, using a control group is unlikely to be feasible in this specific context, as it will be difficult to form an appropriate comparison group (e.g. a group matched in all other ways).

**Non-experimental design**

- The key feature of a non-experimental design is that there is no control group, often due to this not being possible in applied research.
- This may be the most appropriate design for many interventions targeting any differences in performance associated with demographics.

**Non-experimental designs may include:**
- **Pre and post-test design:** compares the situation after the intervention with the situation beforehand and attributes any difference to the effects of the intervention. This type of design assumes that any difference in the two observations will tell you if there was a meaningful change over the time period between them, and assumes that any positive change was caused by the intervention. A variation of the pre and post-test design is the time series design, where the evaluator looks for changes over time to determine trends.

- **Post-test design:** here, the intervention group is observed at one point in time after the intervention. Evaluators may focus on comparing responses of sub-groups based on such characteristics as age, gender, ethnicity, place of medical qualification, or level of exposure to the intervention.

- **Case study:** this is a data collection technique that relies on the examination of a limited number of specific cases, which the evaluator anticipates will be informative about the impact of the programme or intervention. Case studies tend to be appropriate where it's extremely difficult to choose a sample large enough to be statistically relatable to the wider population, or where the cases or projects to be studied are likely to be quite complex.