What is it?:

The key feature of experimental and quasi-experimental methods is that they control for selection bias, which gives confidence that the effects being measured are due to the intervention and not down to other differences between those receiving an intervention and any comparison group, therefore avoiding potentially misleading results from non-experimental studies. An experimental design is a scientific experiment in which the researcher manipulates one or more variables, and controls and measures any change in other variables. In educational research this is usually through a randomised-control trial (RCT). Quasi-experimental designs are an alternative to a pure experiment; they estimate what you'd expect if the intervention had not been implemented (the 'counterfactual') by comparing the results with a contrived comparison group using statistical methods in order to capture causal effects. Quasi-experimental methods are less robust than randomised trials but are often more practical, and can sometimes be applied retrospectively, or where randomisation is not possible or desired.

The most common experimental and quasi-experimental designs are:

Randomised control trial (RCT):	The researchers randomly assign potential participants to either receive an intervention or be exposed to 'business as usual' without the intervention thereby enabling the difference in the outcome to be calculated. Randomisation is usually undertaken at the level of individual participants, but could be another unit such as the school/college (as in the practice example below).
Difference-in-difference:	A variant of pre- and post- research where the difference in the before and after results are compared between the intervention group and another group not receiving the intervention (where the participants have not been randomly assigned)
Propensity score matching:	Compares the results between the intervention group and a statistically created comparison group based on an analysis of the factors that affected people's propensity to participate in the programme.
Regression discontinuity analysis:	Compares the outcomes of the intervention group with those just above and below the threshold at which an intervention is assigned (for example if participation in a programme is based on household income, the comparison group would be people just above the eligibility threshold who might be expected to be similar to those who were eligible to take part).

How can it be used?:

These type of designs are useful in relation to the question of 'what works' because they are the strongest design in terms of proving causality between an intervention and the outcomes observed by showing what would have happened without the intervention all other things being equal. They can therefore help to inform decisions about what interventions it's worth investing in, or to choose between different interventions on the basis of which one works best.

Type of evidence:

RCTs are OfS Type 3 (causal) evidence. RCTs are considered the 'gold standard' for establishing causation between an intervention and its outcomes. Quasi-experimental methods could be Type 2 or 3 depending on the robustness of the design.

Strengths:

An RCT is a very robust design because randomising the selection of participants and 'controls' means that, on average, there should be no difference between the intervention group and control group other than in the intervention itself. Therefore, comparing the outcomes should give you a reasonable estimate of the average effect of the intervention in the population.

Although quasi-experimental designs do not use random assignment to select the intervention and control group, when the comparison group is carefully designed these studies can statistically mitigate for selection bias and therefore provide useful causal insights. The strength of the design depends on how similar the comparison group is to those receiving the intervention.

Weaknesses:

While randomised experiments are common in healthcare, there are ethical, political or logistical constraints that often rule them out in social science research. The main ethical concern associated with RCTs is the undesirability of denying an intervention to a subset of participants who could potentially benefit, although the counter argument is that it is unethical to continue with an intervention if there is no evidence that it works. It is sometimes possible to evaluate an intervention without withholding it from trial participants if you are trying out two or several different ways of delivery.

There are sample size and timeline challenges in a widening participation context. To have statistical power, experiments usually need large number of participants, plus it can take a long time for the expected outcomes to come about (although most studies tend to look at evidence in the short-term rather than the long-term).

There are significant implementation issues to make sure the study is robust including getting participants to agree to participate in the intervention or be in the control/comparison group, and ensuring that the data collection is comprehensive and consistent. Plus there are practical difficulties in ensuring that any control group that is selected is not exposed to the intervention in any way and does not receive any types of similar support.

There can be problems generalising the learning from an experimental methods evaluation to different contexts. For example, although an RCT will provide evidence on the outcomes, it is specific to the particular context and it is not always possible to infer whether a similar intervention would have the same effect in another context, or with a different population. Plus the study doesn't necessarily tell you whether the outcomes would be just as effective with a more light-touch intervention or whether adding other components would have a greater impact.

Mixed methods:

Additional research, often qualitative, may be needed to answer specific questions, including why the intervention worked and what the mechanisms of change for particular types of participants could be. Unless there is an element of process evaluation, experimental methods won't tell you why an intervention worked (which is crucial to those who might want to implement the intervention in future).

Expertise:

High

Requirements:

RCTs that require large samples and long study periods to achieve the desired outcomes can be expensive to undertake because of the rigour required in selecting and maintaining the intervention and comparison group and collecting the data appropriately. The intervention delivery has to be consistent in order to know what is being measured. Quasi-experimental methods also usually require relatively large sample sizes (sometimes there are just not enough participants to make this approach viable). High level statistical analysis skills and access to statistical software analysis packages will be needed to manipulate and analyse the data and generate results.

Ethical Considerations:

It's especially important to consider the ethical principles when running experimental and quasi-experimental methods because there is an element of human experimentation. Sometimes policy-makers, practitioners and the potential participants themselves may question the fairness of giving (or withholding) an intervention to some and not others. If there is sufficient doubt whether the intervention is better than doing nothing at all, or genuine uncertainty about which of two options works best, then this provides good justification to proceed with a trial. RCTs should probably undergo an appropriate ethics review.

You will need to get informed consent on data processing from all the study participants whether in the intervention or control group (this is usually done as part of a recruitment process to get people to agree to take part in the study). Censoring rules will apply to the data to ensure anonymity of the participants.

Work Planning:

In the first instance, what you need is a clear rationale for deciding to undertake an experiment and that it is feasible in your context. It can be hard to decide in advance if an experiment is feasible, the required number of participants, whether randomisation is possible, how outcomes might be measured, the likely intervention effects, and how much it will cost. Estimating how easy it will be to recruit people to the study, how many are likely to drop out, the effect size and the standard deviations of outcome measures will help to decide the sample size needed. Doing a feasibility study can help to answer these questions.

You'll also need to decide on the scope of the study. RCTs don't necessarily have to be difficult and expensive to carry out, but since access and participation interventions are more complex then, for example, clinical interventions, in reality you may need to think about setting up a very focused study in order to make sure strong causal claims can be made. There's often a trade off between the scope of the study and the strength of the results.

The trial design is crucial and it's important to set out a written trial protocol in advance describing how you'll get from the initial research questions to reporting the results. Poorly designed trials will not have the statistical power you need to generate conclusions. The protocol will include the measures you will use to capture the outcomes and impact. The choice of outcome measure is crucial and you'll need to consider if the measures are valid (i.e. measuring what you want to measure), sensitive enough to detect important changes, and reliable (reliable measures should produce the same findings when participants are measured again under the same conditions).

A well- defined theory of change for how an intervention may or may not work will help you to decide on the research questions, and the outcomes you are going to use to test effectiveness with (and to inform any process evaluation). If you're designing your own measurement tool (e.g. a survey with indicators to measure your outcomes) then you need to make sure the tools don't unintentionally favour the intervention participants.

The protocol should also include: the data collection methods and tools, how and when data will be collected (as far as possible aiming to collect data from all participants the same); and the plan for analysing the data, such as the statistical methods to be used and any subgroup analysis.

Time and effort needs to go into implementing the protocol and managing the trial in order to avoid poor implementation and weak data. The practical steps will include deciding on the optimal way to deliver the intervention and capture the target population; developing a recruitment plan and materials for participants including consent forms; briefing any collaborators and materials/plans for staff members who will carry out the study; designing the data collection methods and tools; and collecting the data. As noted above, potential pitfalls will include inconsistent recording of data, drop out of participants, and 'contamination' of the control group amongst other challenges.

As with all studies involving data, you'll need to consider how to record and securely hold raw data from participants. If you're using administrative data to capture variables or measure outcomes (e.g. exam results, HESA data), you'll need to have a data sharing agreement and work within the data providers' data policies and procedures covering storage, sharing and archiving requirements.

Analysis:

The type of analysis will depend on the type of trial you're doing. Often RCTs can be reported simply by looking at the whether the difference in the primary outcome is significantly significant ($p < 0.05$) which means that chances are the observed difference is real. The effect size is also important when considering the importance of the results for practice. If the difference made by the intervention is small and not statistically significant it does not necessarily mean the trial has failed (the findings should be described as inconclusive rather than negative).

You'll need to document the procedures used for analysis, so that they can be understood and used by others in the future (e.g. documenting the process used to create any derived variables to support replication of previous analysis). Making the data available to the wider research community will enable others to verify the results and use the data for secondary analysis.

Reporting:

Because these are very strong designs it's usually possible to provide a concise and clear-cut conclusion of the effectiveness of the intervention without lengthy caveats.

Being as transparent as possible about the design and implementation of your study is important so that informed judgments can be made later on about the quality of the evidence. A format called CONSORT (Consolidation of the Standards of Reporting Trials) which was developed for medical trials is increasingly being used in education in order to increase the transparency of RCT reporting.

Useful Link(s):

CONSORT (Consolidated Standards of Reporting Trials) was developed to alleviate the problems arising from inadequate reporting of randomised controlled trials. The CONSORT Statement is an evidence-based set of recommendations for reporting trials: http://www.consort-statement.org