

TASO: Evaluation Methods

Impact Evaluation Methods

		Strengths	Weaknesses	Data required
Level 1 - Monitor	Rationale/ theory of change	<ul style="list-style-type: none"> Documents the logic of why the intervention should work. 	<ul style="list-style-type: none"> Provides no evidence of whether the logic is correct. 	<ul style="list-style-type: none"> None.
	Secondary research (literature review)	<ul style="list-style-type: none"> Straightforward to accomplish and can be done quickly. Gives a sense of whether there is existing evidence that is applicable to whether an intervention is likely to be effective. 	<ul style="list-style-type: none"> Doesn't take account of context – superficially similar interventions can have very different impacts when implemented. Limited by what research already exists. 	<ul style="list-style-type: none"> None.
	Tracking	<ul style="list-style-type: none"> Gives you information about what your participants have done and where they have ended up. 	<ul style="list-style-type: none"> Does not give any information about how the intervention has influenced their destinations. 	<ul style="list-style-type: none"> Individual level outcomes data for intervention participants.
	Pre/Post comparison	<ul style="list-style-type: none"> Gives a sense of whether outcomes have changed over time for participants. 	<ul style="list-style-type: none"> Limited to outcome measures that can be collected at multiple time-points (e.g. grades, attitudes). Assumes that nothing else has changed between pre- and post- measurement that could affect the outcome (not very plausible). 	<ul style="list-style-type: none"> individual level baseline and outcomes data for intervention participants.
	Qualitative research with participants (interviews/ focus groups/ surveys)	<ul style="list-style-type: none"> Allows you to probe whether participants believe the intervention has had an impact on them, and whether they are displaying the attitudes/ knowledge/ skills the intervention should develop. 	<ul style="list-style-type: none"> Relies on participants' self-report, which can be unreliable and influenced by gratitude/desire to please researcher. Selection into the research may limit generalisability. Gives no sense of change over time or change in outcomes. 	<ul style="list-style-type: none"> Contact data for recruiting participants. Qualitative data from respondents.

Level 2 - Compare	Dosage-response	<ul style="list-style-type: none"> • If participants received the intervention at different levels (e.g. didn't attend all the sessions, or receive all the materials) then you can gauge whether those who received the intervention at a higher level displayed a higher level of the outcome. 	<ul style="list-style-type: none"> • There is very likely a third variable (e.g. motivation, free time) that influences both engagement with the intervention and level of the outcome. 	<ul style="list-style-type: none"> • Outcomes data for intervention participants. • Dosage data (i.e. which parts of the intervention each student received)
	Non-random comparison groups (comparing means)	<ul style="list-style-type: none"> • Straightforward analytically – usually specified as a regression model. • You can start to try and isolate the effect of the intervention by including 'control' variables (e.g. demographics, past grades). 	<ul style="list-style-type: none"> • Only gives a correlation not a causal estimate – there are very likely going to be unobservable differences (e.g. motivation) between intervention and comparison groups. • Something that is effective may look ineffective while something ineffective may look effective. • Comparison is only as good as the model specified, and the comparability of the comparison group to the intervention participants. 	<ul style="list-style-type: none"> • Outcome data for both intervention participants and comparison group. • 'Control' variable data for both groups.
	Non-random comparison groups (with matching)	<ul style="list-style-type: none"> • Improves the comparability of the comparison group to intervention group by discarding comparison units that aren't a close match with an intervention participant. 	<ul style="list-style-type: none"> • As above. • Matching still requires the assumption that there are no unobservable differences between intervention and comparison groups that influence the outcome. • Discards units that don't have a match reduces sample size & ability to detect effects. 	<ul style="list-style-type: none"> • As above – requires more data (particularly in comparison group) to maximise odds of successful matching.
	Qualitative research with intervention and comparison group	<ul style="list-style-type: none"> • Enables the exploration of the attitudes, skills and knowledge of participants, while contextualising them by comparison to non-participants. 	<ul style="list-style-type: none"> • Relies on participants' self-report, which can be unreliable and influenced by gratitude/desire to please researcher. • Selection into the research may limit generalisability. • Can be difficult to recruit non-participants. 	<ul style="list-style-type: none"> • Contact details for participants and comparable non-participants. • Qualitative data from respondents.

Level 3 – Identify	Quasi-experimental methods – difference-in-differences	<ul style="list-style-type: none"> • Compares the trend in outcomes (e.g. pre/post) among participants with those among a comparable non-participating group. • With well-matched comparison group, allows the isolation of a potentially causal difference in the rate of change in the outcome among participants. 	<ul style="list-style-type: none"> • Limited to outcome measures that can be collected at multiple time-points (e.g. grades, attitudes), and across both participating and non-participating groups. • Assumes that the trend in the participant group would have been the same as the comparison group in the absence of the intervention – need to probe/test whether this assumption is justified. 	<ul style="list-style-type: none"> • Baseline and outcomes data for intervention participants and a comparison group.
	Quasi-experimental methods – discontinuity designs	<ul style="list-style-type: none"> • Takes advantage of eligibility cut-off points (e.g. grades requirements, household income) to compare the outcomes of those close to either side of the threshold. • If people can't sort themselves across the threshold, and nothing else changes at that threshold, then we can assume that, for those close to the threshold, any difference in outcomes is as a result of the intervention. 	<ul style="list-style-type: none"> • There needs to be a suitable cut-off point. • People can't be able to sort themselves precisely on either side of the cut-off. • Only gives a 'local' effect for marginal individuals, not for those further away from the threshold. • Requires a lot of data, and particularly a lot of cases around the threshold. 	<ul style="list-style-type: none"> • Eligibility (e.g. grades) data for all intervention and comparison cases. • Outcomes data for all intervention and comparison cases.
	Quasi-experimental methods – Synthetic control	<ul style="list-style-type: none"> • It allows for the systematic selection of comparison groups by using the weighted average of units in the underlying population to best approximate the treatment unit before the intervention in order to explicitly model counterfactual trends after the intervention. • Weights are optimally chosen to minimise the distance between treatment and control group; making it clear which units are contributing to the counterfactual "synthetic control" unit. 	<ul style="list-style-type: none"> • The method relies on the linearity of the model of untreated outcomes. This bias can be particularly large if the characteristics of control units are far from those of the treated. • Outcome variable can be subject to transitory shocks • The method is restricted to clearly defined 'local' effects. • Enlarging the sample pool by including units with idiosyncratic variation in predictors runs the risk of overfitting, in which case the resulting synthetic control might poorly mimic the outcome of the treated effect in the absence of the treatment. 	<ul style="list-style-type: none"> • A large dataset, including a relatively long time series of the outcome prior to the intervention and estimates weights in such a way that the control group mirrors the treatment group as closely as possible.

	Randomised-controlled trial	<ul style="list-style-type: none"> • When well-designed, provides a robust causal estimate of the effect of the intervention on participants. • Can be very flexible in the way they're implemented, to fit to context. 	<ul style="list-style-type: none"> • Requires control over the way people are allocated to participate in the intervention. • Requires a large number of intervention and control cases. • Requires upfront planning for evaluation, whereas other methods can be done post-hoc. 	<ul style="list-style-type: none"> • <i>Before intervention:</i> details of all individuals who are in scope for the intervention, in order to allocate them to intervention and control. • <i>After intervention:</i> outcomes data for all intervention and control cases.
--	-----------------------------	---	---	--

Process Evaluation Methods

	Strengths	Weaknesses	Data required
Focus groups	<ul style="list-style-type: none"> • Helps you collect data from a number of people simultaneously (not as resource intensive as interviews) • Allows you to cover more breadth and depth • An be conducted with those receiving and running the intervention (seperatly) 	<ul style="list-style-type: none"> • Discussions can be easily dominated or side-tracked by a minority of participants • Transcriptions of audio can be expensive and take time • Analysis can be time consuming • Requires an experienced facilitator and coder 	<ul style="list-style-type: none"> • Contact data for recruiting participants • Any additional supporting data dependent on the project: <ol style="list-style-type: none"> a) Institutional data such as HESA Student b) Data on attendance or dosage
Lessons learned workshops	<ul style="list-style-type: none"> • Helps you collect data from a number of people simultaneously (not as resource intensive as interviews) • Participatory, and encourages reflection 	<ul style="list-style-type: none"> • Discussions can be easily dominated or side-tracked by a minority of participants • Transcriptions of audio can be expensive and take time • Analysis can be time consuming • Requires an experienced facilitator and coder 	<ul style="list-style-type: none"> • Contact data for recruiting participants • Any additional supporting data dependent on the project: <ol style="list-style-type: none"> a) Institutional data such as HESA Student b) Data on attendance or dosage

Questionnaires	<ul style="list-style-type: none"> • Can be administered to many people across a large geographic area • Can be distributed in a number of ways (email, phone etc) • Can give you quick feedback • Accommodates different kinds of questions (open and closed) 	<ul style="list-style-type: none"> • Requires expertise in question design. Badly designed questions e.g. leading questions can lead to unreliable data • Likely to have low response rates • Will not always provide limited information on “why” • Doesn’t allow for follow-up 	<ul style="list-style-type: none"> • Contact data for survey distribution • Relevant data that you might wish to use to prepopulate the survey, such as programme strand, ethnicity
Interviews	<ul style="list-style-type: none"> • Allows you to collect in-depth information • Allows you to follow-up on interesting answers 	<ul style="list-style-type: none"> • Can be challenging to collect coherent information across interviews • Resource intensive and therefore often small in number 	<ul style="list-style-type: none"> • Contact information for recruitment
Observations	<ul style="list-style-type: none"> • Allows you to collect rich information • Can help you corroborate feedback gathered through other methods 	<ul style="list-style-type: none"> • Can be time consuming • Not always possible • Evaluator’s subjective views can introduce error 	<ul style="list-style-type: none"> • Contact information of gatekeeper
Administrative Dataset analysis	<ul style="list-style-type: none"> • Can track the student activities and outcomes in datasets across the lifecycle 	<ul style="list-style-type: none"> • Analysis is limited to what is in the dataset and the quality of the data • Data is not always recorded consistently 	<ul style="list-style-type: none"> • Student unique identifier to match between datasets