The Evaluation of the Impact of Outreach

Proposed Standards of Evaluation Practice and Associated Guidance

Dr Claire Crawford, Dr Siobhan Dytham and Professor Robin Naylor

June 2017

THE SUTTON TRUST

WARWICK
THE UNIVERSITY OF WARWICK

offa
office for fair access

Department for Education

Contents

1 The Intrinsic Role of Impact Evaluation within the Overall Outreach Strategy 1

2 Context-setting 2
2.1 Impact Evaluation and Process Evaluation 2
2.2 Outreach and Widening Participation objectives 2

3 Key Stages and Principles in an Outreach Impact Evaluation: a summary 3
3.1 Review and Reflect 3
3.2 Plan 3
3.3 Implement 4
3.4 Evaluate 4
3.5 Review and Reflect 4

4 Standards of Evaluation Practice 5
OFFA Standards of Evaluation Practice 5
4.1 Which Level of standards might be appropriate to which type of outreach activity? 6
4.2 Specific Guidance on each of the Levels in OFFA's Standards of Evidence 7
Level 1 Guidance 7
Level 2 Guidance 7
Level 3 Guidance 13

Endnotes 17

Bibliography and Selected References 18

Appendix 1: Existing Outreach Evaluation Guidance 20

Appendix 2: Summaries of Tracking Resources 23
Appendix 2A: HEAT 23
Appendix 2B: UCAS (STROBE) 26
Appendix 2C: NCOP 28

1 The Intrinsic Role of Impact Evaluation within the Overall Outreach Strategy

We take the view that:

- Evaluation of an HE outreach activity means assessing the impact of the activity on its participants, measured against its intended objectives.
- Without evaluating an activity, one cannot be sure that it constitutes a useful way of allocating outreach resources.
- Evaluation enables the practitioner to be more confident that investment is being made in activities which best meet outreach objectives.
- Evaluation assesses the success of current programmes and thereby feeds into the ongoing selection and design of outreach programmes.
- For this reason, evaluation should be an intrinsic element of the overarching outreach strategy. Evaluation provides an evidence base for what works, and for what doesn't.

The OFFA web-pages provide a number of statements setting out how and why evaluation is an intrinsic part of the 'strategic, evidence-led approach' embodied in access agreements.[i] There is also an extensive literature on evaluation principles and methodologies with a corresponding multiplicity of toolkits and guidance available online for WP practitioners.[ii] [iii] There are many sources providing discussion and suggestions for the adopting of principles to be followed in the design of evaluation frameworks.[iv] A summary of evaluation principles and key stages in the development of evaluation strategy is outlined in Section 3. It is our intention that the Standards and the Guidance proposed in this document be tested in a programme of suitably-designed case studies and revised in the light of findings arising from those studies.

2 Context-setting

2.1 Impact Evaluation and Process Evaluation

This Guidance focuses on the evaluation of **impact**. This centres on questions such as, "Is our outreach activity having a transformative effect on participants, as measured, for example, by an increase either in attainment or in their awareness of the benefits of higher education?" In addition to a primary concern with impact evaluation, we are aware that **process** evaluation is also valuable in enabling practitioners to address related questions such as, "How can we refine the nature of our outreach activities in order to enhance their impact or make them more cost-effective?" An impact evaluation of, say, a summer school assesses the extent to which the summer school enables participants to achieve better outcomes along dimensions which reflect the objectives set for the summer school. Through process evaluation we assess how we might make improvements to the design of the summer school in order for it to be able to achieve the same outcomes but at lower cost or in ways which convey additional benefits. Process evaluation is important for a number of reasons, including the insights it can produce for enhancing impact. For the purposes of the current guidance, however, our focus concerns impact evaluation. Nonetheless, we believe that much of what we have to say about impact evaluation is also relevant for the evaluation of process. We also note that the conduct and use of both these forms of evaluation require institutional management to be effective and appropriately configured, and this should be taken into account in strategic planning.

2.2 Outreach and Widening Participation objectives

This Guidance is concerned with outreach as a strategy to raise aspirations and attainment of young people from groups that are currently under-represented in higher education with the intention that they might be more likely to apply to higher education. As such, outreach is just one, albeit crucial, element of an overall widening participation agenda, which will also take account of progression through higher education and beyond into the graduate labour market. We believe, however, that our Guidance can be extended to cover the broader goals of widening participation more generally.

3 Key Stages and Principles in an Outreach Impact Evaluation: a summary

HEIs are encouraged to view the design of evaluation strategies as an intrinsic element of the design of outreach itself. Reviewing outreach strategy in the light of evaluation evidence and planning outreach strategy in ways which will generate reliable evidence of impact are crucial principles and suggest the following key stages and questions which outreach teams might consider in the formulation of strategy.

3.1 Review and Reflect

In the context of our overall strategic aims and targets as an institution, what are our outreach activities trying to achieve? And how do we know if we are being successful?

3.2 Plan

In designing our outreach interventions, how can we ensure that we will be able to collect evidence of a quantitative and/or qualitative nature to be able to evaluate the impact of the interventions: i.e., to know if we are being successful?[v] In designing interventions, we need to **plan in advance** the nature of the evaluation methods and nature of the required data to be collected (see Box below on "**Building effective evaluation into the development of WP strategy**"). Evaluation design principles around the collection and analysis of data include:

Ethics: are there ethical issues involved in collecting the necessary data?

Consent: have all individuals/agencies involved in providing data given their consent? This and other considerations are likely to vary according to the age of the participants.

Feasibility: is it certain that we will be able to collect the evidence/data we will need? Are there technical barriers?

Transparency: are all individuals/agencies involved as fully aware as is appropriate regarding the nature of the intervention and data collection and use?

Timing: what is the time horizon for the objectives of the intervention and hence for the collection and analysis of data?

Security: is there a data management plan which meets data security and privacy requirements?

Standards: will the data collected permit the implementation of evaluation methodologies which will meet desired/necessary standards of evaluation practice?

3.3 Implement

In conducting the intervention, we have ensured in advance that our chosen methods and procedures for data collection have been built-in. This includes any protocols for just which data are to be collected (including how and when) and how they are to be recorded.

3.4 Evaluate

Evaluation methods should be implemented as planned so as to meet the desired Standards of Evaluation Practice. Additional analysis might follow according to the nature of the evidence and results.

3.5 Review and Reflect

The process is an ongoing one, based on learning and enhancement.

Building effective evaluation into the development of WP strategy is best achieved if the objectives of strategy are clearly and comprehensively defined and matched to specific indicators against which outcomes can be measured.

The objectives and the indicators should be made clear when programmes/activities are at the design stage to ensure that methods and processes are put in place for the collection of the data relevant to the chosen indicators.

- For example, if the objective of an outreach activity is to raise attainment of Year 9 pupils, what data will be required for the evaluation of the impact of the intervention?
- What processes and arrangements will be set up to ensure that the relevant data can be collected (which might include collection before, during and after the intervention)?

If no objectives have been set, or no outcome indicators identified or if relevant data are not or cannot be collected, we should ask ourselves the question, “Why are we conducting this outreach activity?”

Evaluating outreach activities requires forward planning: planning which is incorporated into the design of the outreach activities themselves.

4 Standards of Evaluation Practice

There are many sources of 'standards of evidence' available to organisations wishing to adopt structured criteria for the assessment of evidence of impact of interventions.[vi] The OFFA Standards of Evaluation Practice below are an adaptation of the 5-Level Standards of Evidence developed by Nesta (the National Endowment for Science, Technology and the Arts) and The Social Innovation Partnership (TSIP). Nesta's Standards of Evidence were developed to 'embed evidence and learning' in the Centre for Social Action Innovation Fund in a 'structured but flexible way'. Two key characteristics of Nesta's approach are the following:

- "Evaluations are better engaged with, of a higher quality and therefore more useful when they are 'owned' by the innovations, rather than the funder or evaluator". Hence the innovators themselves are able to drive the evaluations.

- Innovators are 'on their own journey, progressing from a place of limited insight to a point at which they feel confident using evidence to design, manage and refine' their outreach strategies.

OFFA Standards of Evaluation Practice

OFFA Level 1:	The HEI can provide a narrative to motivate its selection of outreach activities in the context of a coherent outreach strategy Evidence: The HEI can refer to evidence of impact elsewhere and/or in the research literature on outreach effectiveness
OFFA Level 2:	In addition to a narrative account, the HEI has collected data on impact and can report evidence that those receiving an intervention treatment have better outcomes, though this does not establish any direct causal effect Evidence: Quantitative and/or Qualitative evidence of a pre/post treatment change or a treatment/non-treatment difference
OFFA Level 3:	The HEI has implemented an evaluation methodology which provides evidence of a causal effect of an intervention Evidence: Quantitative and/or Qualitative evidence of a pre/post treatment change on a treated group relative to an appropriate control or comparison group

We view Level 1 above as, broadly, already an expectation within the OFFA access agreement.

4.1 Which Level of standards might be appropriate to which type of outreach activity?

While one might initially suppose that only the most resource-intensive outreach interventions would justify the most robust types of evaluation, in principle almost any outreach activity might be amenable to evaluation at even a Level 3 standard of evaluation practice. For example, McGuigan *et al*. (2012) conduct a (Level 3) randomised field experiment of what the authors describe as 'a fairly light-touch information campaign', based on an intervention in which treated Year 10 pupils in London schools received (i) an email with a link to a website with information on education decisions and (ii) a one-page leaflet. In addition, the treated pupils' teachers received related presentation materials. The authors report statistically significant results, which are consistent with robust (Level 3) evidence from the US on the significant and causal impact on college enrolment, for example, associated with the provision of information on financial aspects of attending college (see Ilie and Vignoles 2017)). Hence, even relatively light-touch interventions which are not resource-intensive can be thought of as the potential subject of Level 3 evaluations. Nevertheless, it would be reasonable if, typically, higher standards of evidence were attached to those outreach activities which are more resource-intensive. The matrix below sets out an indication as to which Levels of the Standards of Evaluation Practice HEIs might aspire for different types of outreach activity, though this should not be regarded as restrictive and, in particular, will vary according to the HEI's outreach objectives.

	Standard of Evaluation Practice		
Type of Outreach Activity	**Level 1**	**Level 2**	**Level 3**
Long-term or multi-activity intervention	A	A	B
Summer school or other HE residential programme	A	A	C
Mentoring	A	A	C
Campus visit or open day (cross department)	A	B	n/a
One-off subject-specific taster session or masterclass	A	B	n/a
One-off school visit	A	B	n/a
HE fair	A	B	n/a

Code:
A = Expected
B = Commended
C = Highly Commended

It is important to stress that this matrix is intended to be illustrative only, as there will be huge variation in practice in the nature of each of the activities identified in the table. For example, if relatively few students are involved in the mentoring activity and/or if it is very light-touch mentoring, then it is unlikely that a Level 3 standard of evidence is appropriate and so this would have a 'n/a' (not

applicable) coding in place of the 'C' in the table. As the 'owners' of the evaluation, it would be for the individual HEI to determine its allocation of evaluation effort across its outreach activities. Similarly, when evaluation evidence across the sector is more developed, extensive and robust, what is expected of HEIs by way of evaluation should respect the validity of appeals to existing evidence where this is directly relevant for specific types of outreach intervention.

4.2 Specific Guidance on each of the Levels in OFFA's Standards of Evidence

Level 1 Guidance

4.2.1 Our starting point is the view that **Level 1** of our proposed Standards of Evidence is already an expectation within the current OFFA access agreement. In addition, the guidance provided below for Levels 2 and 3 should prove helpful in the development of the narrative expected at Level 1.

Level 1 Descriptor: The HEI can provide a narrative to motivate the selection of outreach activities in the context of a coherent outreach strategy. The form of this narrative should be permitted to reflect an evaluation approach appropriate to the HEI's particular context. It might be based, for example, on 'the articulation of a clear theory of change' (see Harrison and Waller, 2016).

Level 2 Guidance

4.2.2 *Level 2 Descriptor*: In addition to a narrative account, the HEI has collected data on impact and can report evidence that those receiving an intervention treatment have better outcomes, though this does not establish any direct causal effect.

Evaluation which is flexible and built-in: As set out above, we see impact evaluation as an intrinsic part of outreach strategy, to be built in at the design stage of the outreach programme and based on a clear enunciation of what outreach activities are trying to achieve. For a Level 2 standard of evaluation, the starting point will be to determine the objectives of the outreach strategy and translate these into outcomes that can be measured or recorded either quantitatively or qualitatively, or both. The design and planning will have to set out a clear arrangement for how the outcome measures will be gathered at appropriate time points, reflecting such considerations as those suggested in Section 3 above. These considerations will vary with factors such as: the outreach objectives, the characteristics of the participants, the nature of the outreach activities and the sets of resources available for evaluation.

HEIs will differ in the extent to which they wish to 'look inside the black box' of the transformation process by which outreach activities might be affecting individual participants. At one extreme, there is a simple 'production function' or 'input-output' approach in which the HEI is interested merely in questions such as, "Do participants in our mentoring scheme achieve higher A-level grades?" This might be characterised as a 'What Works?' approach. Alternatively, some HEIs might be more deeply interested in the theoretical mechanisms through which any transformation might be working or might be expected to work: this can be useful in enabling organisations to clarify and establish the objectives they are setting and the associated evidence they are gathering for their outreach activities.[vii] We see these approaches as complementary.

The planning stage will have identified outreach objectives and outcome measures which will depend on factors such as those described above. Broadly, intended objectives are likely to be grouped into one or other of two categories:[viii]

- *Raising participants' attainment*
- *Raising participants' aspirations*

Associated with each of these broad objectives, more specific outreach targets and associated outcome measures, of either a quantitative or qualitative nature, will have been identified. As an example, suppose that the review and planning stages have highlighted the importance of working with Year 10 pupils in order to generate a greater number of credible applicants to the institution. This might be through raising Key Stage 4 (GCSE) and Key Stage 5 (A-level) attainment together with the HE aspirations of local state-school pupils. In this case, the outcome measures for pupils participating in outreach activities might include:

- *Evidence on participants' attainment*
  - Teacher assessments of pupil performance in Years 10 and 11
  - GCSE attainment
    - For example, GCSE grades in specified subjects
  - A-level attainment
    - For example, A-level grades by subject
  - Education choices at end of Year 11
    - For example, staying on at school or college; A-level subject choices
  - Progression through school or college in Years 12 and 13

- *Evidence on participants' aspirations*
  - Future study intentions
  - Future career expectations
  - Perceptions of university life
  - Awareness of HE study options
  - Awareness/familiarity with HE social and living arrangements
  - Awareness of graduate career opportunities
  - Evidence of educational choices at end of Year 13
    - For example, applying to HE; applying to selective HEIs

For the purposes of impact evaluation at Level 2, one would then want to be able to measure the extent to which either (i) the participants' attainment and aspirations had changed before and after any outreach interventions or (ii) the participants' attainment and aspirations post-intervention compared favourably with those of non-participants, albeit in ways which (for Level 2 purposes) did not establish any causal impact of the interventions. We note that the attainment measures mostly generate data which lend themselves to quantitative analysis while the aspirations evidence is of a more qualitative nature. Therefore a mixture of methods is likely to be appropriate in evaluation analysis. We also note that qualitative data can be used both for developing a narrative regarding the underlying 'transformative' nature of any impact on participants and as a basis, after suitable coding, for quantitative analyses.

We now discuss in more detail each of the Level 2 evaluation approaches referred to above.

4.2.2i Impact evaluation based on participants' attainment and aspirations changing before and after outreach interventions [Difference over time]

This approach will require data based on the sorts of evidence listed above in relation to either or both attainment and aspirations (of participants only), collected for periods both before and after the outreach intervention. This is one of the reasons for stressing the importance of building evaluation planning into the design of outreach strategies from the very start, thereby enabling the gathering of evidence on participants' aspirations prior to their participation in the intervention – through, for example, survey questionnaires which are conducted both before and after intervention.

Collecting the relevant data on participants is likely to require **tracking the participants** through and beyond the period of time in which they are receiving any intervention. Qualitative data are most likely to be collected within the lifetime of the intervention. Quantitative evidence on attainment is more likely to be needed beyond the intervention and, potentially, for a number of years later.[ix] For example, evaluation of a mentoring scheme for Year 10 pupils is likely to require data on GCSE results over a year later and on A-level results typically 3 years later.

Tracking Options

There are various ways in which we might track participants' attainment and HE decisions beyond the intervention. Some of the data might be obtained through follow-up surveys of participants, but here there are problems associated with recall and response. Some data might be obtained from school partners. However, there are a number of ways of tracking pupil attainment and related outcome data based on administrative data and these are potentially of huge value as a source of data for outreach evaluation. These data sources include the following:

- HEAT and regional access trackers
- UCAS: STROBE
- Partnership and Collaborative tracking services

Description and discussion of tracking options is provided in Appendix 2 on Accessing Tracking Data. The key point is that these resources have the potential to enable the evaluator to access administrative data with information on participants' attainment and thereby permit before and after evaluation. Consider the following examples.

Example 1

Participants in your mentoring scheme were selected on the basis that they were assessed by their teachers as not on track to perform well at GCSE and/or A-level and as unlikely to apply to university. Evidence from the National Pupil Database can be used to track pupils through different educational key stages and hence we can see whether, for example, their post-intervention GCSE results exceeded their pre-intervention teacher assessments. From HESA data, we can see whether participants gained admission to university. From UCAS data, we can see the universities to which they applied and hence gauge their level of ambition. Our Level 2 impact evaluation might record statistics such as the increase in the percentage of participants achieving A*- C grades at GCSE relative to teacher expectations prior to the intervention.

Example 2

You are providing masterclasses to Year 12 and Year 13 pupils from a set of local schools which have accepted your invitation to register their pupils in your scheme. From the start of Year 12, prior to the first of the masterclasses, you have teacher assessments of each pupil's expected A-level grades. Your outcome measure is based on each pupil's actual grades at the end of year 13 relative to the original expected grades.

Limitations of the Level 2 before and after evaluation approach

Like all Level 2 evaluations, the before and after approach described above suffers from the problem that we cannot be confident that any improvement in participants' outcomes is ***caused by*** the intervention. We cannot rule out that participants would have anyway gone on to achieve better-than-expected outcomes even in the absence of the intervention. In our first example, perhaps non-participating class-mates of participants also performed better than expected as a result of, say, extra resources being channelled into attainment across the year group. There are potentially many competing reasons for why participants might have performed better than expected other than as a direct consequence of the intervention. In our second example, perhaps those schools choosing to register pupils with your masterclass series were doing so as part of a wider set of supporting measures. In this case, any improvement you observe might be a result of any or all of the supporting measures. A before and after comparison of participants' outcomes might be regarded as 'suggestive of' or 'consistent with' a possible impact of the intervention, one requiring further analysis of a more robust nature.

We also note other reasons to be cautious in interpreting evidence based on before and after comparisons of participants. Consider an evaluation based on the impact of outreach on young people's HE aspirations. To the extent that students' aspirations anyway tend to decline during secondary school, a simple before and after measure of the impact of an evaluation will tend to generate a downward estimate of any true or causal effect.

4.2.2ii *Impact evaluation based on participants' outcomes being better than those of non-participants following outreach interventions: benchmarking* [difference relative to a comparator group]

In discussing the Level 2 before and after approach above, we noted a key limitation arising from the fact that improved (eg 'better than expected') outcomes of participants might reflect, say, a whole school or year group improvement covering not just observed participants but also unobserved non-participants. In such a situation, we would be wrong to attribute an improvement to our outreach intervention. This suggests that a way forward might be to look at the performance of the 'treated group' of participants relative to that of a 'non-treated' comparator group of non-participants. As we will see, however, for this approach to generate a credible evaluation strategy (moving our evaluation standard toward Level 3), great care will be needed in the selection of the comparator or 'control' group.

Consider the following example. You are designing a large-scale residential summer school for pupils who will be entering school Year 12. You set various parameters regarding eligibility (such as attendance at a state school, neighbourhood characteristics etc) and you invite applications. You then select those applicants who most closely match your entry criteria. You have designed your outcome measures to be based on university application decisions in Year 13 and on A-level grades: you are satisfied that you will be able to access the necessary data through tracking your participants using

NPD and UCAS data. Your approach will be to compare the participants' outcomes with those of a comparator group of non-participants. Who are these non-participants? There are various possibilities:

Comparator Group 1: Rejected applicants: an internal control group

You have received 400 applications and have accepted 200 onto the summer school. Your control group consists of the 200 rejected applicants.

We note that an important issue here is that the consent issues you might have with your treated participants are likely to be more problematic in the case of the non-treated non-participants. You are working only with the participants, from whom obtaining consent along with sufficiently detailed information enabling you to track should be feasible: but how will you obtain consent and relevant information on non-participants? This will need to be thought through in the planning stage: the drafting of the application form, for example, is likely to be crucial.

Assuming that consent and other data issues have been addressed, you will be able to obtain the selected outcome measures of both the treated and the non-treated groups and compare them. In so doing, you are conducting a Level 2 impact evaluation study. Why might this approach not meet the criteria set for a Level 3 impact evaluation?

Suppose you find that the treated pupils have *better outcomes* than the non-treated. Can you conclude that you have established a causal effect of your summer school intervention? No: other explanations are just as credible. Such as: (i) self-selection effects – were those selected onto the programme able to present better cases (eg personal statements) than those rejected because of greater school support or stronger personal motivation than those rejected? If so, the treated group might be expected to do better regardless of having participated in the summer school; (ii) acceptance/rejection effects – perhaps the difference in performance stems not from any transformative effect of attending the summer school, but merely from the differential psychological/motivational impact of being accepted onto the programme versus being rejected.

Suppose instead that you find that the treated pupils have inferior outcomes to the non-treated non-participants. Can you conclude that the summer school has had an adverse impact on the participants? No: as we have described it, your selection procedure was to admit onto the programme those applicants whose characteristics were less likely to be associated with high attainment (you are targeting relatively disadvantaged young people). Hence, in the absence of the intervention, they are likely to have performed at a lower level than the non-participants. The programme might have reduced this gap. Or not. With the evidence collected, it's not possible to say with any confidence. Either way, the internal control group method described above fails to meet Level 3 criteria as it does not give a convincing basis for a *causal* interpretation of impact.

Comparator Group 2: An external control group

You are concerned about the adverse motivation effect on rejected applicants and so to evaluate the impact of your summer school on the 200 participants, you use as a control group not rejected applicants but, instead, 200 pupils attending schools with which you have links (and are able to obtain all necessary pupil or school-level data) but whose pupils were not part of the pool from which the summer school was drawn. Having tracked performance data (either from access to school-held records or through administrative data sources), you are able to compare outcomes (eg A-level grades)

of participants with those of pupils in the control group. Again, this is a legitimate approach for a Level 2 study.

Such an analysis would, however, fall short of a Level 3 impact evaluation because you cannot rule out that school or pupil selection effects might explain any performance differences you find. Schools/pupils are *choosing* to sign up to your intervention: you are not picking participants randomly from the potential population. The participating schools/pupils are likely to be unrepresentative of all schools and hence to be different, on average, from your control group: for example, the schools are likely to be more aspirational for their pupils than on average and this is likely to enhance pupil performance, relative to your benchmark, even were they not to have participated in your outreach programme. Consequently, you could not legitimately claim to have uncovered a causal impact of your summer school from a finding of better performance relative to this benchmark. The fundamental problem is that non-random selection of people across the treatment and non-treatment groups creates potential selection biases and hence apparent impacts of the treatment cannot be regarded as causal.

Analysis of differences

In a Level 2 comparison, the evaluator would be advised to control for the influence on the outcome variable of confounding factors. This can be achieved through the use of regression analysis. Consider the case above in which you have outcome data on 200 summer school participants and on 200 non-participants. A simple evaluation measure would be to calculate the average value of the outcome of each of the two groups and compare them. This would generate the 'raw' difference in the mean outcome across the two groups. But this raw mean difference does not allow for the possible impact on outcomes associated with any confounding factors. For example, there may be compositional differences across the two groups in characteristics, such as, say, gender, which might have their own impact on outcomes. Regression analysis can be used in order to 'control for' the influence of observed confounding factors in estimating the impact of the treatment.

Individual-level versus aggregate analysis

In much of our discussion, we have assumed that evaluation is based on comparisons of outcomes of individual participants against outcomes of individual non-participants. If the outreach involves many schools, then it can also be valid to conduct the analysis at school-level, comparing outcomes of participating schools with outcomes of schools not participating.

Summary of potential evidence appropriate for a before-and-after Level 2 intervention:

(i) Teacher assessment of expected participant end-of-year performance (before, possibly during, and after intervention)
(ii) Evidence from survey questionnaires on participant intentions, expectations, perceptions regarding HE (before, possibly during, and after intervention)
(iii) Data on each participant's decision on whether or not to apply to university
(iv) Detailed data on each participant's university application (e.g., by type of HEI, by degree course)
(v) Data on each participant's grades in public exams, such as GCSEs and A-levels

Other data collection considerations:

(vi) How do the data requirements and evaluation strategy generally impact on design of the interventions?
(vii) How many participants are covered by the intervention (as this will affect the likely statistical power of the analysis of impact)?
(viii) How can we obtain pupil/parent/school consent to access required school-based evidence? [x]
(ix) How can we access data which tracks participants beyond the intervention? What are the consent issues here? What are the data management issues?
(x) Are we using individual pupil-level data or, if we have participants from multiple schools, are we using school-level data?
(xi) How can we code the data in ways appropriate to the intended evaluation methodology?

Additionally, when comparing outcomes between a treated group of participants and an non-treated group:

(xii) What might represent a feasible control group? Do we have a potential control group within the programme?
(xiii) How might we generate a control group? What options are available? What are their relative merits?
(xiv) Many of the data and data-related issues identified in (i) – (xi) above apply not only to participants, but also to non-participants – ie to the control group.

Level 3 Guidance

4.2.3 *Descriptor*: The HEI has implemented an evaluation methodology which provides evidence of a causal effect of an intervention

In section 4.2.2 above, we set out Guidance on Level 2 evaluation criteria and discussed the value of adopting a control group approach. We also noted how one reason for Level 2 evaluations to fall short of the more demanding requirements of a Level 3 standard arises through non-random selection of individuals across treatment and control groups. We described a number of examples designed to show limitations of specific control group designs at Level 2. To meet the Level 3 standard, evaluations have to meet the criterion of establishing, with confidence, that the intervention has had a *causal* impact on participants' outcomes.

Level 3 approaches

4.2.3i Many would argue that the gold standard for the design of impact evaluation is the **randomised control trial**.[xi] This is designed to overcome the various possible problems we highlighted with the schemes described in section 4.2.2. The key element of the design is that, by selecting individuals randomly into treatment and control groups, there are no systematic differences across the two groups in either observed or unobserved characteristics. Hence, there are no self-selection issues to be worried about, for example. In the absence of any treatment (intervention programme), there is

no reason to expect there to be any systematic differences in outcomes across the two groups. Accordingly, any differences which are found can be attributable to the impact of the programme. To minimise the effects of random noise, the greater the number of individuals in the two groups the greater the statistical power of the test.

In practice, there are a number of challenges to be confronted in implementing an RCT. These include the sets of issues described in relation to data needs in section 4.2.2. In addition, there is the key issue of how to make the random assignment of individuals across the treatment and control groups. As well as the matching and self-selection issues, which RCTs are designed to overcome, there is also the need to avoid problems such as the acceptance/rejection or inclusion/exclusion effect: ideally, non-participants should not understand that they are non-participants in some event that is occurring. This can raise ethical issues – though it is notable that within medical research there is a widespread understanding of the legitimacy of the principle of random assignment to treatment.

A common way around the ethical issue in the context of educational interventions is to sequence treatment across potential participants. For example, one recruits a number of schools into an intervention programme and assigns randomly some of the schools to the group whose pupils will receive a treatment and others to a control group whose pupils will not be treated. The schools are informed in advance that those not selected for the programme this year are aware that they will have the same or an equivalent intervention on a subsequent occasion.

The evaluation then proceeds on the sort of basis described under the Level 2 guidance set out in section 4.2.2, tracking and comparing outcomes of treated and non-treated individuals, but based on a control group which overcomes the limitations highlighted there.

4.2.3ii Impact evaluation based on a **before-and-after change** in participants' attainment and aspirations **relative to that of non-participants** [difference-in-difference]

We note that 4.2.2i concerned *before and after changes **within** the group of participants* while 4.2.2ii was described as referring to *differences **between** participants and non-participants*. A difference-in-difference (D-i-D) approach takes account of both dimensions, addressing the question, "Did the treated schools/pupils *improve by more than* the non-treated schools/pupils?" This offers a way of overcoming a major limitation of the simple before-and-after approach, which assumes that any improvement in performance over time for participants is a consequence of the intervention on participants: instead, we are now explicitly checking that the improvement did not also occur in non-treated cases. The D-i-D approach does assume that treated pupils, had they not been treated, would have experienced the same change in performance as the non-treated. This is the 'common trends' assumption, the validity of which cannot be proved but can be given suggestive support through evidence of similarity in trends in periods prior to the intervention.[xii] The D-i-D approach can be regarded as meeting a Level 3 Standard as the common trends assumption is generally regarded as more acceptable than those underlying the simple before-and-after approach described in section

4.2.2i Implementation of the D-i-D method is broadly similar to that described in sections 4.2.2i and 4.2.2ii, but with a focus on the collection of data before and after intervention for both treatment and control group cases. Data on outcomes of the control group might be school-based records on pupils attending schools with which the HEI has links, as described previously, or might be based on

administrative data obtained either directly from NPD-HESA datasets or might be based on tracking resources such as through HEAT or UCAS-STROBE, etc.

In order to reduce possible measurement and evaluation problems associated with non-random selection of schools/pupils into outreach interventions as well as to increase the likelihood of common trends between treatment and control groups, selection of the control group should be based on both groups being as similar as possible in characteristics. This can be achieved through access to data and/or data reports produced by HEAT and UCAS-STROBE, for example. More sophisticated and robust than a comparison of outcomes of the treatment group with outcomes of some control group of schools/pupils with average characteristics is a comparison between the treatment group and a control group designed to share characteristics similar to those treated (for example, in free school meal or attainment measures). The extent to which the control group can be designed to be similar to the treated will depend on what data are observed both within the treatment group and in the administrative data available. As with the Level 2 approach, regression analysis should be used to estimate the effects of treatment rather than a simple comparison of raw differences in outcomes. The regression approach takes account of the potential influence of observed confounding factors.

4.2.3iii For an RCT to be implemented, individuals and/or schools have to be assigned randomly, as we have discussed. For outreach activities, this is not always possible. For example, partnering schools or universities themselves might have strong preferences for particular individuals or groups of individuals to participate in outreach activities, perhaps according to WP or related characteristics.

Propensity score matching (PSM) offers one method which attempts to correct for non-random selection.[xiii] The key idea here is that the outcomes of individuals in the treatment group are compared to the outcomes of individuals who were not treated but who were as likely to have been treated as were the treated. The analysis proceeds in steps. First, across all individuals, treated or non-treated, one obtains the predicted probability (or 'propensity score') for each to have been treated, based on observed characteristics associated with participation. The predicted probability will run on a scale from 0 to 100%. Second, each of the treated individuals is matched to non-treated individuals with similar propensity scores. For the method to be reliable, it is helpful if there is a reasonable balance across the treatment and control (non-treated) groups in terms of both the propensity scores and observed characteristics. Third, based on the new sample consisting of treated individuals matched by propensity scores to the control group of non-treated individuals, the outcomes of the treated and the non-treated are compared.

One can think of PSM as one of the 'second best' approaches for identifying causal effects of an intervention. One of its major limitations is that matching can be achieved only on observable characteristics and so, unlike a well-designed RCT, one cannot be sure that one has not introduced systematic differences in unobservable characteristics through the nature and design of the outreach activity itself. Complementary analysis and discussion can help to strengthen the case that a PSM-based approach has generated robust evidence of impact. We note that the Access Project (TAP) has achieved a Nesta Level 3 evaluation standard based on a PSM-based analysis.[xiv]

4.2.3iv A further route by which one might attempt to uncover a causal impact of an intervention is through an approach based on **Regression Discontinuity Design (RDD)**. This can be thought of as an attempt to approximate an RCT approach. Suppose that there is some hurdle or critical value of a

criterion which determines whether pupils will be selected for an intervention. For example, suppose that there was a particular level of family income at or below which individuals were offered participation in an outreach activity but above which they were not. Given that, on average, family income is associated with educational attainment, then – in the absence of the outreach activity – one would expect those whose family income is at or below the threshold to have lower attainment and also to differ in other characteristics from those from higher income families in ways which might correlate with educational attainment. In other words, we don't seem to have a valid control group. But suppose that we were considering an activity with very many participants. Then many of these individuals might be only fractionally above the threshold and many others will be below it by similarly vanishingly small amounts: in other words, there will be a lot of individuals who are essentially equivalent in family income (and are unlikely to be systematically different in other ways) but fall essentially randomly one side or the other of an arbitrary income threshold. An RDD compares the outcomes of the treated group of individuals who just meet the criteria with those of the non-treated group who just fail to meet the criteria.

The D-i-D, PSM and RDD methods, as we have discussed them, lend themselves to evaluation based on quantitative data generated from administrative records, such as NPD, HESA and UCAS datasets. Data on the outcomes of the control groups are generated within these datasets and can be collected beyond the lifetime of the intervention through official records, including archived records of attainment prior to the intervention period. The control group can be created post facto. In the case of the RCT approach, the identities of individuals in each of the treatment and the control groups are established prior to the implementation of the intervention. Evidence on both groups can be collected before, during and after the programme, and this can include qualitative data.

Implementation

Each of the design, data and analysis elements of a Level 3 standard of evaluation appears quite daunting. Nevertheless, some HEIs are already implementing some of these approaches and are keen to confront potential challenges and to develop and enhance further their evaluation practice and strategies. Other HEIs are at a different point in their adoption of evaluation methods and approach. In either case, consideration should be given at the design stage regarding the question of whether to commission a 3rd party to supply expertise in any aspect of the evaluation. Working with expert evaluators can assist HEIs in developing greater competence in independent evaluation as they progress.

Data issues

These are as set out under the draft guidance for Level 2 with, in addition, the following considerations:

(i) Which Level 3 evaluation method?
(ii) Hence what are the data requirements?
(iii) How are we going to access the required data? Within programme or HEAT etc?
(iv) How does this impact on design of the interventions?
(v) Are we going to conduct the analysis in-house or use a 3rd party evaluation service?

Endnotes

[i] See Evaluating your access activities and expenditure (https://www.offa.org.uk/universities-and-colleges/guidance/evaluating-your-access-activities/) and Setting your access agreement strategy (https://www.offa.org.uk/universities-and-colleges/guidance/setting-access-agreement-strategy/).

[ii] See Appendix 1 for a non-exhaustive list of some of the resources available.

[iii] Links include: HEFCE/Progression Trust toolkits for practitioners (https://www.heacademy.ac.uk/system/files/resources/evaluation_3rd_0.pdf) and Kirkpatrick's evaluation model (http://www.kirkpatrickpartners.com/OurPhilosophy/TheKirkpatrickModel).

[iv] See, for example, (i) "Measuring Success: a guide to evaluation for AimHigher" by Sue Hatt, 2007, (ii) Theory, Practice and Impact in Widening Participation: the NERUPI Framework, (iii) the Education Endowment Foundation, Approach to Evaluation, EEF, and (iv) HEFCE/Progression Trust toolkits for practitioners.

[v] Broadly, quantitative evidence refers to readily measurable and comparable data on outcomes such as attainment (eg an improvement in A-level grades) or aspiration (eg an increased probability of applying to HE). Qualitative evidence includes a variety of types of evidence such as: perceptions of the nature of university life, awareness of graduate career options, confidence in overcoming HE challenges. These might be derived from questionnaires, interviews, focus groups or observation studies. Through coding procedures, many qualitative measures can be translated into quantitative data.

[vi] Nesta's "Standards of evidence: an approach that balances the need for evidence with innovation" by Ruth Puttick and Joe Ludlow, October 2013, provides links and traces the origins of the Nesta Standards of Evidence to the GLA's Project Oracle, which was designed around the evidence base of youth programmes in London. See also "What counts as good evidence? Provocation paper for the alliance for useful evidence" by Sandra Nutley, Alison Powell and Huw Davies, University of St Andrews, 2013.

[vii] The NERUPI framework provides a very rigorous theoretically-informed methodology for linking WP aims and objectives to impact evidence (see "Widening Participation habitus and capital: exploring a framework to evaluate WP interventions," Annette Hayton and Andrew Bengry-Howell).

[viii] We note that the two objectives are quite likely to be mutually-reinforcing.

[ix] This is especially true if we are interested in the progression of outreach participants through university and into the graduate labour market.

[x] An important set of issues concerns the development of productive and collaborative relationships with schools as this will impact significantly on both the effectiveness of outreach and also on the feasibility and quality of evaluation.

[xi] For a detailed discussion of RCTs and how to conduct them in the context of public policy impact evaluation, see: "Test, learn, Adapt: developing public policy with randomised control trials," by Laura Haynes, Owain Service, Ben Goldacre and David Torgerson, Cabinet Office Behavioural Insights Team, 2012. See also papers by Ilie and Vignoles, 2017 and by McGuigan, McNally and Wyness, 2012.

[xii] Note that this imposes additional data requirements in having information on outcomes and other features of schools/pupils in years prior to the introduction of the intervention activity.

[xiii] See "The use of propensity score matching in the evaluation of active labour market policies," by Alex Bryson, R. Dorsett, and S. Purdon, DWP Working Paper, London: DWP, 2012.

[xiv] See "The Access Project: evaluation support," by Jake Anders, Lucy Stokes and Heather Rolfe, NIESR, February 2015.

Bibliography and Selected References

Alkin, M. C. "Debates on Evaluation", Thousand Oaks, USA: Sage Publications, 1990.

Allen, M. "In defence of targeting? The Open University's Community Partnerships Programme," Widening Participation and Lifelong Learning, 12(Special Issue), pp. 138-53, 2010.

Anders, J., Stokes, L. and Rolfe, H., "The Access Project: evaluation support," NIESR, February 2015.

Babudu, P., Nicholls, J. and Trevithick, E., "Innovations, good evidence and lessons learned: embedding evidence in to the Centre for Social Action Innovation Fund," TSI/Nesta, 2016.

Baker, S., Brown, B. and Fazey, J. A. "Individualization in the widening participation debate," London Review of Education, 4(2), pp. 169-82, 2006.

Bryson, A., Dorsett, R. and Purdon, S., "The use of propensity score matching in the evaluation of active labour market policies," DWP Working Paper, London: DWP, 2012.

Delaney, C. F. "Critical Realism," In: Audi, R. (ed.) The Cambridge Dictionary of Philosophy, (2nd ed.), New York: Cambridge University Press, 1999.

Doyle, M. and Griffin, M. "Raised aspirations and attainment? A review of the impact of AimHigher (2004-2011) on widening participation in higher education in England," London Review of Education, 10(1), pp. 75-88, 2012.

Dytham, S. and Hughes, C., "Widening Participation and Outreach Research and Evaluation: Where are we now?" Briefing Document for the OFFA-Warwick Symposium Evaluating Outreach, http://wrap.warwick.ac.uk/id/eprint/88772 , 2017.

Gorard, S. and Smith, E. "Beyond the 'learning society': what have we learnt from widening participation research?," International Journal of Lifelong Education, 25(6), pp. 575-94, 2006.

Harrison, N. and Hatt, S. "'Disadvantaged Learners': Who Are We Targeting? Understanding the Targeting of Widening Participation Activity in the United Kingdom Using Geo-Demographic Data From Southwest England," Higher Education Quarterly, 64(1), pp. 65-88, 2010.

Harrison, N. and Waller R., "Evaluating outreach activities: overcoming challenges through a realist 'small steps' approach," Perspectives: Policy and Practice in Higher Education, DOI: 10.1080/13603108.2016.1256353, 2016.

Hatt, S., "Measuring Success: a guide to evaluation for AimHigher," https://www.heacademy.ac.uk/system/files/ah_measuring_success.pdf, 2007.

Haynes, L., Service, O., Goldacre, B. and Torgerson, D., "Test, learn, Adapt: developing public policy with randomised control trials," Cabinet Office Behavioural Insights Team, 2012.

Hayton, A. and Bengry-Howell, A., "Theory, evaluation, and practice in widening participation: a framework approach to assessing impact," London Review of Education, 14(3), pp. 41-53, 2016.

Ilie, S. and Vignoles, A., "Fair Access to Elite Universities: Literature review and focus group report," Mimeo, Faculty of Education, University of Cambridge, 2017.

Jennings, P. L. "Hidden Treasures: The contribution of critical realism to evaluation," www.aston.ac.uk/EasySiteWeb/GatewayLink.aspx?alId=171885, 2013.

Kettley, N. "The Past the present and the future of widening participation research," British Journal of Sociology of Education, 28(3), pp. 333-47, 2007.

Maxwell, J. A. "Evidence: A Critical Realist Perspective for Qualitative Research," In Denzin, N.K. and Giardina M D (eds) Qualitative Enquiry and Social Justice, USA: Left Coast Press, 2009.

McGuigan, M., McNally, S. and Wyness, G., "Student awareness of costs and benefits of educational decisions: effects of an information campaign," CEE Discussion Paper 139, Centre for the Economics of Education, London School of Economics, 2012.

Nutley, S., Powell, A. and Davies, H., "What counts as good evidence? Provocation paper for the alliance for useful evidence," RURU, School of Management, University of St Andrews, 2013.

Pawson, R. and Tilley, N. "Realistic Evaluation," London: Sage Publications Ltd, 1997.

Pawson, R. and Manzano-Santaella, A. "A Realist Diagnostic Workshop," Evaluation, 18(2), pp. 176-191, 2012.

Puttick, R. and Ludlow, J., "Standards of evidence: an approach that balances the need for evidence with innovation," Nesta, October 2013.

Reay, D., David, M. D. and Ball, S. J. "Degrees of choice: Social class, race and gender in higher education," Stoke on Trent: Trentham Books, 2005

Scriven, M. "A Summative Evaluation of RCT Methodology: An Alternative Approach to Causal Research," Journal of Multi-disciplinary Evaluation, (5)9, pp. 11-24, 2008.

Shaw, J. "The diversity paradox: does student diversity enhance or challenge excellence?" Journal of Further and Higher Education, 33(4), pp. 321-31, 2009.

Smith, N. L. "Foreword," In: Cousins, J.B. and Chouinard, J.A. (eds.) Participatory Evaluation Close Up: An Integration of Research-based Knowledge, USA: IAP-Information Age Publishing Inc., 2012.

Stufflebeam, D. L. and Coryn, C. L. "Evaluation Theory, Models, and Applications: Research Methods for the Social Sciences," (2nd ed.), USA: Jossey Bass, 2014.

Weiss, C. H. "Have We Learned Anything New About the Use of Evaluation?" American Journal of Evaluation, 19(1), 21-33, 1998a.

Weiss, C. H. "Evaluation," (2nd ed.), USA: Prenice Hall, 1998b.

Appendix 1: Existing Outreach Evaluation Guidance

- Action on Access (2005) Widening Participation: A Rough Guide for Practitioners http://actiononaccess.org/wp-content/files_mf/roughguide.pdf

- Aimhigher Greater Merseyside (no date) Evaluation Toolkit. A resource for evaluating Aimhigher. http://www.heacademy.ac.uk/resources/detail/aim_higher/AHGTM-Evaluation_Toolkit-PBR

- AMOSSHE Value & Impact online toolkit (2011) http://www.amosshe.org.uk/value-and-impact-toolkit#toolkit

- Analyse this! (website) http://archive.learnhigher.ac.uk/analysethis/index.html

- Andrews, J., Clark, R. and Thomas, L. (2012) What Works? Compendium of effective practice in higher education retention and success: HEA https://www.heacademy.ac.uk/resources/detail/what-works-student-retention/Compendium_Effective_Practice

- Department for Business, Innovation and Skills (2013) National Strategy for access and student success in higher education https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/299689/bis-14-516-national-strategy-for-access-and-student-success.pdf

- Education Endowment Foundation (EEF) DIY Evaluation Guide https://educationendowmentfoundation.org.uk/resources/diy-guide/getting-started/ https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_DIY_Evaluation_Guide_(2013).pdf

- Forum for Access and Continuing Education: 12 March 2013 the national conference Evidence Based Access (FACE) - Evidence Based Access Agreements: Target setting, evidence building, monitoring and evaluation http://www.f-a-c-e.org.uk/conference-and-events/events/evidence-based-access-agreements-12-03-13.htm
    - Targeting and Target Setting - Mike Thompson, West Midlands Aim Higher
    - Evaluating Impact and Progress - Sharon Smith, Kent and Medway Progression Federation
    - Monitoring and Evaluation - Dr Penelope Griffin, University of Nottingham

- Gale, T., Sellar, S., Parker, S., Hattam, R., Comber, B., Tranter, C. and Bills, D. (2010) Interventions early in school as a means to improve higher education outcomes for disadvantaged (particularly low SES) students, A design and evaluation matrix for university outreach in schools, The Department of Education, Employment and Workplace Relations, Commonwealth of Australia http://dro.deakin.edu.au/view/DU:30040776

- Hayton, Annette and Andrew Bengry-Howell (presentation). The Activities Matter: Exploring a framework to evaluate the impact of university-led outreach interventions Podcast: https://itunes.apple.com/gb/podcast/srhe-society-for-research/id594177334 'The Activities Matter: Exploring a framework to evaluate the impact of university-led outreach interventions' PowerPoint slides:

https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjYtbD2xoPUAhUqBMAKHSgqBbEQFggqMAA&url=https%3A%2F%2Fwww.srhe.ac.uk%2Fdownloads%2Fevents%2F165_HaytonBengryHowellAWP090215.pptx&usg=AFQjCNHyhG4I5e15SSPPVzNLj5yBSzMH-w

- Higher Education Academy (HEA) HE outreach to widen participation: toolkits for practitioners. Evaluation https://www.heacademy.ac.uk/resources/resource2322

- HEA - A brief guide to data sources https://www.heacademy.ac.uk/resources/detail/aim_higher/Aimhigher_guide_to_data_sources

- HEA: HE outreach to widen participation: toolkits for practitioners. Targeting https://www.heacademy.ac.uk/resources/resource2324

- HEA, Hatt, S. (2007) Measuring success: a guide to evaluation for Aimhigher https://www.heacademy.ac.uk/resources/detail/aim_higher/AH_measuring_success

- Higher Education Access Tracker (HEAT) Training and Guidance pages http://heat.ac.uk/support-training/

- Hefce (2007) Higher education outreach: targeting disadvantaged learners http://webarchive.nationalarchives.gov.uk/20100202100434/http:/www.hefce.ac.uk/pubs/hefce/2007/07_12/07_12.pdf

- Hefce (2008) Guidance for Aimhigher partnerships http://webarchive.nationalarchives.gov.uk/20100202100434/http://www.hefce.ac.uk/pubs/hefce/2008/08_05/

- Hefce (2008) Further guidance on evaluation planning http://www.lancaster.ac.uk/fass/events/capacitybuilding/docs/HEFCE%20Further%20guidance%20on%20evaluation%20planning%20_April%202008_.pdf

- Hefce (2010) Widening participation strategic assessments: guidance on developing evaluative approaches to widening participation activities and commitments. Circular Letter 24/2010 and attached Annexes A and B. http://www.hefce.ac.uk/pubs/year/2010/cl,242010/

- Hefce: Literature review of research into widening participation to higher education http://www.hefce.ac.uk/pubs/rereports/Year/2013/wplitreview/Title,92181,en.html

- Hefce - Evaluation Capacity Building ECB 'Toolkit': HEFCE commissioned CSET (now part of the Higher Education Research and Evaluation Centre) and REAP (Researching Equity Access and Participation) in the Department of Educational Research at Lancaster University to undertake evaluation capacity building in widening participation: http://www.lancaster.ac.uk/fass/events/capacitybuilding/toolkit/index.htm

- Horizon Scanning Report (2013): Widening Participation, Retention and Success Research Synthesis http://www6.plymouth.ac.uk/files/extranet/docs/WID/PedRIO_WP_Horizon_Scanning_Report_FINAL.pdf

- Hutchison, Dougal and Ben Styles (2010) A guide to running randomised controlled trials for educational researchers https://www.nfer.ac.uk/publications/RCT01/RCT01.pdf

- Jones, R. (2008) New to widening participation? An overview of research, Evidencenet, HEA. https://www.heacademy.ac.uk/resource/new-widening-participation-overview-research

- The Kirkpatrick Model http://www.kirkpatrickpartners.com/OurPhilosophy/TheKirkpatrickModel

- National Foundation for Educational Research (2013). How to... Use focus groups: Get the most from them (How to Guides). Slough: NFER. https://www.nfer.ac.uk/publications/RESM04/

- National Foundation for Educational Research (2014). How to… Develop a questionnaire survey: Ask the right questions (How to Guides). Slough: NFER. https://www.nfer.ac.uk/publications/RESM06

- National HE STEM Programme 'Using Data: An evidence-based approach to improving transition, induction and retention' http://www.ncl.ac.uk/ltds/assets/documents/6068_HE-STEM_Data_Evaluation_AW_low_res.pdf

- OFFA: Evaluating your access activities and expenditure https://www.offa.org.uk/universities-and-colleges/guidance/evaluating-your-access-activities/

- Sutton Trust Research 'Evaluating Access' http://www.suttontrust.com/researcharchive/evaluating-access/ This report reviews national and international research on widening participation and access programmes to find out which methods are most likely to help disadvantaged pupils get into higher education.

- University of Hertfordshire: Outreach and Widening Participation Evaluation Framework http://www.herts.ac.uk/__data/assets/pdf_file/0016/111517/UH-OWP-Evaluation-Framework-Overview.pdf based on the work of Bengry-Howell and Hayton (2015)

- Widening Participation Research and Evaluation Unit (WPREU) (at University of Sheffield) https://www.sheffield.ac.uk/schools/outreach-programmes/wpreu
    - Guidance available here: https://www.sheffield.ac.uk/als/wp/wpevaluation
    - Good Practice When Designing Evaluation Questionnaires: https://www.sheffield.ac.uk/als/wp/wpevaluation/goodpractice
    - Evaluation Focus Groups and Interviews: https://www.sheffield.ac.uk/als/wp/wpevaluation/interviews

Appendix 2: Summaries of Tracking Resources

This appendix provides a brief summary of selected resources and services available in the HE sector for tracking participants of outreach activities as part of the evaluation of the impact of outreach.

2A HEAT Page 23

2B UCAS (STROBE) Page 26

2C NCOP Page 28

Appendix 2A: HEAT

Background

The Higher Education Access Tracker (HEAT) Service is a monitoring and evaluation service for member institutions. Originally a self-sustaining HEI subscription service, it has received three years funding from HEFCE to expand HEI membership across England and will return to a self-sustaining service from December 2018. The HEAT service is developed by members for members, collaborating on the sharing of data, analysis and research evidence.

Tracking and monitoring

Recording of activities and of participants

The HEAT service enables member institutions[1] to register their outreach activities, recording securely and confidentially the nature of those activities and the characteristics of individual participants, using common protocols and coding fields. Sharing across members can also be put in place through the system, to facilitate partnership recording and reporting (e.g. for NCOP consortium).

Tracking of participants

Outreach participants can be tracked in a variety of ways.

1. Surveys (within and outside HEAT)

> *Within HEAT*: There is a survey tool within the HEAT service which enables HEIs to build up longitudinal information on participants of both a qualitative and a quantitative nature
>
> *Outside HEAT*: The HEAT evaluation is being designed to enable HEIs to upload externally-generated survey information so that this sits alongside data and reports generated within the HEAT database

[1] NCOP evaluators can also access HEAT services.

2. HE outcomes within HEAT institutions

The sharing of data among HEAT subscribers enables participants in outreach activities in one institution to be tracked across any other HEAT subscribing institution to which they might apply.

3. Attainment and HE outcomes through 'fuzzy' matching to NPD-HESA datasets[2]

HEIs can 'locate' outreach participants within NPD-HESA data through fuzzy matching and hence track their attainment at Key Stages and into HE.

Consent

HEIs subscribing to HEAT are asked to present a privacy notice to participants which tells them that their data will be shared with HEAT and other agencies such as HESA. HEAT has developed an animation which HEIs can show to students, explaining why their data are of benefit for research and evaluation. Partly in view of the GDPR coming into force next year, HEAT are developing an electronic registration tool that can be used in schools via iPads and phones, with a built-in privacy notice and opt-in mechanism. These tools can be accessed by any member institution.

Comparator groups

Data on aggregate outcomes of comparator groups can be obtained within NPD-HESA data by matching against participant characteristics: enhancement of ease and quality of matching is under review between HEAT management and the Department for Education. HEAT is also negotiating access to Department for Education NPD/Attainment data for a counterfactual analysis based on control group data.

Implementation

The developing HEAT evaluation planning tool supports HEIs in building their evaluation into an overall outreach strategy. HEIs have the option to record an extensive range of descriptors into the HEAT evaluation planning tool. The aim is to have this tool "live" by Autumn 2017. This includes:

- The objectives of the activity (eg detailed target group(s), student and school-level factors, nature of identified need)
- The anticipated outcome measures (eg learning/behavioural outcomes, attainment results, KS4/5 Report, UCAS Report, HESA Track Report)
- Delivery context
- Descriptors of Activity (resources, inputs, outputs)
- Measures of extent of engagement by participants
- The type of evaluation (eg impact or process evaluation; post/pre-post/control group, type of data)

[2] Note that HEAT data can also be linked to UCAS data through fuzzy matching, but this generates aggregate data on outcomes of participants, not individual-level data.

Evaluation

The HEAT service enables and enhances impact evaluation at various levels of standard of evaluation practice. Level 1 narrative discussion can be enhanced through reference to reports based on the HEAT tracking service.

Before-and-after evaluation (at a basic Level 2 standard) is enabled through, for example, the capacity to monitor the HE applications/outcomes of participants within the HEAT tracking service or through fuzzy-matching to HESA data and compare outcomes to pre-intervention assessments/expectations of participants/teachers. Similarly, before-and-after evaluation based on post-intervention attainment relative to pre-intervention assessment is possible through fuzzy-matching to NPD data. Difference-in-difference can be achieved through comparison of outcomes of participants with those of a comparator group.

Impact Evaluation within HEAT is being enhanced through ongoing developments, increasing the sector's understanding of the complexity of outreach participants from diverse backgrounds, differing levels of attainment at KS2 and KS4 and differing levels of outreach engagement. Within the developing HEAT evaluation planning tool, subscribing HEIs will be able to obtain an Evaluation Report built around data entered regarding the characteristics of the outreach activity and specified outcomes. The data entry is within a rich set of pre-specified fields and protocols, common across HEAT users, thereby potentially, enabling comparisons over time and across activities and institutions.

The HEAT tracking service has demonstrated scalability through its capacity to grow significantly the number of subscribing institutions.[3] Being based on the principle of the sharing of data across institutions, its capacity to enable tracking of outreach participants is enhanced as membership grows. Were it to cover all institutions, then tracking would be potentially comprehensive across the sector – and evaluation would be meaningfully comparable through a common protocol-based data entry system. Importantly, by sharing data across members, HEAT can enable a student led analysis, exploring the relationship between their outreach engagement (across more than one HEI) and attainment and progression outcomes. Extending membership to third sector organisations may also increase the completeness of a participant's outreach experience. By working with a combined dataset across the membership, HEAT analysis allows for a cross-institution interrogation of data for impact evidence and more opportunity for intersectional analysis. This would not be possible for many members independently, due to smaller populations, or to the fact that they work with a specific WP population, or attainment group.

[3] Membership is 69 at the time of writing.

Appendix 2B: UCAS (STROBE)

Background

From the UCAS website:

"STROBE is a UCAS service that can track individuals into the UCAS applications system, and report anonymously on their outcomes or characteristics at aggregate levels.

STROBE is currently operated by the UCAS charity as a sustainable service, priced to cover its operational and development costs. The basic pricing for STROBE is on a per record submitted basis, at a base of £5 per record submitted. This charge is reduced if only a low proportion of records are linked into the UCAS data. STROBE users who are organisational customers of the UCAS Undergraduate scheme (higher education providers and apply centres) have a separate pricing arrangement. The minimum charge for an initial STROBE execution will be £400. For potential users who have many data records but require only summary information, or a link into the EXACT data service, a simplified version of STROBE (with less configurability) is available at a fixed cost of £2,000.

For analysis of recognised widening participation and fair access activities where the pricing of STROBE is preventing access, we have a public benefit programme. Please contact us for a discussion of how this works."

Tracking and monitoring

HEIs wishing to access data on HE applications of their outreach participants can request various aggregated output measures. The STROBE service takes individual level data on an HEI's outreach participants (such as name, date of birth, home postcode) and links these to individuals in the UCAS records to calculate anonymous aggregate statistics on the group of participants.

A STROBE Report can provide statistics on a range of application outcomes, covering the entire application process from application through to offer making and replies, and finally to acceptance into HE. It can also give breakdowns by characteristics such as high/medium/low tariff of universities applied to.

Consent

From the UCAS website:

"To use STROBE, you will need to have named personal data, for which you have the correct authority and permission to submit into the service for processing and receive anonymous statistical reports back. No personal information is ever disclosed through the STROBE service. Personal data transferred into STROBE for processing is destroyed at the end of the process, or after a specified period of time if the user has requested that it be retained for an updated analysis at a later point."

As tables provided in STROBE Reports are at an aggregate level of information with built-in disclosure control, users can share the results without restriction or risk of disclosing personal information.

Comparator groups and benchmarking

As all the STROBE-generated outputs and definitions are to the same standards and definitions as national UCAS reporting, they can be directly compared to them for reference. Various 'comparator groups' can therefore be used as benchmarks for comparison in STROBE Reports: for example, either the full UCAS database of applications in the cohort or a 'potential applicant database cohort' similar to the UCAS-linked outreach participants in terms of age, socio-demographics and nearest-school FSM eligibility.

Implementation

A STROBE output report to an HEI on anonymous aggregate statistics of fuzzy-matched outreach participants consists of: core tracking information, summary charts giving an overview of key statistics, and four standard tables, each focusing on a specific part of the application journey.

- **Applications** – the number of applicants and applications, including overall application rate, and the route through which applications were made.
- **Offers** – the number of offers received and the proportion of the cohort receiving an offer, split by whether the offers are conditional or unconditional.
- **Replies** – the number of offers replied to and whether they were firm or insurance.
- **Acceptances** – the number of people accepted, the acceptance rate, and the route through which acceptance was gained.

In addition, the STROBE user can:

- Specify a set of higher education providers for which outcomes are reported (for example, mission groups or individual providers)
- Request separate reports for different groups of individuals (for example, by gender and by outreach activity)

Evaluation

A STROBE Report can be the basis for impact evaluation of outreach activities. The aggregate statistics on the (fuzzy) matched participants presented in the report tables can be used in a before-and-after evaluation (at a basic Level 2 standard) by comparing the post-intervention STROBE-generated aggregate statistics for the treatment group with pre-intervention assessments or expectations of participants/teachers.

A difference-in-difference type of evaluation (up to a Level 3 standard) can be achieved through comparison of the aggregate outcomes of participants with those of a comparator group, such as the UCAS applications database cohort or the potential applicant database cohort. In the STROBE Report, the cohort of 'treated' students submitted by the HEI to UCAS is compared to many cohorts of similar size and composition and the relative outcomes of the treated cohort are presented as a percentile of the overall sample. The percentile recorded is presented with an indication of its statistical significance.

Appendix 2C: NCOP

Background

From the NCOP website:

"The national collaborative outreach programme aims to support the most disadvantaged young people in England to progress into higher education (HE). It will run from 2016-17 to 2019-20. The programme consists of 29 consortia undertaking outreach activity in geographical areas where the HE participation of young people is both low and much lower than expected based on GCSE-level attainment.

HEFCE will provide £30 million in 2016-17 to establish the consortia and start outreach activity. From 2017-18, funding for the programme will be set at £60 million per annum. Funding will be provided for two years in the first instance, from January 2017 to December 2018. Funding for a further two years to December 2020 will be subject to consortia making satisfactory progress towards meeting the Government's goals."

Tracking and monitoring

Recording of activities and of participants

Consortia maintain records of participant engagement in outreach activities across the collaborating institutions.

Tracking of participants

Consortia are required to track their individual target learners through the Higher Education Access Tracker (HEAT) or a similar mechanism.

Evaluation

Evaluation will take place at both local (consortium) level and nationally. At the national level, a four-year impact evaluation will be conducted by CFE in partnership with The Behavioural Insights Team and economists from the University of Sheffield and LSE in order to understand 'what works, in what context and why.'

The evaluation strategy will have many aspects and will be based on a mixed method approach involving primary research, surveys, interviews, case studies, and analysis of HEFCE monitoring data. An outline of the key characteristics of the evaluation strategy for NCOP can be found at: http://www.hefce.ac.uk/media/HEFCE,2014/Content/Student,access,and,success/NCOP/Evaluation_table.pdf