



Leading education
and social research
Institute of Education
University of London



CAYT

Centre for Analysis of Youth Transitions

Evaluating the impact of youth programmes

Claire Crawford and Anna Vignoles

June 2013

CAYT Repository

- CAYT is a DfE sponsored research centre
- Reliable and independently validated information on the effectiveness of programmes
- Repository of reviews with judgements on quality
 - <http://www.ifs.org.uk/centres/caytRepository>
 - http://www.ifs.org.uk/caytpubs/types_impact_study.pdf
- CAYT can offer individual support for organisations who would like advice on evaluation.

Programme for today

- Principles of evaluation
- Sources of useful data
- A case study
- How to access DfE data?
- Discussion

PRINCIPLES OF EVALUATION

Why evaluate?

- Good-quality evaluations generate reliable results which can be used and quoted with confidence.
- Evaluation enables programmes to be improved and justifies reinvestment.
- Shows whether or not resources are being used effectively.

Source: Magenta book

What is evaluation?

- Evaluation seeks to understand:
 - how a programme was implemented
 - what effects it had, and for whom
 - how and why.
- The earlier evaluation is considered, the more likely a high quality evaluation can be undertaken

What sort of evaluation?

- A wide range of factors need to be considered:
 - Expected impacts and outcomes
 - Timing
 - Resources available for evaluation

What type of evaluation?

- Process evaluation

- *How* a programme is implemented; whether it is run properly; how it is perceived by participants and deliverers. TELLS YOU NOTHING ABOUT IMPACT.

- Impact evaluation (short run)

- Impact of programme on measurable output

- Outcome evaluation (long run)

- Impact of programme on outcomes – generally longer term.

How to evaluate?

- Methodologies for process evaluation
 - Surveys of participants, in depth interviews/ focus groups, case studies
 - Qualitative or quantitative research on implementation process (e.g. what was spent, what was delivered etc.)
- Methodologies for impact evaluation
 - Quantitative/ statistical

Principles of research design

- Correlation ≠ Causation
 - Unemployed people are more likely to have poor mental health. Young unemployed people are more likely to engage in anti social behaviour.
 - Does not *necessarily* mean that unemployment *causes* poor mental health or anti social behaviour.
- Key issue in any impact evaluation is to establish a **causal** relationship.

Impact Evaluation Problem

- The fundamental problem is the estimation of **counterfactual events**
 - **What would have happened to participants had they not been exposed to the program?**
- Comparing outcomes of participants with outcomes of a counterfactual or comparison group allows us to estimate the causal impact of a policy
 - Different impact evaluation methods differ because of the way they estimate this missing counterfactual

Principles of research design

- Counterfactual requires
 - Collecting data on participants
 - Collecting data on a comparison group of non-participants
- Comparison or counterfactual group needs to be *very similar* to the participant group
 - Should be same as those receiving the programme in all factors that help to explain both programme participation and outcomes of interest

Principles of research design

- Examples of good comparison groups:
 - If there is a waiting list for your programme and you allocate places randomly (or using criteria not related to outcomes of interest), could use those on the waiting list
 - If your programme is only available to those living in certain areas, could use those in other similar areas
- Examples of less good comparison groups:
 - If your programme is designed to help women with children into work, comparing them to women without children is unlikely to be helpful, as their circumstances differ greatly

Foundation level: Score 0

- Studies that describe the intervention and collect data on activity associated with it.
- To improve:
 - Collect some “before and after” outcome data for those receiving the intervention.
 - Collect some “after” data for the group receiving your services and compare these outcomes with those of comparable young people using other sources of data.

Descriptive, anecdotal, expert opinion: Score 1

- Studies that ask respondents or experts about whether the intervention works.
- To improve:
 - Collect some “before and after” outcome data for those receiving the intervention.
 - Collect some “after” data for the group receiving your services and compare these outcomes with those of comparable young people using other sources of data.

Statistical correlation: Score 2

- A statistical relationship (correlation) is established between whether someone receives the programme and their subsequent outcomes.
- Outcomes of those who receive the intervention are compared with those who do not get it.
- To improve:
 - Does not allow for those in the programme being different from those not in the programme
 - Collect some before and after data.

Before and after study: Score 3

- Compares outcomes for programme participants before and after an intervention.
- To improve:
 - If you have before-after data you can measure the change in a particular outcome after the programme.
 - If possible compare change in outcome for programme participants with change in outcome for similar group of young people who did not receive the programme
 - Could use administrative data to generate information for a control group (more on this later . . .)

Control group: Score 4

- Before and after evaluation strategy and a clear comparison between groups who do and do not receive the services or programme.
- You have most of the data you need. Contact an expert and they will be able to apply statistical methodologies to improve the robustness of your results, e.g. using matching methods.

Control group and quasi experimental model: Score 5

- Before and after evaluation strategy, statistically generated control groups and statistical modelling of outcomes.
- Short of a randomised control trial (RCT), this methodology is the most robust.
- To improve:
 - Ensure comparison group is as similar as possible to those receiving the programme.

Randomised Control Trial: Score 6

- Random assignment to the programme.
- The gold standard. It is challenging to run a RCT, with cost, ethical and practical issues arising.
- Even with a RCT you have to think about how generalisable it is to other situations
 - e.g. if the RCT only applied to males, it cannot tell you how well the programme would work for females.

The importance of timing

- Evaluation methods available depend on whether you are thinking about evaluating a programme that has already started (or finished) or one that is yet to start
- Wider range of options available to you if your programme has yet to start
 - e.g. could consider running an RCT or collecting “before” data on outcomes not available in administrative data

Nearly there.....

- Impact evaluation is not sufficient to tell you whether to invest in a particular programme
- Cost-benefit analysis (CBA) needed for that
 - CBA estimates the value of the benefits arising from your programme against its costs
 - Benefits valued using other studies and data sources
 - Needs good impact evaluation
- Impact evaluation not sufficient to tell you why a programme worked (or did not work)
 - Process evaluation required for that

Key references on evaluation

- Magenta book http://www.hm-treasury.gov.uk/data_magentabook_index.htm
- ***The Green Book***, GSR (Government Social Research). Publishing Research in Government. January 2010. HM Treasury http://www.hm-treasury.gov.uk/data_greenbook_index.htm
- Baker, J. L. (Ed) (2000) *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Washington D.C: World Bank.

SOURCES OF DATA

Data sources

- Administrative data to save costs and avoid problems with primary data collection
 - E.g. Using education data on GCSE achievement of your participants rather than trying to undertake your own survey
- Administrative data for long run follow up
 - E.g. Using education data on HE participation as a way of measuring longer run outcome from a programme.

Data sources

- Creating comparison or “control” groups using additional data sets
 - Programme designed to improve employment.
 - Data on participants’ employment rates before and after the programme.
 - Compare to similar group of young people using Labour Force Survey during the same time period.
 - Programme designed to improve school attendance.
 - Before and after data on participants’ absence rates.
 - Compare to similar group of students using administrative education data.
 - Can be done after the event.

Useful administrative data – can often be linked

- National Pupil Database and School Census
 - Achievement, absence, SEN
- Higher Education Statistics Agency
 - HE participation, type of HE,
 - Destination of Leavers from Higher Education (DHLE) – earnings
- Individual Learner Record
- DWP benefits data
- Ministry of Justice data on offending
- Health data including mental health

<http://www.adls.ac.uk/find-administrative-data/research-themes/>

Useful survey data

- Longitudinal Study of Young People in England
 - Follow up to cohort 1, new cohort going into the field
 - <http://discover.ukdataservice.ac.uk/catalogue/?sn=5545&type=Data%20catalogue>
- Millennium Cohort Study
 - Young but useful soon
 - <http://www.esds.ac.uk/longitudinal/access/mcs/l33359.asp>
- Labour Force Survey
 - Transitions into the labour market and employment
 - <http://www.esds.ac.uk/findingData/qlfs.asp>
- Understanding Society
 - <https://www.understandingsociety.ac.uk/>

CASE STUDY

Introduction

- Social Mobility Foundation (SMF):
 - Aims to support high achieving young people from low income backgrounds into top universities and professions
- Runs a variety of residential and mentoring programmes for final year A-level students who:
 - Are entitled to free school meals (FSM), or;
 - Attend a school where at least 30% of pupils are FSM eligible and are the first in their family to go to university

SMF programmes

- Aspiring Professionals Programme:
 - Offers mentoring, university visits, application support, and internships; support continues whilst pupils are at university
- J.P. Morgan Residential Programme:
 - Two weeks work experience plus workshops on softer skills and social activities for young people outside London

Aims of evaluation

- Qualitative/process evaluation:
 - To find out about the experiences of young people and their mentors: whether they found the programmes helpful; which bits they found most useful; suggestions for improvement
- Impact evaluation:
 - Identify the **causal impact** of each programme on education and employment outcomes
 - Will focus on this today

Reminder of the best available evaluation options

- Randomised control trial:
 - Amongst eligible participants, randomly allocate some to the “treatment group” and some to the “control group”
 - Not possible for SMF programmes because decision to evaluate was taken after the programmes started
- Quasi-experimental methods:
 - Compare outcomes of programme participants with outcomes of an *otherwise identical* control group

Choosing a suitable control group

- Aim is to pick a group who are as similar as possible to young people who participated in your programme
- Best option often to focus on young people who were prevented from participating for some reason
 - e.g. not enough places, lived in an area that you didn't serve
- Important because this helps to ensure that they have similar “unobservable characteristics” to participants
- Less desirable to use as a control group young people who were offered your services but chose not to use them

Choosing a control group for the SMF programmes

- Open to all young people who meet criteria and no capacity constraints in early years of implementation
- But programmes still relatively small
- Use as a control group all young people who are eligible for the programmes but do not participate

Is choosing a suitable control group enough?

- Unlike RCTs, quasi-experimental methods may not produce identical treatment and control groups
 - Collect detailed info on young people in both groups and use statistical methods to ensure you compare like with like
 - Reliability of results depends heavily on how much you know about young people before they join your programme
 - Of particular importance are characteristics that affect programme participation and/or outcomes of interest

What data do we need for a robust evaluation?

- Information to identify your chosen control group
 - For SMF evaluation, whether young people are (or would have been) eligible for the programmes
- As much information as possible about characteristics that might determine whether young people participate in the programme and their outcomes of interest
 - e.g. prior attainment, education/career aspirations, etc
- Outcomes of interest

Challenges for SMF impact evaluations

- Decision to evaluate was taken after programmes started
 - No data collected from potential control group
 - Limited “before” data collected from participants
- Means we will need to fill in some gaps:
 - (Vital) Before and after data for a suitable control group ***from the same dataset***
 - (Nice but not necessary) More “before” data on participants

What do we know about SMF participants?

- Reasonably detailed pre-programme information:
 - Gender, ethnicity, month of birth, where they live, schools attended for GCSEs and A-levels, A-level subjects/grades
- Some outcomes already collected:
 - Which universities applied to (and to study which subject), where received offers from, which offer accepted
- And plans for others to be collected in future:
 - Whether finished degree (and where/which subject), degree class, whether in work some time later (and in a profession)

Datasets from which to construct a control group

- Need data which contains as much of this info as possible
- Considered a number of options, but best bet seemed to be linked NPD-HESA-DLHE data (if access is granted; otherwise NPD-HESA and HESA-DLHE data separately)
 - Seemed to provide greatest overlap with information known about participants and included outcomes of interest to SMF

What is included in linked NPD-HESA-DLHE data?

- Vast majority of info required to determine eligibility
 - The NPD contains information on:
 - Whether a pupil is eligible for FSM
 - % of pupils in a school who are eligible for FSM
 - HESA data contains information on:
 - Whether a young person's parents have HE qualifications
 - So linked NPD-HESA data not perfect (FSM eligibility not entitlement and no family history of HE) but pretty good
 - And can check how good in other data sources:
 - e.g. using LSYPE data, information available in NPD-HESA data accurately predicts eligibility in 97% of cases

What is included in linked NPD-HESA-DLHE data?

- Identical background characteristics (mainly from NPD)
- Possibility of observing most outcomes of interest to SMF:
 - HESA data covers:
 - Whether went to university (and if so where and which subject)
 - Whether completed degree and if so which degree class
 - DLHE data covers:
 - Outcomes 6 months (and for a subset 2.5 years) after graduation, e.g. whether studying post-graduate qualifications or working (and if so in which occupation and salary band)

Summary for SMF evaluations

- Before and after data on participants collected via online surveys by SMF
- Control group of young people who are eligible for the programmes but did not take them up
 - Equivalent information for these individuals available from linked NPD-HESA-DLHE data
- Use statistical modelling to ensure treatment and control groups are as similar as possible
 - And hence that comparing their outcomes identifies the causal impact of the SMF programmes

Summary of key points

- Choose an appropriate control group:
 - For a programme that has not yet started, consider an RCT
 - Otherwise think about suitable alternatives
 - Perhaps young people who were prevented from participating?
- Collect as much pre-programme information as possible
- Make sure you observe your key outcomes of interest
 - Info needs to be identical for treatment and control groups
- Apply statistical models and check results are robust