TASO Transforming Access and Student Outcomes in Higher Education

Project report:

Learning about evaluation with small cohorts

Dr Julian Crockford, Sheffield Hallam University
Rain Sherlock, TASO

December 2023

CONTENTS

EXECUTIVE SUMMARY

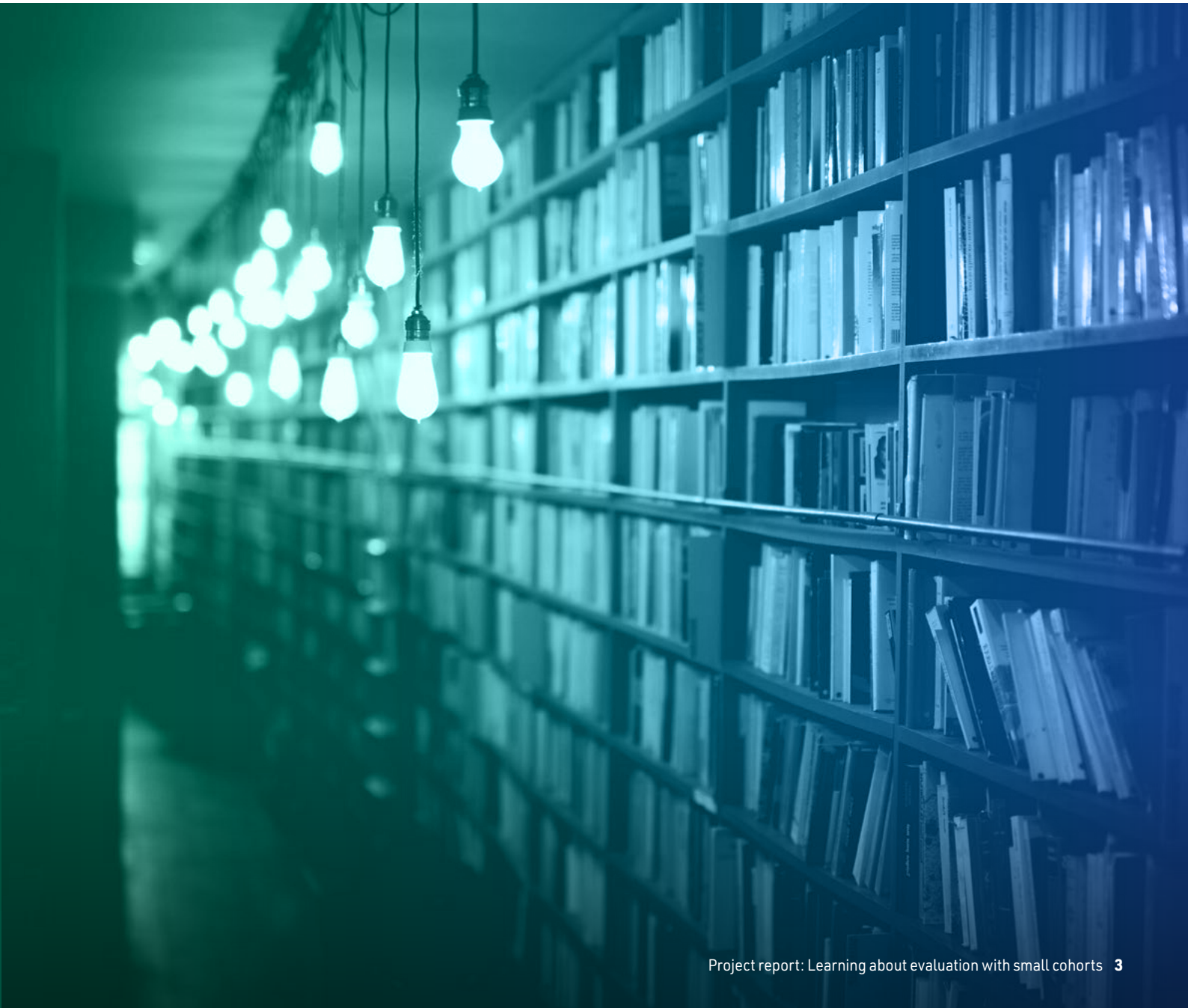
Introduction

The *Impact evaluation with small cohorts* project was intended to help higher education providers (HEPs) overcome methodological and theoretical challenges in evaluating the impact of interventions with small cohorts of participants, where traditional large-scale quantitative evaluation methods are not suitable.

An initial consultation phase revealed that respondents from across the sector face a range of challenges when evaluating small-cohort interventions. As well as the primary challenge – that small cohorts represent inherently small sample sizes, for which large-scale quantitative methods are unsuitable – respondents reported a range of challenges, including low response rates to evaluation instruments and difficulties in isolating the influence of external factors and in identifying and recruiting target groups where they were not included in institutional datasets. Furthermore, respondents reported uncertainty about how to define the term 'small' in the context of cohort size and how appropriate evaluation methods would align with other regulatory evaluation requirements.

The project advisory team, from Manchester Metropolitan University's Policy Evaluation and Research Unit (PERU) developed guidance on the use of eight evaluation approaches suitable for use in small-cohort contexts. This guidance was supplemented by hypothetical case studies showing how each method might be applied in a higher education (HE) evaluation context.

Six project teams representing different HEPs tested four evaluation methodologies: Realist Evaluation, Contribution Analysis, Most Significant Change and Qualitative Comparative Analysis.



Summary of project team reflections

The pilot project teams identified a series of challenges that arose during the implementation of these methodologies:

- The need to navigate complex terminology and concepts associated with new and unfamiliar methodologies.
- A range of routine challenges commonly associated with evaluation in this space, including engaging participants, data availability, selecting appropriate and robust evaluation measures, reliance on self-reported data and difficulty isolating the impact of the target intervention from external factors.
- The high level of resources and relatively long time frame demanded by these evaluation approaches for effective implementation.
- The need to develop specialist knowledge and experience to implement small-cohort methodologies effectively.

The pilot teams also identified a variety of benefits produced by small-cohort evaluation approaches:

- Increased knowledge was gained about the nature and functioning of the target programme, the change mechanisms associated with certain outcomes, the key constructs and concepts underpinning the intervention design, and the target participant groups.
- Although none of the methodologies produced Type 3 (causal) impact evidence, as currently defined by the Office for Students *Standards of Evidence* guidance, all methodologies generated valuable evidence that increased the evaluators' confidence in impact claims and their understanding of the relationship between the activity and the outcome.

The following lessons were learned from pilot projects:

- Small-cohort evaluation approaches illuminate the complex nature of many programmes and interventions delivered in this space.
- The focus on theory-driven evaluation, and the need to understand intervention change mechanisms, is valuable and could positively impact both intervention and programme design and support the development of a knowledge base of what works.
- There is potential value in combining different small-cohort and traditional large-scale quantitative or counterfactual impact-evaluation methodologies to produce a broader and richer range of evaluation outcomes.
- A quality assurance framework and reporting guidance are required to support the consistent and effective implementation and reporting of small-cohort evaluations.

Recommendations

Recommendation 1: The pilot projects engaged with four small-cohort methodologies. Pilots of small-cohort methodologies not included in this project (Process Tracing, General Elimination Methodology and Comparative Case Study) should be encouraged and supported.

Recommendation 2: Given the challenges experienced by the project teams in adapting and implementing small-cohort evaluation methodologies, sufficient consideration should be given at the planning stage to the significant time and resources required to engage effectively with unfamiliar evaluation methodologies.

Recommendation 3: Given the currently limited availability of HE-specific examples, small-cohort evaluation methodologies conducted in other disciplinary areas should be identified and signposted to evaluators working in the HE sector. To make these resources more accessible, they should be translated into terms or contexts relevant to this sector.

Recommendation 4: Conducting high-quality, robust evaluation is time-consuming and resource-intensive. HEPs should invest in further evaluation capacity to facilitate stronger evaluation practice across their Access and Participation Plan portfolio to ensure they are providing students with the best possible support.

Recommendation 5: Working in isolation on complex, technical or unfamiliar evaluation methodologies brings a risk of incorrect applications, flawed conclusions and stalled projects. HEPs should seek opportunities for peer support, in which practitioners and evaluators with experience in a particular approach can support or advise less experienced colleagues.

Recommendation 6: Evaluators across the sector are generating valuable knowledge about change mechanisms, target groups and evaluation implementation. The use of small-cohort methodologies is likely to increase this knowledge.

A central repository of emerging sector knowledge and evaluations should be considered.

Recommendation 7: The design and implementation of robust, effective and valid evaluation measures are integral to good evaluation practice. Irrespective of the methodology used, evaluation practitioners should ensure that sufficient attention and consideration are given to identifying appropriate measures.

Recommendation 8: The outcomes of the pilot projects suggest that small-cohort evaluation approaches may be productively combined with Type 3 impact-evidence methodologies. Further research or pilot studies should be conducted to assess the potential of combining approaches in this way.

Recommendation 9: There is currently no suitable quality-assurance framework to guide future work in this area. A formal quality-assurance framework and reporting guidance should be developed to improve the rigour of future small-cohort evaluation projects.

SECTION 1 - INTRODUCTION AND BACKGROUND

The Centre for Transforming Access and Student Outcomes in Higher Education (TASO) aims to improve lives through evidence-based practice in higher education (HE). Our goal is to eliminate equality gaps for disadvantaged and under-represented groups, enabling all students to have the same opportunity to enter HE, be awarded a good degree and progress into further study or employment.

Background and project context

The *Impact evaluation with small cohorts* project aimed to address the methodological and theoretical challenges encountered by HEPs who want to evaluate the impact of widening-participation and student-success activities with small participant cohorts. The project includes both pre-entry (widening participation) and post-entry (student success) activities.

At the heart of impact evaluation is the need to establish a connection between cause and effect in order to explain how and why activities lead to changes in desired outcomes. While experimental and quasi-experimental evaluation methods[1] can provide the sector with the confidence required to state causal inference,[2] it is not always possible to use these methods. Small and specialist providers, in particular, face difficulties when attempting to generate causal impact evidence, and even in larger institutions it can be difficult to use existing evaluation methods with smaller cohorts of students, such as those receiving highly targeted support or involved in specialist projects.

Some of the challenges faced in utilising experimental and quasi-experimental evaluation methods are:

- Small sample sizes: Some specialist providers, such as music schools, may only reach a small number of students and young people. Equally, larger HEPs may wish to target interventions at a small group of students, such as those from the Gypsy, Roma and Traveller community.
- Cost of increasing the sample size: Not all HEPs have the budget to conduct interventions on a large scale, which can limit the production of causal evidence to a small number of institutions. This may result in a lack of representation across the wider sector.
- Resourcing: In addition to the cost of increasing sample sizes, some HEPs may have small teams who both deliver widening-participation activities and evaluate them. In this scenario, it is likely to be very challenging to adopt experimental and quasi-experimental methods.
- Complex and multi-intervention programmes: HEPs often deliver interventions that target multiple outcomes and take a holistic approach to supporting students and young people. It is therefore important to have a range of impact evaluation methods on which we can draw to ensure that we capture programme complexity while establishing what works.

The HE sector needs to adopt a wider range of impact-evaluation methodologies that are practical enough to be used by a diverse audience, including small and specialised providers, while maintaining the rigour to give providers the confidence to make bold statements about what works. This project aimed to develop and test a range of evaluation designs that could be used to assess the impact of small-group interventions.

The project was divided into three phases:

1. Consultation: we engaged with those in the HE sector to understand the challenges they face and how they view potential impact evaluation methods in this context.
2. Methodology: we developed general guidance about small-cohort evaluation and specific guidance about implementing small-cohort evaluation methods in this context. Along with this guidance, we include case studies that describe pilot projects. In cases where a particular methodology has not yet been implemented in practice, we provide hypothetical case studies.
3. Translation and testing: we piloted small-cohort evaluation methodologies with a range of individual HEPs, including providers with small student populations and larger providers working with small student cohorts.

[1] Experimental and quasi-experimental research designs test causal hypotheses. In both experimental (i.e. randomised controlled trials) and quasi-experimental designs, the programme or policy is viewed as an 'intervention' in which a treatment – comprising the elements of the programme/policy under evaluation – is tested for how well it achieves its objectives, as measured against a prespecified set of indicators. See the TASO website for evaluation webinars on experimental and quasi-experimental research designs.

[2] Causal inference is key to impact evaluation. An impact evaluation should allow the evaluator to judge whether the intervention under evaluation caused the outcome being measured. Causation differs from correlation. If you are unsure of this difference, please watch this TASO causality webinar.

Consultation phase

The consultation process involved online surveys with four larger HEPs, 12 small specialist HEPs, four sector organisations, three What Works Centres, and six members of the TASO Evaluation Advisory Group.

Additional consultation was conducted with the National Educational Opportunities Network (NEON) Evaluation and Monitoring Group, the Association of Colleges, and the Mixed Economy Group.

Larger HE providers described using 'mixed methods' approaches to evaluate small- cohort interventions. They identified several challenges in this approach, including low response rates to evaluation instruments, the difficulty of disentangling the specific impact of target interventions from other activities in which participants might be involved, and difficulties in identifying and recruiting target groups where these are not included in institutional datasets.

Small and specialist providers reported challenges in accessing services such as the Higher Education Access Tracking service (HEAT) that would enable them to track longitudinal impacts, alongside the limitations posed by restrictions on the numbers of staff and other resources available for evaluation, and limited staff expertise in evaluation.

Similarly, relevant sector organisations also had limited evaluation skills and resources. They often focus on a single cohort of beneficiaries, such as care-experienced learners, but note potential diversity even within small cohort groups.

Evaluation practitioners reported some confusion around the definition of 'small cohorts' and what qualifies as 'small' in this context. They also expressed concerns about whether the outcomes of these types of evaluation would meet the regulatory requirements of the Office for Students, especially in terms of evaluation robustness and the need to produce Type 3 (causal) evaluation evidence. Furthermore, providers were unsure about how to integrate these forms of evaluation within their Access and Participation Plan intervention strategies.

Methodology phase

During the methodology phase, the PERU at Manchester Metropolitan University developed a guidance report, Impact Evaluation with Small Cohorts: Methodology Guidance.[3] Given the current lack of implementation case studies relevant to the HE sector, the evaluation advisory team developed hypothetical case studies to show how the evaluation methodologies might be used in this context. These interim case studies will be replaced by real case studies as the methodologies are implemented and documented.

Translation and testing phase

The translation and testing phase was designed to assess the value of these evaluation methodologies and approaches in evaluating HE programmes across the access and student-success stages of the student lifecycle.

Six pilot-project teams participated in this phase of the project. These teams were from two institutions with a limited HE offer in the context of Further Education (City College Norwich and University Centre Leeds); one small specialist provider (Leeds Arts University); one small HEP (Plymouth Marjon University); and two larger HEPs (University of Leeds, Lifelong Learning Centre and University of Suffolk).

The *Methodological Guidance* provided detailed information on eight evaluation methodologies suitable for programmes with small cohorts:

- **Realist Evaluation**
- Process Tracing
- General Elimination Methodology
- **Contribution Analysis**
- **Most Significant Change**
- **Qualitative Comparative Analysis**
- Comparative Case Study
- Agent-based Modelling

Agent-based modelling was considered potentially too technical for pilot implementation in this context; of the remaining seven methodologies, four were selected for a pilot (marked in bold above).

To date, the three additional methodologies – Process Tracing, General Elimination Methodology and Comparative Case Study – have not been tested and their usefulness and relevance to the HE sector are unknown. The piloting of the remaining small-cohort methodologies should be encouraged and supported. (Recommendation 1).

[3] For the remainder of the report, this will be referred to as the *Methodological Guidance*.

PILOT PROJECT OUTLINE

The pilot project was delivered between July 2022 and May 2023. See Table 1:

Table 1: Outline of pilot project stages

Project Stage	Timescale	Activity
Project set up	July – August 2022	The project teams confirmed their project plans and obtained research ethics approval where applicable.
Methodological workshops	August – September 2022	A series of workshops provided opportunities to delve deeper into selected methodologies and foster collaboration between the teams.
Drafting initial Theories of Change and research protocols	September – November 2022	The project teams developed initial Theories of Change for their programmes and interventions. These were reviewed and feedback was provided by the evaluation advisory team.
Evaluation implementation	November 2022 – February 2023	The teams recruited evaluation participants and implemented evaluation activities.
Analysing data and redrafting Theories of Change	February – March 2023	Teams continued to implement evaluation and conducted data analysis. Where required by the methodology, programme Theories of Change were updated.
Write up and reporting	April – May 2023	Project teams wrote up the evaluation outcomes, prepared a case study and prepared a reflective report.

Throughout the project, the teams received support from the evaluation advisory team, comprising a TASO project manager, evaluation methodology advisers including Professor Chris Fox and colleagues from the PERU at Manchester Metropolitan University, and Dr Julian Crockford, Student Engagement Evaluator and Researcher at Sheffield Hallam University. This team provided support through scheduled and ad hoc meetings, workshops, document reviews and guidance whenever needed.

The *Methodological Guidance* outlines the evaluation context for, and procedural information on, eight different methodologies appropriate for evaluating programmes with small participant cohorts. Pilot project teams were invited to select the most appropriate methodology based on their context, needs and target intervention. Some teams also referred to the selection guidance provided by an online resource developed by Befani (2020)[4] and recommended in the *Methodological Guidance*. A summary of each team's reasons for selecting a particular methodology is provided in Appendix 1.

The final project outputs, such as local pilot reports and case studies, can be found on the TASO website.

- Impact Evaluation with Small Cohorts: Methodology Guidance
- City College Norwich, Pilot Project Report and Case Study
- Leeds Arts University, Pilot Project Report and Case Study
- University Centre Leeds, Pilot Project Report and Case Study
- University of Suffolk, Pilot Project Report and Case Study
- Plymouth Marjon University, Pilot Project Report and Case Study
- University of Leeds, Lifelong Learning Centre Pilot Project Report and Case Study

[4] The full tool is available at https://www.cecan.ac.uk/news/choosing-appropriate-evaluation-methods-a-tool-for-assessment-and-selection-version-two/

Pilot projects

Table 2 below outlines the focus groups of the six pilot evaluation projects.

Table 2: Breakdown of pilot partners, target intervention, target group and methodologies tested

Pilot partner	Intervention	Student lifecycle stage	Target group	Evaluation methodology	Summary of methodology
City College Norwich	Higher Education Tutorial Supervisor	Student Success (Academic)	Students with self-declared learning disabilities/difficulties	Contribution Analysis (CA)	Contribution Analysis involves creating a Theory of Change, mapping change mechanisms and then testing, challenging, and refining them with various stakeholder groups.
Leeds Arts University	Creative Pathways Programme – Creative Arts Outreach Programme	Access	16–18, students from groups under-represented in HE	Contribution Analysis (CA)	
University Centre Leeds	Gypsy, Roma and Traveller outreach programme	Access	16–18 Gypsy, Roma and Traveller young women in the community	Realist Evaluation (RE)	Realist Evaluation is founded on the philosophical position that the direct observation of causal effects is not possible, but outcomes can be understood by testing them in the context of theorised context and mechanism models.
University of Suffolk	Student Micro-Placements	Student Success (Career Progression)	Current students from disadvantaged groups or under-represented groups, including mature students or those who may not be able to access traditional placement opportunities	Realist Evaluation (RE)	
Plymouth Marjon University	Student Colleagues Scheme – a careers development programme	Student Success (Career Progression)	Current students meeting Office for Students criteria for under-representation and/or disadvantage (POLAR 4 and disability)	Transformative Evaluation / Most Significant Change (TE/MSC)	Transformative Evaluation involves collecting and selecting participant stories in partnership with intervention stakeholders. It is a variant of the Most Significant Change methodology.
University of Leeds, Lifelong Learning Centre	Jumpstart – An outreach programme for adult learners	Access	Adult learners from low-participation neighbourhoods and underperforming schools	Qualitative Comparative Analysis (QCA)	Qualitative Comparative Analysis involves devising, calibrating and assessing case attributes against programme outcomes to calculate their contribution and interaction.

Additional reflection on the evaluation outcomes and knowledge produced by each of these evaluation approaches is included in Appendix 2 - Evaluation methodology map.

SECTION 2 – PILOT PROJECT REFLECTIONS AND LESSONS LEARNED

This section summarises the evaluation reports and case studies provided by each pilot team, as well as the outcomes of a reflective workshop held at the end of the project. Section 3 presents the evaluation advisory team's overall reflections.

The first section below focuses on the challenges faced by the project teams in implementing the methodologies examined. The following section reports on the benefits of small-cohort evaluation approaches as identified by the project teams.

Challenges associated with small-cohort evaluation

The purpose of the *Impact Evaluation with Small Cohorts* project was to test the small-cohort evaluation approaches outlined in the *Methodological Guidance* within an HE context by focusing on interventions that occur at various stages of the student lifecycle.

During the pilots, each team encountered a range of practical and implementation challenges while implementing these approaches, including:

- The need to negotiate new and unfamiliar terminology and concepts
- Practical and implementation challenges, including time and resource limitations, and the specific skills required by these methodologies
- Routine evaluation challenges, including in engaging participants, accessing relevant data, deriving robust and effective measures, the need to rely on self-reported data, and isolating the intervention impact from other influences.

Negotiating new and unfamiliar terminology and concepts

Key learning points:

Many of the small-cohort evaluation methodologies do not fit into established evaluation paradigms. Even experienced evaluators who are familiar with traditional approaches may initially find it challenging to adapt to new terminology, concepts and methodological approaches.

The difficulty is exacerbated by the current paucity of sector-relevant case studies describing these methodologies in use.

Most of the small-cohort evaluation approaches included in the *Methodological Guidance* are associated with a different evaluation paradigm from the 'traditional' forms of evaluation typically used across the HE sector, and broader policy evaluation landscape, and with which many evaluators are familiar. This paradigm shift initially challenged some of the pilot teams, who were used to working in a more 'traditional' evaluation mode. Once this learning curve was negotiated, however, many of the teams valued these methodologies as a fresh way of approaching the evaluation process.

Some of the initial challenges stemmed from the use of unfamiliar terminology or concepts in the small-cohort evaluation approaches. Most of the small-cohort methodologies described in the *Methodological Guidance* begin with the formulation of a detailed Theory of Change to map how the programme or intervention 'works' to deliver change. Some pilot-team members found this approach challenging, as it differed from the logic-model framework with which they were familiar. They noted that concepts such as 'change mechanism' were seen as ambiguous or unclear by those unfamiliar with them. At the start of the project, members of the University of Suffolk team, for example, were unsure how the concept of change mechanisms would fit in the context of the complex programme they were evaluating, which comprised multiple elements, stakeholders and strands of work (University of Suffolk report, p.25). An early advisory workshop was dedicated to helping teams develop a programme-relevant understanding of what change mechanisms might look like and how they fit into a broader understanding of how interventions deliver their outcomes.

Many of these evaluation methodologies also rely on specific technical skills or knowledge. The University of Leeds' Lifelong Learning Centre project team observed that the technical nature of Qualitative Comparative Analysis evaluation, which is based on set theory and involves Boolean logic, required them to seek external expertise during the initial stages of the programme implementation, alongside developing their own understanding of the approach (University of Leeds, Lifelong Learning Centre report, p.32).

Across the different projects, each team reported some initial challenges in understanding and engaging with their chosen evaluation approach. In some cases, the process of familiarisation with new terms, concepts and methodologies delayed the data-collection and evaluation-implementation phases. This highlights the importance of allocating sufficient time for understanding and developing familiarity

with a new evaluation approach during project planning (Recommendation 2).

Some project teams also reflected that implementation might have been easier had there been more domain-relevant examples on which to draw. Indeed, the current lack of HE-specific case studies for these methodologies was one driver of this project. For some of the small-cohort methodologies, examples or implementation case studies exist in other disciplinary areas, such as international development, public health or public policy. A better understanding of this work in other disciplines, and some support in recontextualising or translating it into an HE context, could help support HE-based evaluation practitioners in implementing unfamiliar evaluation methodologies in the future (Recommendation 3).

Practical and implementation challenges

Key learning points:

Project teams described several factors which constrained their ability to implement small-cohort evaluation methodologies effectively:

- limitations in time (particularly in the context of the project timeframe)
- limitations in available resources
- limitations in existing skill sets.

There may be opportunities to mitigate some of these limitations through, for example, taking a strategic approach to disseminating evaluation outcomes and findings about specific change mechanisms and target groups. Sharing knowledge in this way could reduce the risk of multiple teams tackling the same issues from scratch rather than building on each other's learning.

Nevertheless, it is important to recognise that small-cohort evaluation methodologies can be time-consuming and resource-intensive and that they demand specific skills and expertise. The pilot projects received a high level of expert support during implementation. Teams without this additional support and guidance may find the process even more challenging.

During the implementation of their pilot projects, project teams faced various practical and logistical challenges. Some of these were specifically related to the nature of small-cohort evaluation, some to the knowledge and experience demanded by these methodologies and others to the complex nature of the target interventions.

Limited timeframe of the pilots

All the project teams observed that the overall project structure required rapid implementation and compressed reporting timeframes, leading the evaluation process to be often conducted in a much shorter timeframe than would normally be the case. Many project teams found this limiting and suggested that it prevented them from fully implementing their chosen methodology. For example, the City College Norwich team, using a Contribution Analysis approach, noted that they would have preferred to conduct multiple phases of iterative review and development of their contribution story, but the limited timeframe restricted them to a single cycle. Similarly, the Plymouth Marjon University team had to reduce the gap between story collection cycles to four weeks, instead of the usual two to three months typical of the Transformative Evaluation approach they employed. This reduced their opportunities to assess change and shifting perspectives across time.

When reviewing the pilot outcomes, it is important to consider the impact of this condensed timeframe. Few teams felt they had delivered their chosen methodology 'by the book'. The resulting compromises may have limited evaluation outcomes and the knowledge and evidence they generated. However, as these methodologies become more widely adopted across the sector, more fully implemented case studies will be produced, and our understanding of their potential contribution will increase.

Nonetheless, despite the short timescale of the project, each pilot team was able to learn, develop and implement their chosen methodologies, and produce valuable outcomes that demonstrated the value of small-cohort approaches across a range of HE settings.

Resourcing

The implementation of the pilot projects was often also constrained by limitations in available resources. Adequate resourcing of evaluation activity has frequently been flagged as a concern for HE-based evaluators (Crawford et al. 2017, Harrison et al. 2015, Harrison and Waller 2017). The *Methodological Guidance* also cautions that gathering in-depth qualitative data for some small-cohort evaluation approaches can be as time-consuming and resource-intensive as data-collection methods in traditional

counterfactual impact evaluations (p.71). This proved to be the case, and accessing sufficient resources emerged as an issue for all project teams, perhaps reflecting broader issues of evaluation resourcing across the sector.

The pilot projects reveal the extent of staff resources required for these methodologies. The City College Norwich team observed that they were able to focus in detail on only one causal chain out of the five included in their Theory of Change (City College Norwich report, p.10). They concluded that 'Contribution Analysis is not a quick evaluation process, but a process that can take months of data collection and analysis, both of which likely need to be repeated and revisited' (p.17). The level of resources required is likely to be particularly problematic for small evaluation teams. Indeed, the University Centre Leeds team concluded that Realist Evaluation placed a significant demand on staff time and resources, and its implementation was likely to pose a real challenge for small teams, such as those in small and/or specialist providers (University Centre Leeds report, p.38).

Many small-cohort evaluation methodologies involve an iterative process of testing and challenging change models and assumptions. As with time limitations, a lack of resources negatively affected the teams' ability to conduct repeated review cycles. Contribution Analysis, for example, includes a phase in which different stakeholder groups review, challenge and enrich a Theory of Change. The Leeds Arts University project team noted that they were unable to challenge their own thinking and assumptions effectively because their small team shared an understanding of how the programme was intended to work and they had neither time nor opportunity to test this against alternative perspectives. In this sense, small teams as well as small participant cohorts can challenge evaluation.

In the reflective workshop at the end of the project, one pilot team noted that receiving project funding had made it possible for them to conduct a more complex and intensive evaluation than would usually be the case. This reinforces the findings of others that evaluation across the HE sector is often under-resourced. The outcomes of the project reinforce recommendations that HEPs should continue to invest in further evaluation capacity to facilitate stronger evaluation practice. (Recommendation 4).

Specific skillsets required by these methodologies

The methodologies associated with small-cohort evaluations often require evaluators to draw on specific evaluation skills or experience. These often extend or differ from the skills and knowledge required for more 'traditional' evaluation approaches and are likely to require additional staff development. For example, the University of Leeds' Lifelong Learning Centre team required external support to develop the specific technical skills required to implement Qualitative Comparative Analysis. The University of Suffolk team already had qualitative research skills but was required to adapt them to the specific demands of Realist Evaluation.

The pilot teams received support and guidance from the project advisory team, which includes representatives from TASO, Manchester Metropolitan University and Sheffield Hallam University. The project outcomes suggest that without this additional support and guidance, evaluators working in isolation could face challenges in understanding and interpreting methodological guidance, especially where this contained specific technical components. This could result in the incorrect application of methodologies, flawed conclusions or stalled projects. A potential mitigation may be for the sector to develop a programme of peer support, in which evaluators with experience of using a particular methodology can guide and support less experienced colleagues (Recommendation 5).

The risk of a flawed evaluation resulting from the use of unfamiliar methods could be exacerbated if evaluation teams are working in isolation, tackling similar problems in silos or failing to capitalise on prior work and learning elsewhere. To mitigate this risk, effective processes should be put in place to enable evaluators to access cumulative learning and knowledge and build on and develop existing work. The theory-driven approach typical of many small-cohort evaluation methods can generate valuable knowledge about change mechanisms, target participant groups and specific evaluation approaches (see below). This knowledge should be collated and made available to other practitioners and evaluators, perhaps through a central repository, to ensure that learning and evidence can contribute to an iterative sector-wide development process. This would enable programme designers and evaluators to build on the work of others and collaboratively develop a sector-wide understanding of what works in specific contexts and for specific beneficiary groups (Recommendation 6).

Routine evaluation challenges

Key learning points:

Although small-cohort evaluation approaches offer an alternative approach to 'traditional' forms of evaluation, they are prone to the same types of challenges presented by other forms of evaluation.

The pilot teams report encountering difficulties such as:

- engaging programme participants in the evaluation process
- obtaining sufficient data about participants to support effective evaluation
- establishing robust and rigorous evaluation measures
- having to rely on self-reported data.

This suggests that while specific small-cohort evaluation approaches can address some of the challenges associated with evaluating small-cohort interventions, they do not offer a solution to all forms of evaluation challenges.

In other cases, however, such as the challenge of isolating target interventions from external influences, many small-cohort evaluation approaches offer an advantage over other methods by building this in as an integral methodological step.

Many project teams noted that in implementing small-cohort evaluations they encountered some of the same kinds of practical and logistical challenges that can impede more 'traditional' evaluation approaches.

Engaging participants

Evaluation across the HE student lifecycle is often hindered by limited participant engagement with the evaluation process, which can result in outcomes such as low rates of response to surveys (Gorard & Smith, 2006, Harrison et al., 2015). Low response rates risk introducing a range of biases, such as non-response bias, into the evaluation process. Given the small size of the participating cohorts, the need to ensure sufficient participant engagement with evaluation can be still more pressing for the piloted methodologies.

The Leeds Arts University team observed that, even though they were confident their creative arts outreach intervention had been successfully implemented, student participation and engagement were variable (Leeds Arts University report, p.18). Low response rates to data-collection activities further limited the conclusions they could draw from their evaluation outcomes. They suggested that some evaluation 'weaknesses result[ed] from participant attendance, survey completion, and access to complete destinations data' (Leeds Arts University report, p.38). The University Centre Leeds team faced similar challenges with inconsistent cohort participation and engagement with evaluation when evaluating their outreach programme for Gypsy, Roma and Traveller young people.

While small-cohort evaluation approaches can partially mitigate the impact of low participant engagement – as can triangulating data collection and individual case studies – they do not directly address the underlying challenge of engaging participants in evaluation activities.

Accessing relevant data

Another widely documented evaluation challenge, particularly in the context of widening-participation and access interventions, is that of accessing reliable and complete data about participants (Crawford et al., 2017; Harrison et al., 2018; Passy & Morris, 2010). This is especially problematic for programmes with long-term outcomes, where participants may move through different education stages and data collection regimes, making it difficult to reliably track their progress and outcomes.

Several pilot teams in this project faced this challenge. For instance, the Leeds Arts University team decided to focus on short- and medium-term outcomes in the evaluation of their outreach programme, because they were unable to track students across subsequent educational stages (Leeds Arts University report, pp.15, 28–30). The University Centre Leeds team, also delivering an access intervention, and the University of Suffolk team, whose intervention aimed to impact student career outcomes, faced similar issues.

Deriving robust and effective evaluation measures

The robustness of evaluation outcomes depends at least partially on the validity of the measures used, irrespective of the evaluation methodology (Austen et al., 2021; Harrison & Waller, 2017; Howson, 2019). Some of the project teams observed that the small-cohort evaluation methodologies piloted did not mitigate challenges commonly experienced when creating and implementing robust and valid measures.

For example, the University Centre Leeds project team were concerned about their ability to evaluate the intangible outcomes of an outreach programme for Gypsy, Roma and Traveller young people. One of the key mechanisms for change in their programme relied on the development of an effective relationship between key stakeholders (participants, programme delivery officers and community representatives). The team noted that defining an effective relationship in this context was complicated, subjective and difficult to measure robustly. They did not feel that the Realist Evaluation approach they adopted provided a solution (University Centre Leeds report, p.16).

Similarly, some pilot teams had initially posited increased participant confidence as a crucial measure of programme success. The advisory team and pilot teams discussed the difficulties of using this as a measure, given that – like the 'effective relationships' concept described above – it is a broad construct, subjective and difficult to define or measure accurately. The University of Suffolk team, who conducted a Realist Evaluation of a student micro-placement programme, mitigated some of these challenges by recoding 'confidence' as a theme in their qualitative data. Sub-coding enabled them to break 'confidence' down into a series of domain-specific attributes, such as 'the confidence to … apply for jobs' or to 'speak to new people' (University of Suffolk report, p.21).

The issue of devising robust and meaningful measures was particularly pronounced for the University of Leeds' Lifelong Learning Centre team, where the design and calibration of effective attribute measures were integral to the Qualitative Comparative Analysis methodology they were piloting. One of their variables, for example, described participants' prior experience of education. This was coded as a binary value, with a cut-off threshold set at 10 years out of education. The project team based this on the advice from programme practitioners and 'on good practice basis that it split our data nearly evenly' (University of Leeds, Lifelong Learning Centre report, p.19). The team acknowledged that the 10-year threshold risked appearing 'somewhat arbitrary' (p. 31). Such challenges usually associated with operationalising complex situational variables could be viewed as reducing the robustness and consistency of evaluation outcomes, despite the otherwise technical and statistical integrity of the Qualitative Comparative Analysis approach. As the team observes, the formulation and calibration of attributes and measures was complex and potentially less 'objective' than was ideal (p.6).

In most cases, pilot project teams opted for a pragmatic compromise on measure definitions and variables; they recognised that establishing robust measure definitions is often difficult in HE evaluation contexts and that it is often necessary to reach a workable compromise in order to progress a project. Nonetheless, the strength and reliability of evaluation claims often rely on the quality of the measures used. Therefore, consideration of the important role played by robust, effective and valid measures – regardless of the methodology used – should remain at the heart of good evaluation practice when small-cohort methodologies are tested or implemented (Recommendation 7).

The need to rely on self-reported data

For some pilot teams, the use of a small-cohort evaluation methodology did not reduce their reliance on participants' self-reported data, and this caused some concerns about validity. This concern is shared with other forms of HE-based evaluation, as observed by Harrison and Waller (2017):

> Evaluations often rely on easily collected self-reports of attitudes and future intentions from young people (or teachers and parents) – measuring the measurable. Validity here is very uncertain, especially given priming and social desirability effects. (p.86)

The potential limitations and biases associated with self-reported data were acknowledged by the Leeds Arts University project team. They were concerned that the use of a pre- and post-test survey design to collect evaluation data risked introducing the Dunning–Kruger effect. This effect is produced when respondents who are inexperienced in a particular domain over-estimate their abilities and then downgrade their self-assessment when they become more familiar with the challenges involved. The Leeds Arts University team found that at the beginning of the programme many participants reported a high level of Likert-scale agreement with positive outcomes for certain intended programme objectives. This left them little room for improvement as the programme progressed (Leeds Arts University report, p.19).

The team also noted the risk of introducing social desirability effects, where participant responses reflect a desire to create a favourable impression on the programme team rather than genuine sentiment.

In some cases, the pilot teams' unease about the use of self-reported data was mitigated by triangulating it with data gathered from other perspectives (e.g. participants and other programme stakeholders). This process is effectively built into Contribution Analysis and Realist Evaluation, which employ a range of viewpoints, perspectives, experiences and knowledge to test and develop initial assumptions about change mechanisms and causal models. The University Centre Leeds team, for example, observed that the iterative thematic analysis used in their Realist approach encouraged them to triangulate different data sources. Qualitative data gathered directly from participants was combined with observations and reflections from key stakeholders, such as the programme delivery officer. The team also felt that this helped minimise researcher bias (University Centre Leeds report, p.25).

Small-cohort evaluation methodologies do not provide a solution to the challenges of creating robust and valid measures or the reliance on self-reported data. The experience of the pilot teams suggests that careful thought is required in the project design to ensure robust outcomes, irrespective of whether the methodology is designed for small-cohort interventions or larger programmes.

Isolating the intervention impact from other influences

Commentators have noted the difficulty of isolating the impact of target interventions from other external influences or interventions (Harrison & Waller 2017; Robinson & Salvestrini 2020). However, some pilot project teams suggested that small-cohort methodologies can help mitigate this issue by focusing on both internal factors and those external to the programme.

The Leeds Arts University project team, for example, observed that the focus in Contribution Analysis on 'contribution' rather than 'cause' navigates this issue by explicitly acknowledging that participants may be subject to a wide range of potential influences outside the immediate outreach intervention. As part of their 'contribution story', the team catalogued a range of external influences with the potential to reinforce or work against the programme aims (Leeds Arts University report, p. 10). Similarly, City College Norwich evaluated the Higher Education Tutorial Supervisor (HETS) role by considering the extent to which external factors – such as other forms of tutorial support, library resources and students' independent engagement in study – are likely to have contributed to the programme objectives. This enabled the project team to identify and isolate the specific contribution of the HETS to the programme.

Many small-cohort evaluation methodologies, such as Contribution Analysis, Qualitative Comparative Analysis, General Elimination Methodology, Process Tracing and Comparative Case Study, require evaluators to consider 'alternative hypotheses' or 'rival causal' explanations for observed outcomes. In so doing, these methodologies make the evaluation outcomes more robust by setting appropriate limits on the claims made about the target intervention and explicitly addressing the influence of external factors or variables. This is particularly important in complex environments such as HE where participants are likely to be part of an open system and are therefore exposed to alternative influences and factors that may affect the intended outcomes.

Benefits associated with small-cohort evaluation

As discussed in the preceding section, the project teams encountered various practical and methodological challenges when adopting small-cohort evaluation approaches. However, they also identified several positive outcomes of these methods when evaluating access, student success and progression interventions.

Some of these benefits specifically address the challenges of evaluation in the context of small cohorts, while others make a broader contribution to thinking about interventions across this domain. Indeed, the pilots suggest that small-cohort approaches also have the potential to contribute to programme development, implementation and delivery across the student lifecycle.

Knowledge production

Key learning points:

Small-cohort approaches appear to be effective in generating valuable knowledge about programmes and interventions.

In pilot projects, these evaluation approaches helped project teams develop their knowledge of:

- target participant groups
- the change mechanisms that contribute to intended outcomes
- the complex nature of their target programmes
- the underlying constructs and concepts upon which programmes are built.

There is a risk, however, that this important knowledge could remain concealed within individual evaluation reports. We therefore recommend that consideration be given to how these knowledge outcomes can be made accessible to and be used by peers, practitioners and evaluators alike.

The *Methodological Guidance* distinguishes between two types of evaluation approach: 'effects of causes' and 'causes of effects' (pp.8–13). The first type, which is associated with Type 3 (causal impact) evaluations, uses randomisation and quasi-experimental methods to hold external causal factors in balance between the treatment and control groups. These approaches tend to prioritise producing evidence that the programme or intervention caused the effects being measured over generating evidence of *how* and *why* they work. At the same time, effective trial design can incorporate some elements of theory development and implementation that can help to start unpicking these questions. In contrast, theory-driven 'causes of effects' approaches, such as small-cohort methodologies, prioritise the development of theories about *how* and *why* programme effects are produced. They try to identify potential causal factors and provide an explanation for the outcomes observed.

The usefulness of this knowledge depends on the purpose of the evaluation. Different stakeholders may have different expectations and requirements of the evaluation findings depending on the decisions they intend to make. Some decisions – for example, those about resource allocation or whether to continue or expand programme delivery – require clear impact evidence. Others – for example, about whether a programme is transferable and likely to deliver the same outcomes in different geographical areas or when delivered to different target groups – may need a broader range of knowledge about how and why a programme works and the contexts that influence this.

As they take a comprehensive theory-driven approach and explore how and why the observed outcomes are produced (or not), the small-cohort evaluation approaches piloted by the project teams proved to be effective in producing knowledge about how programmes cause change.

Each team's report provides further information about the specific knowledge and evidence produced by their chosen small-cohort methodology. Appendix 3 includes a summary chart indicating the strengths and weaknesses of the different methodologies in terms of knowledge production, as assessed by project teams. It should be noted that many of these methodologies are flexible and can be adapted to different contexts and requirements. The table in Appendix 3 is, therefore, a reflection on the *implementation* of the methodologies by each team, which in turn reflects that team's priorities and interests, rather than on the methodology itself.

In the section below, we focus in more detail on three key areas in which these evaluation approaches appear to support the development and enhancement of knowledge about programmes and interventions.

Knowledge of participant groups

The pilot teams found that applying a small-cohort methodology enabled them to learn more about their target participant groups.

Some of these methods directly engage participants as evaluation informants (see below), as they are asked to review or challenge the project team's assumptions about how an intervention works. For example, the Plymouth Marjon University team collected participants' narratives of their experience of university-based work placements using a Transformative Evaluation approach. This approach supported an inductive process of story collection that was able to generate a better and richer understanding of participants' experiences and perspectives than more prescriptive approaches such as surveys or semi-structured interviews, where participants' responses can be led by evaluators' interests or preoccupations.

Similarly, Realist Evaluation approaches, with their focus on the context in which change mechanisms work, encourage close attention to be given to the circumstances and attributes of intended participants. Working with Gypsy, Roma and Traveller young people and community representatives as co-creators of an access programme, the University Centre Leeds project team reflected that this approach enabled them to focus in detail on the experiences and preferences of the participating cohort.

> *"Realist evaluation has afforded us real benefits in understanding the outcomes and success of interventions for very specific under-represented groups such as GRT, and with small n cohorts. It has assisted us in illustrating why we are doing what we are doing [...] whilst highlighting the importance of context for these groups". (University Centre Leeds report, p.36)*

The team noted that this process of co-creation also helped develop the Project Officer's knowledge of the experiences, needs and contexts of Gypsy, Roma and Traveller young people, and this learning could be carried forward into future phases of project delivery and development (p.35).

The knowledge produced by small-cohort approaches about programme target groups is likely to be valuable to other practitioners working with the same cohorts in other programmes and interventions. There is a risk that this knowledge may be hidden in individual evaluation reports unless it is extracted and made available to peers looking through a target cohort lens. We recommend that a method of compiling and sharing this information is developed (Recommendation 6).

Knowledge about change mechanisms

Most small-cohort evaluation approaches involve an open and inductive process, including collecting the perspectives of evaluators, delivery practitioners and other programme stakeholders to build the programme theory. Involving a range of stakeholders in the process can generate new insights and a better understanding of how programmes and interventions work to deliver their outcomes.

This process also helps evaluators to identify and map new change mechanisms or causal chains. Change mechanisms are the 'causal mechanism by which the programme is expected to achieve its outcomes'.[5] Change mechanisms can occur at various levels or points in the intervention; some will be obvious from the outset and part of the programme design while others may appear as different stakeholder perspectives are collected. The University of Suffolk team, evaluating a student work-placement intervention, reported identifying a new change mechanism as a result of collecting different perspectives on the programme. They found that the commitment and dedication of their programme's 'anchor', the person supporting participants through their micro-placement was often a key success factor (University of Suffolk report, p.19). The team also found that Realist Evaluation strategies increased their awareness of the impact of change mechanisms on beneficiaries across the entire programme. In developing their Theory of Change and its integral context-mechanism-outcome models, several additional important causal mechanisms were revealed, in areas such as the targeting and recruitment of potential participants, which had not previously been considered active ingredients in programme outcomes (University of Suffolk report, p.8).

These evaluation methodologies require evaluators to pay close attention to the workings of programmes and their component parts. The increased knowledge and appreciation of programme complexity fostered in this way is apparent when comparing each of the project teams' initial Theory of Change with its final iteration. In each case, the final iteration of the Theory of Change was significantly more detailed and nuanced in its description of change mechanisms, and the framing of the intervention was more expansive than the first version (see below for a brief discussion of this process or, for a more

[5] For more information about change mechanisms and how they work within a Theory of Change, see the TASO resources on Theory of Change.

detailed example, see the project team reports). This increased understanding of programme complexity also provided a good foundation for improving and developing more carefully calibrated and relevant evaluation measures.

Knowledge about key constructs and concepts

As the project teams developed a deeper understanding of their target programmes and interventions, they also gained a better grasp of the key concepts and constructs upon which their programmes were built.

Initially, several of the target programmes in the pilots aimed to increase participant confidence as a primary or secondary outcome. However, as they refined their Theories of Change, some project teams found their initial concept of confidence to be too abstract, vague, subjective or difficult to measure accurately. This meant that 'confidence' was unable to function as a meaningful outcome indicator. The University Centre Leeds project team, for example, decided to refine their initial concept of 'confidence' and reconfigured it as a programme-specific construct encompassing knowledge, openness, empowerment and agency. This, in turn, enabled them to develop more a detailed and nuanced understanding of programme change mechanisms (University Centre Leeds report, p.10).

In this way, many of the project teams created a virtuous circle in which close attention to the underpinning conceptual framework of their target interventions increased their awareness of what they were evaluating and the kinds of change they expected to see. This, in turn, helped them refine and develop their evaluation measures and indicators.

Engaging participants in the process

Key learning points:

Many small-cohort evaluation approaches require a close engagement with programme participants and beneficiaries. This involvement of the participants in evaluation and data collection can contribute rich perspectives on programmes and interventions.

The combination of an increased attention to detail and a more highly developed understanding of the conceptual underpinnings of target programmes resulted in a more detailed understanding of how complex programmes and interventions delivered their outcomes. This knowledge could be further enhanced by including the experiences and perspectives of programme participants and beneficiaries.

A crucial element of the University Centre Leeds outreach programme was the close involvement of Gypsy, Roma and Traveller participants in the co-design of programme sessions. This proved to be a valuable evaluation exercise and resulted in the collection of detailed data on the perspectives, needs and preferences of the target cohort. Similarly, the Transformative Evaluation/Most Significant Change methodology used by Plymouth Marjon University is based on a self-reflective narrative approach. The storytelling and story-selection activities represented an opportunity for both participants and practitioners to reflect on the programme and their relationship with it. This process produced a range of valuable insights into how the programme functioned.

Involving participants and stakeholders as collaborators in the evaluation process is not exclusive to small-cohort evaluation approaches, but the opportunity for close involvement with programme participants when exploring how and why an intervention works can enable evaluators to collect valuable evaluation data and increase their understanding of how a programme works to deliver its intended outcomes.

Effective evaluation with a small number of cases

The primary goal of the *Impact evaluation for small cohorts* project was to explore the potential of evaluation methodologies explicitly designed for programmes with small numbers of participants. Perhaps because it was taken as a given, this aspect was rarely mentioned in team reflections. Nonetheless, each team was able to identify a range of meaningful evaluation outcomes for their target programme despite the small participant cohort.

The size of the participant cohorts involved in these evaluation projects varied between nine and 25 – see Table 3 below.

Table 3: Summary of pilot evaluation sample sizes

Project team	Evaluation methodology	Cohort size
University of Leeds, Lifelong Learning Centre	Qualitative Comparative Analysis	14 (programme participants)
Leeds Arts University	Contribution Analysis	18 (programme and evaluation participants)
University Centre Leeds	Realist Evaluation	12 (programme and evaluation participants)
Plymouth Marjon University	Transformative Evaluation / Most Significant Change	17 (evaluation participants)
University of Suffolk	Realist Evaluation	25 (programme participants)
City College Norwich	Contribution Analysis	9 (evaluation participants)

Each project team reported that their evaluation findings were valuable and significant despite the limited size of the programme and evaluation participant cohorts. At the same time, however, some teams suggested that a larger participant cohort could help them increase the robustness of their evaluation outcomes. The University of Leeds' Lifelong Learning Centre project team, for example, concluded that increasing the number of cases in its Qualitative Comparative Analysis, through its continued use in evaluation activities in subsequent years, would increase the robustness of the findings by increasing the diversity of the participant cohort (University of Leeds, Lifelong Learning Centre report, pp.30–31).

Recommendations for future implementation

In the reflective workshop held at the end of the project, project teams were asked what advice they would give themselves if they were starting their projects now. A summary of their responses is listed below:

- It is important to recognise the value of a team with a range of evaluation and research skills. This diversity promotes meaningful discussions and boosts motivation.
- It is crucial to acknowledge that the process will be difficult at times but that it will become easier. Hindsight reveals the progress being made, even if it is hard to see at the time.
- It is beneficial to be open to engaging with external networks and to seek support from both internal and external sources.
- From the outset, it is important to recognise the complex and time-consuming nature of this kind of project and the need for effective project management.

Summary of evidence gaps and gains

Table 4 below summaries each project team's reflection on what their evaluation methodology helped them learn and the areas it failed to cover. For more detail on the project teams' reflections on the specific methodologies piloted, please see Appendix 4

Table 4: Summary of what has been learned, evidence gaps and limitations of the method

Project team - programme	Methodology	Evidence and knowledge produced	Evidence/knowledge gaps and limitations
Plymouth Marjon University – Student Placement	Transformative Evaluation	• Increased understanding of the programme's impact on participants • Aspects of the programme which function or do not function in line with expectations	• Transformative Evaluation produced partial evidence about participants' development of skills framework attributes. • Because they were generated by participants, some Theories of Change and change mechanisms lacked sufficient development and detail compared to methods which included a broader range of stakeholder perspectives.
City College Norwich – HETS: Additional support for students with LDD	Contribution Analysis	• Increased confidence in the role of the HETS in closing gaps in student outcomes • A greater understanding of where and how the HETS role contributed • Greater understanding of external contributing factors to programme outcomes • Supported the conclusion that positive programme outcomes would be less likely without the presence of the HETS role.	• Contribution Analysis may not be a suitable approach where stakeholders require a rapid response or findings. • The approach to Contribution Analysis adopted by the team did not enable them to quantify the individual contribution of each causal factor.
Leeds Arts University – Creative arts outreach programme	Contribution Analysis	• Evaluation reveals the complex nature of the programme. • The time and timing of sessions are important success factors. • Increased participation in sessions is associated with positive outcomes. • Self-reported data suggests that participants' awareness of possible options, study requirements and application process all increased.	• The retrospective evaluation limited potential evaluation conclusions. • The small number of participants exacerbated issues of missing data and inconsistent attendance. • It is challenging to gather sufficiently robust evidence to effectively confirm programme outcomes. • The Theory of Change could have been strengthened by iterative testing. • The evaluation process is time- and resource-intensive.
University of Suffolk – Student micro-placement scheme	Realist Evaluation	• The programme appears to be successful in increasing student self-reported employability and job application skills. • The evaluation supported the programme improvement recommendations.	• The team felt there was less focus on external factors and alternative hypotheses than in other evaluation approaches.
University Centre Leeds – Outreach programme for Gypsy, Roma and Traveller young people	Realist Evaluation	• The relationship between participants, programme officer and community representative is a crucial success factor. • Continuing participants can support new participants and encourage peer participation. • Participants reported increased knowledge of potential future options. • The Programme Officer learned more about the target group and how to design effective and relevant programmes for them. • There is a positive correlation between participant engagement and outcomes. • Participants developed a greater openness to considering future opportunities.	• The inductive, embedded and informal data-collection process means that evaluation stakeholder perspectives dominate. • The evaluation data-gathering process was negatively impacted by inconsistent participant engagement.

Project team - programme	Methodology	Evidence and knowledge produced	Evidence/knowledge gaps and limitations
University of Leeds, Lifelong Learning Centre – Mature student outreach	Qualitative Comparative Analysis	• None of the hypothesised conditions or attributes were necessary (in presence or absence) for the programme outcomes. • Specific combinations of attributes correlated to positive programme outcomes.	• Limited variation in participant characteristics restricted the robustness of conclusions. • In-depth knowledge of the participants/cases was required if attributes were to be effectively calibrated.

SECTION 3 - CONCLUSIONS AND REFLECTIONS

The following reflections are those of the evaluation advisory team representatives.

Small-cohort evaluation approaches and Type 3 causal impact evaluation

Key learning points:

The small-cohort evaluation methodologies are effective in providing Type 1 (narrative) and Type 2 (empirical) evidence. As it is currently defined in the Office for Students *Standards of Evidence*, however, none of the approaches deliver Type 3 (causal) evidence, primarily due to the absence of a counterfactual, which is required by the current Type 3 definition.

Nonetheless, as piloted, these methodologies produce different forms of evidence by triangulating multiple stakeholder perspectives, seeking confirmation from programme participants, assessing the contribution of different or external causal factors and/or focusing on the operation of small-scale change mechanisms. They also generate detailed, albeit sometimes partial, knowledge about how and why programme outcomes are achieved. This is a valuable evaluation outcome in itself.

The Office for Students' access and participation standards of evidence guidance distinguishes between three 'types' of evidence:

- **Type 1 (narrative)** evidence is underpinned by a clear narrative rationale for the activity and/or Theory of Change and evidence base.
- **Type 2 (empirical)** evidence includes data on impact and can report evidence that those receiving an intervention have better outcomes, although it does not establish any direct causal effect. Impact evidence in this category may include qualitative or quantitative pre- and post-intervention designs that capture change across time and counterfactual differences (e.g. between participant and non-participant cohorts).
- **Type 3 (causal)** impact evidence is produced using more structured approaches involving a robust counterfactual design (usually using randomisation or quasi-experimental approaches to construct an effective counterfactual from within existing data).

Type 3 evaluations provide more confidence in the relationship between intervention and outcomes than Type 2 because they utilise more robust designs. TASO aligns with this typology and seeks to support the sector to produce more Type 3 evidence.

Given the definition of causality outlined in the existing standards of evidence, the small-cohort evaluation approaches piloted in this project, precisely because they involve cohorts too small for large-scale, quantitative designs, would be unlikely to meet the criteria for Type 3 impact evidence.

Nonetheless, this is not to suggest that they do not provide other valuable forms of evidence. The outcomes produced by small-cohort evaluation methodologies can increase evaluators' confidence in the relationship between intervention and outcome. These evaluation methods use a range of techniques:

- Triangulating a range of programme stakeholder perspectives on the causal relationship between intervention and outcome (Contribution Analysis, Realist Evaluation)
- Triangulating causal claims with the perspectives and experiences of intervention participants (Transformative Evaluation/Most Significant Change, Realist Evaluation, Contribution Analysis)
- Analysing the relationship between assumed contributary factors and programme outcomes (Qualitative Comparative Analysis)
- Identifying and testing causal mechanisms within the programme or intervention (Realist Evaluation).

All the pilot evaluation methodologies meet the criteria for Type 1 'narrative' impact evaluation because they include 'a coherent explanation of what we do and why' (Office for Students 2023, p.5). Many also meet the criteria for Type 2 'empirical enquiry', particularly where they include evaluation designs such as pre- and post-intervention data collection to provide evidence that a change or difference occurred 'compared to what otherwise might have happened' (Office for Students 2023, p.2).

Based on the Office for Students' *Standards of Evidence* guidance, however, none of the piloted methodologies currently meets the criteria for Type 3 (causal) evidence. TASO will continue to consider how evidence from these evaluations fits into the broader evidence landscape alongside more traditional large-scale quantitative or counterfactual evaluation methods.

The pilots have been successful in demonstrating the range and richness of the evaluation outcomes produced by small-cohort evaluation approaches. These approaches also provide detailed insights into how and why certain programmes or interventions work (or do not work) in particular contexts. As with other evaluation outcomes, however, there is a challenge in ensuring that this knowledge – even where it is provisional, partial, fragmentary or shows a null result – is readily available, accessible and fit for use by delivery and practitioner staff across the sector.

A potential solution is to develop a collaborative, sector-wide process for sharing outcomes. This could support the iterative development and refinement of knowledge about what works, in what contexts, and for whom across different HEPs and evaluators. Ultimately, this would increase the impact of the collective work conducted by the HE sector across the entire student lifecycle.

Sharing learning and findings could potentially encourage both evaluators and delivery practitioners to consider this knowledge in a modular way. By drawing on detailed evaluation outcomes relevant to their intervention, target group or operating context, they may be able to focus more directly on individual activity components or change mechanisms. This would support a more granular approach to evaluation and enable practitioners and evaluators to consider how various elements of a programme contribute to intervention outcomes. This was the outcome for the project teams piloting Realist Evaluation, Contribution Analysis and Qualitative Comparative Analysis methodologies.

This modular approach to programme design may also mitigate concerns about what Younger et al. (2019), Robinson and Salvestrini (2020) and others have described as 'black box' interventions. These feature 'multiple components [...] meaning that a randomised controlled trial or quasi-experimental design cannot by itself indicate which elements of each programme may have been instrumental in causing any identified effects' (Younger et al. 2019, p. 750). By dividing programmes and interventions into component change mechanisms, the methodologies associated with small-cohort evaluation make it possible to consider the relationship between different programme components and particular outcomes. This approach helps to open up the 'black box' of complex interventions and expose its workings to view.

Small-cohort evaluations support the development of Theories of Change

Key learning points:

Small-cohort evaluation approaches typically start with developing an outline of how and why interventions are expected to be effective (usually through a Theory of Change). This process can help practitioners and evaluators explore, test and comprehend how activities and programmes achieve their intended outcomes.

In the course of their pilot projects, each project team significantly developed their thinking and understanding of the target intervention. The progress made by each team can be seen by reviewing their final project report and comparing their Theory of Change model at the start of the project to the final version, which is much more detailed and complex. For example, the University Centre Leeds team began with a 'logic model' structure that encouraged them to consider the programme's needs, aims, activities and outputs. The final enhanced Theory of Change was much more complicated and included significant detail about the change mechanisms involved (University Centre Leeds Report, p. 11). This Theory of Change development process is an area where small-cohort evaluation methodologies can add significant value to both the evaluation and development of HEP interventions and programmes.

Methodologies can be combined

Key learning points:

Based on the pilot studies, the small-cohort evaluations do not meet the current standards for Type 3 causal evidence. Nonetheless, the outcomes of this project suggest there is potential in combining them with Type 3 causal evaluation designs, thereby enhancing the value and utility of both.

Although we have suggested that small-cohort evaluation methodologies do not currently meet the criteria laid out in the *Standards of Evidence* for Type 3 causal impact, the pilot projects have shown that they have the potential to produce other types of valuable evidence and knowledge about how programmes work to deliver impact. The pilot project outcomes suggest that these methodologies may also be usefully combined with experimental or quasi-experimental designs to enhance evaluation outcomes. Small-cohort methodologies could, for example, support the development of appropriate and effective evaluation measures and strengthen the evidence base underpinning Type 3 causal evaluation designs.

How such combinations might work in practice in an HE context should be considered as an area for future research or pilot projects. Different small-cohort methodologies could, for example, be piloted at different stages of the evaluation process (Recommendation 8).

Table 5: Summary of how different methodologies could be used at different stages in an evaluation

Stage	Purpose	Methodologies
Theory/knowledge generation	• Develop Theory of Change • Identify change mechanisms • Map causal chains • Investigate rival explanations	• Contribution Analysis • Process Tracing • General Elimination Methodology • Realist Evaluation
Challenge and refine Theory of Change/ mechanisms	• Gather stakeholder perspectives • Explore rival theories • Gather evidence to support Theory of Change	• Contribution Analysis • Process Tracing • General Elimination Methodology • Realist Evaluation • Most Significant Change
Investigate change-mechanism interactions	• Explore the interaction of different change mechanisms	• Contribution Analysis • Process Tracing • Qualitative Comparative Analysis • Realist Evaluation

Acknowledging complexity

Key learning points:

The theory-driven process underpinning many of the small-cohort evaluation approaches piloted in this project reveals the complexity of the target interventions and their impact on participants. Simply acknowledging this complexity can enhance the robustness and effectiveness of the evaluation process.

The 'causes of effects' approach that characterises many of the small-cohort evaluation approaches described in the *Methodological Guidance*, and the granular approach to understanding change mechanisms they encourage, requires a shift away from linear conceptions of causality. Many of the approaches piloted encourage us to view a programme as a network of interrelated change mechanisms or causal chains rather than a single step-by-step process. This reconfigures programmes as a complex system in which different components interact and contribute to programme outcomes.

Each evaluation methodology produces a distinct version of complexity: for instance, the Realist Evaluation approach piloted focuses on the context of individual participants, treating each person as a unique case, with a specific context and patterns of interaction and, consequently, a unique journey through the programme. As such, programme change mechanisms may operate at different times and produce particular outcomes for different participants. From this perspective, a programme no longer operates as a simple linear process with defined stages and outcomes, but as a complex system with the potential for different interactions and outcomes to occur at different points.

This complexity is compounded when considering the impact of external factors on programme outcomes. Methodologies such as Contribution Analysis, Realist Evaluation, Qualitative Comparative Analysis, Process Tracing and General Elimination Methodology require evaluators to consider alternative hypotheses or rival explanations for observed outcomes. The Leeds Arts University team, for example, identified several external factors with the potential to impact their target programme when developing their Theory of Change. They note that 'these factors could intersect, multiply in influence, or even act to cancel each other out, impacting on each participant's context in a range of different ways' (Leeds Arts University report, p.14). The team concluded that the process of developing the Theory of Change highlighted the complexity of their programme. Similarly, a Qualitative Comparative Analysis approach is intended to capture and assess the contribution of different factors, whether designed into the programme or interposed from outside. As the team at the University of Leeds' Lifelong Learning Centre noted, their analysis process aimed to eliminate extraneous factors and determine which programme elements are causally associated with programme outcomes.

The theory-driven process of mapping target programmes into a Theory of Change reveals the inherent complexity of these programmes. Acknowledging and working with this complexity can improve the robustness and effectiveness of evaluations in the access and participation space.

An appropriate quality assurance framework may be needed

Initially, the project advisory team found it challenging to develop a reliable quality assurance framework and standardised proforma that could be used to guide each of the projects. This was mainly due to a lack of consistency in the approach, underlying philosophy and intended outcomes of each evaluation approach. While this was not necessarily problematic for the current project, it does raise concerns about effective quality assurance, reporting and communication in future evaluations of small-cohort approaches.

The Centre for Evaluating Complexity Across the Nexus (CECAN) has published an evaluation tool designed to support evaluators in robustly selecting and implementing alternatives to large-scale counterfactual approaches (Befani 2020). The report's author suggests that 'evaluation quality should be assessed on the basis of at least eight dimensions: conceptual framing, transparency, appropriateness, validity, credibility, transferability, reliability, and structure' (p.4). This highlights the need for a meta-evaluation framework to support the effective and robust implementation and reporting of future small-cohort evaluation projects (Recommendation 9).

Considerations for the future use of small-cohort methodologies

Overall, the outcomes of the project pilots indicate that:

- Small-cohort evaluation approaches can provide valuable and relevant findings for the HE sector. They are effective in supporting knowledge generation about how and why programmes work. They can also increase evaluators' confidence in the relationship between intervention and outcomes.
- These evaluation methodologies are not straightforward to deliver, however, and realistic consideration needs to be given to the timeframe, resources and skill sets required.
- Pilot teams have successfully implemented these methodologies within project constraints, but further work is needed to fully realise their potential.

- Some pilot methodologies – Qualitative Contribution Analysis, in particular – make significant technical, knowledge and skills demands, and external support and guidance are likely to be essential for evaluators who do not already possess the skills and experience required.
- Small-cohort approaches are still at an early stage of development in the HE context and more work is needed to ensure that an appropriate quality assurance framework and reporting guidance are developed to support evaluators working with these methodologies.
- The small-cohort approaches included in the *Methodological Guidance* differ in nature from the counterfactual Type 3 causal evaluation designs described in the Office for Students' *Standards of Evidence* guidance. However, the pilots suggest they may be productively combined with this kind of large-scale quantitative impact evaluation to produce robust causal evidence. Further research and piloting are necessary to determine how this combination will work in practice.
- These approaches raise the profile and importance of a range of evaluation stakeholders – including participants, practitioners and programme stakeholders, among others. The additional knowledge that results from their involvement could be a real asset in increasing the complexity and nuance of future evaluations and developing the evidence base of what works, when, for whom and in what contexts. This should, in turn, support more nuanced and flexible programme implementation.

RECOMMENDATIONS

Recommendation 1:

The pilot projects engaged with four small-cohort methodologies. Pilots of small-cohort methodologies not included in this project (Process Tracing, General Elimination Methodology and Comparative Case Study) should be encouraged and supported.

Recommendation 2:

Given the challenges experienced by the project teams in adapting and implementing small-cohort evaluation methodologies, sufficient consideration should be given at the planning stage to the significant time and resources required to engage effectively with unfamiliar evaluation methodologies.

Recommendation 3:

Given the currently limited availability of HE-specific examples, small-cohort evaluation methodologies conducted in other disciplinary areas should be identified and signposted to evaluators working in the HE sector. To make these resources more accessible, they should be translated into terms or contexts relevant to this sector.

Recommendation 4:

Conducting high-quality, robust evaluation is time-consuming and resource-intensive. HEPs should invest in further evaluation capacity to facilitate stronger evaluation practice across their Access and Participation Plan portfolio to ensure they are providing students with the best possible support.

Recommendation 5:

Working in isolation on complex, technical or unfamiliar evaluation methodologies brings a risk of incorrect applications, flawed conclusions and stalled projects. HEPs should seek opportunities for peer support, in which practitioners and evaluators with experience in a particular approach can support or advise less experienced colleagues.

Recommendation 6:

Evaluators across the sector are generating valuable knowledge about change mechanisms, target groups and evaluation implementation. The use of small-cohort methodologies is likely to increase this knowledge. A central repository of emerging sector knowledge and evaluations should be considered.

Recommendation 7:

The design and implementation of robust, effective and valid evaluation measures are integral to good evaluation practice. Irrespective of the methodology used, evaluation practitioners should ensure that sufficient attention and consideration are given to identifying appropriate measures.

Recommendation 8:

The outcomes of the pilot projects suggest that small-cohort evaluation approaches may be productively combined with Type 3 impact-evidence methodologies. Further research or pilot studies should be conducted to assess the potential of combining approaches in this way.

Recommendation 9:

There is currently no suitable quality-assurance framework to guide future work in this area. A formal quality-assurance framework and reporting guidance should be developed to improve the rigour of future small-cohort evaluation projects.

REFERENCES

Austen, L., Hodgson, R., Heaton, C., Pickering, N. and Dickinson, J., 2021. *Access, retention, attainment and progression: an integrative review of demonstrable impact on student outcomes*. Advance HE.

Befani, B., 2020. *Choosing appropriate evaluation methods: A tool for assessment and selection* (Version 2). London: CECAN. Available at https://www.cecan.ac.uk/wp-content/uploads/2020/11/Final_Choosing-Appropriate-Evaluation-Methods-1.pdf

Crawford, C., Dytham, S. and Naylor, R., 2017. The evaluation of the impact of outreach: *Standards of evaluation practice and associated guidance.* The Office for Fair Access. Available at https://pure.northampton.ac.uk/en/publications/improving-the-evaluation-of-outreach-interview-report

Dalkin, S.M., Greenhalgh, J., Jones, D., Cunningham, B. and Lhussier, M., 2015. What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science,* 10(1), pp.1–7.

Gorard, S. and Smith, E., 2006. Beyond the 'learning society': what have we learnt from widening participation research? *International Journal of Lifelong Education,* 25(6), pp.575–594.

Harrison, N. and Waller, R., 2017., Evaluating outreach activities: overcoming challenges through a realist 'small steps' approach. Perspectives: *Policy and Practice in Higher Education,* 21(2–3), pp.81–87.

Harrison, N., Waller, R. and Last, K., 2015. *The evaluation of widening participation activities in higher education: A survey of institutional leaders in England.*

Harrison, N., Vigurs, K., Crockford, J., McCaig, C., Squire, R. and Clark, L., 2018. *Understanding the evaluation of access and participation outreach interventions for under 16 year olds.*

Hernán, M.A. and Robins, J.M., 2020. *Causal Inference: What If.* Boca Raton: Chapman & Hall/CRC.

Howson, C.B.K., 2019. Final evaluation of the Office for Students Learning gain pilot projects. *Report to the Office for Students.* King's College.

Markus, H. and Nurius, P., 1986. Possible selves. American Psychologist, 41(9), p.954.

Mayne, J., 2008. Contribution Analysis: An approach to exploring cause and effect. Brief 16, *Institutional Learning and Change (ILAC) Initiative.* Available at https://nonprofitbuilder.org/storage/377/Contribution-analysis-An-approach-to-exploring-cause-and-effect-ILAC.pdf

Office for Students, 2023. *Standards of evidence and evaluation self-assessment tool.* Available at https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/evaluation/standards-of-evidence-and-evaluation-self-assessment-tool/

Oyserman, D., Bybee, D. and Terry, K., 2006. Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology,* 91(1), p.188.

Passy, R. and Morris, M., 2010. *Evaluation of Aimhigher: learner attainment and progression. Final report.* Slough: NFER.

Porter, S. and O'Halloran, P., 2012. The use and limitation of realistic evaluation as a tool for evidence-based practice: A critical realist perspective. *Nursing inquiry,* 19(1), pp.18–28.

Robinson, D. and Salvestrini, V., 2020. *The impact of interventions for widening access to higher education: A review of the evidence.* Centre for Transforming Access and Student Outcomes.

TASO (Centre for Transforming Access and Student Outcomes), 2020. *How response rates can affect the outcomes of a study and what to do about it.* Available at https://taso.org.uk/news-item/how-response-rates-can-affect-the-outcome-of-a-study-and-what-to-do-about-it/

White, H. and Sabarwal, S., (2014). Quasi-experimental design and methods. *Methodological Briefs: Impact Evaluation,* 8, pp.1–16.

Younger, K., Gascoine, L., Menzies, V. and Torgerson, C., 2019. A systematic review of evidence on the effectiveness of interventions and strategies for widening participation in higher education. *Journal of Further and Higher Education,* 43(6), pp.742–773.

APPENDICES

Appendix 1: Summary of the project teams' rationales for methodology selection

The table below summarises the reasons reported by each project team for selecting a particular methodology to pilot.

Pilot partner	Methodology selected	Reason for selection
University Centre Leeds	Realist Evaluation	• Guidance from Befani (2020). • The team's interest in exploring how programme activities can cause or contribute to programme impacts.
Leeds Arts University	Contribution Analysis The team also drew on General Elimination Methodology and Process Tracing.	• Guidance from Befani (2020). • The team's interest in exploring how programme activities can cause or contribute to programme impacts.
Plymouth Marjon University	Transformative Evaluation / Most Significant Change	• Access to the originator of Transformative Evaluation (Professor Sue Cooper) who is based at the university and provided advice and guidance. • The team's interest in exploring and developing the potential value and application of this methodology in an HE setting.
City College Norwich	Contribution Analysis	• The team's interest in exploring the correlation between positive intervention outcomes, the introduction of the HETS role and the extent to which this role contributed to these outcomes.
University of Leeds, Lifelong Learning Centre	Qualitative Comparative Analysis	• The team's interest in exploring and understanding the configuration of contributing factors that are correlated with positive programme outcomes.
University of Suffolk	Realist Evaluation	• The team's interest in understanding how the intervention worked to deliver its outcomes.

Appendix 2: Evaluation methodology map

Methodology	Starting point	Consultation/ perspective gathering	Theory of Change development approach	Iterative development process	Additional evidence gathering	Development of alternative hypothesis / rival explanations	Tests for theory/ change mechanisms
Contribution Analysis	Usually, an initial Theory of Change or logic model to map programme assumptions.	Test plausibility of Theory of Change with stakeholders.	Construct Theory of Change and identify change mechanisms/ causal chains/ contribution story.	The contribution story is tested with a range of stakeholders.	Gather evidence for the plausibility of the causal chain using surveys, interviews, focus groups and external literature.	Collect evidence in support of or against rival explanations/ external factors.	Assess plausibility of Theory of Change. Potentially use additional assessments, such as hoop and smoking-gun tests.
Realist Evaluation	Initial Theory of Change or logic model, designed by practitioners	Draw on literature review, multi-sector forums, community representatives and the views of programme practitioners and participants.	Gather thinking and evidence about CMO configurations to map assumed interaction between context, mechanisms and outcomes	Iterative process – thematic analysis of evidence gathered to construct CMOs. Test emerging theories with a range of stakeholders.	Evidence gathering – process of developing CMOs. Interviews, focus groups, workshops with participants and community representatives, and gathering practitioner reflections. Thematic analysis of data.	The focus on context and mechanisms allows for a range of causal models to be identified and included in the model.	An iterative process of evidencing and testing the CMO configurations
Transformative Evaluation / Most Significant Change	The inductive nature of Transformative Evaluation means that it is not always necessary to have an initial Theory of Change or model of causal chains.	This is at the heart of the Transformative Evaluation process, which relies on the contributions of programme participants. Other stakeholders review and select participant narratives.	There is no explicit focus on creating a formal Theory of Change. Instead, the Transformative Evaluation process develops a curated series of contribution and change stories and stakeholder reflections.	Long-term Most Significant Change evaluations involve iterative processes of story-gathering to capture change and developments across time.	Some Most Significant Change approaches include a story-validation phase to test the accuracy of participant stories.	This is not a focus of Transformative Evaluation or Most Significant Change approaches, which attempt to capture participant experiences in an unmediated way.	The process of story selection and, where relevant, validation may provide some element of testing.
Qualitative Comparative Analysis	Qualitative Comparative Analysis normally begins with a Theory of Change or logic model to identify likely conditions or attributes for analysis.	This is not necessarily a feature of Qualitative Comparative Analysis. The attribute calibration process can involve consultation with key stakeholders to assess and calibrate attributes	The Qualitative Comparative Analysis process does not necessarily draw on a conventional Theory of Change. This is replaced by a truth table or parsimonious solution charting the relationship between attributes and outcomes.	The analysis phase is likely to involve iterative phases of attribute modification to eliminate contradictions and create clear outcomes. This process may involve reviewing initial assumptions about attributes.	The process of defining and calibrating attributes is often the key evidence-gathering approach.	The process of gathering and testing attributes constitutes the assessment and testing of a range of potential alternative attributes.	The analysis phase represents a robust and logical test for the impact and contribution of the different attributes.

Appendix 3: Strengths and weaknesses of the piloted methodologies

P = partial contribution (some evidence of knowledge contribution, but not a key or comprehensive knowledge outcome)

S = strong contribution (evidence of substantial knowledge contribution)

	Understanding of individual change mechanisms	Understanding interaction of different change mechanisms	Identifying alternative causes	Identifying new causal mechanisms	Understanding of the strength of contribution	Understanding of target group	Demonstrating robust evidence of impact	Supporting recommendations for future programme development
City College Norwich – Contribution Analysis	S	S	S	S	P	S	P	S
Leeds Arts University –Contribution Analysis	S	P	S	S	P	S	P	S
University Centre Leeds – Realist Evaluation	S	P	P	P	P	S	P	P
University of Suffolk – Realist Evaluation	S	P	P	P	P	S	P	P
Plymouth Marjon University – Transformative Evaluation	S	S	P	P	S	S	P	P
University of Leeds, Lifelong Learning Centre – Qualitative Comparative Analysis	P	S	P	P	S	P	P	P

Appendix 4: Project team reflections on specific methodologies

Key learning points from project team reflections:

Employing small-cohort evaluation methodologies can involve engaging with unfamiliar concepts, theory and terminology, resulting in an initially steep learning curve.

These methodologies are also time- and resource-intensive, but cross-sector knowledge-sharing and collaborative work can help mitigate these challenges.

These approaches encourage close and detailed attention as to how a programme works to deliver outcomes, and this can generate a nuanced understanding of programme complexity. Where stakeholders and participants are closely involved in evaluation, this can also increase their understanding and awareness of the complexity and operation of the programme.

The implementation of these approaches may become easier as more HE-relevant examples and case studies are developed.

While some small-cohort methodologies are valuable in identifying internal and external causes of programme outcomes, they may not be as effective as other approaches in quantifying the strength or influence of these causal mechanisms.

As some of these methodologies concentrate on factors or variables at the individual level, they may limit opportunities to evaluate institutional or structural factors.

This appendix draws together the pilot teams' reflections on their chosen small-cohort methodology and their recommendations for future practice.

REALIST EVALUATION

- The specific terminology and concepts associated with Realist Evaluation are initially challenging to understand. In particular, the central concepts of *context* and *mechanism* both assume methodology-specific definitions. Initial challenges in working with these concepts can be exacerbated by the current lack of domain-relevant case studies and examples to illustrate their use.
- The University of Suffolk team observed that the specific methodological requirements of some elements of the Realist Evaluation approach meant that researchers on their team had to adapt their pre-existing qualitative research skills to conduct Realist interviews and focus groups.
- The Realist Evaluation focus on contexts, mechanisms and outcomes provides a useful lens through which to view programmes from a new perspective. This alternative view reveals how complex programmes can be, with multiple stages, stakeholders and change mechanisms to consider.
- Both project teams reported developing new and richer understandings of their interventions as a direct result of employing Realist Evaluation.
- In their conclusion, the University Centre Leeds team recommended that practitioners wanting to employ this evaluation approach should allocate sufficient preparation time to ensure they have a working understanding of the various concepts and terminology associated with Realist Evaluation.

CONTRIBUTION ANALYSIS

- As with Realist Evaluation, the project teams piloting Contribution Analysis described initial challenges in engaging with unfamiliar terminology and the concepts that are at the heart of the methodology.
- The project teams found the initial process of breaking their programmes down into contribution components challenging. However, once achieved, this helped them develop a detailed understanding of the active change mechanisms that were contributing to programme outcomes. It also helped them to see the relationships and interactions between the different activity strands that made up their target programme.
- The project teams found it useful to consider contributory factors external to the programme and how they either supported or hindered the intended programme outcomes.
- Contribution Analysis is a time- and resource-intensive methodology, and neither project team felt able to fully explore all the contribution chains identified. The Leeds Arts University team suggested that this limitation might be mitigated if the HE sector were to collaborate on building a library of change mechanisms upon which other practitioners could draw and build (Recommendation 6).
- The City College Norwich team observed that, although Contribution Analysis enabled them to identify key causal mechanisms, it did not provide 'out-of-the-box' support for exploring or assessing variations in the strength of these mechanisms. They noted that this could be a potential limitation of the Contribution Analysis approach.

TRANSFORMATIVE EVALUATION/MOST SIGNIFICANT CHANGE

- The project team concluded that Transformative Evaluation provides a useful framework for engaging with stakeholders, and especially participants. By bringing together different stakeholders, the process encourages collaboration that may not otherwise occur.
- The evaluation process increased the university's understanding of the relationship between the programme operation and the outcomes it delivered.
- To be effective, Transformative Evaluation requires evaluation participants to make a commitment to its underpinning methodology and to supporting specific outcomes, such as staff professional development.
- Some combinations of story collectors and participants (e.g. staff and students respectively) required considered management to avoid potentially disruptive power imbalances.
- The project team note that Transformative Evaluation may not generate sufficiently robust evidence to meet the requirements of some evaluation stakeholders, such as the Office for

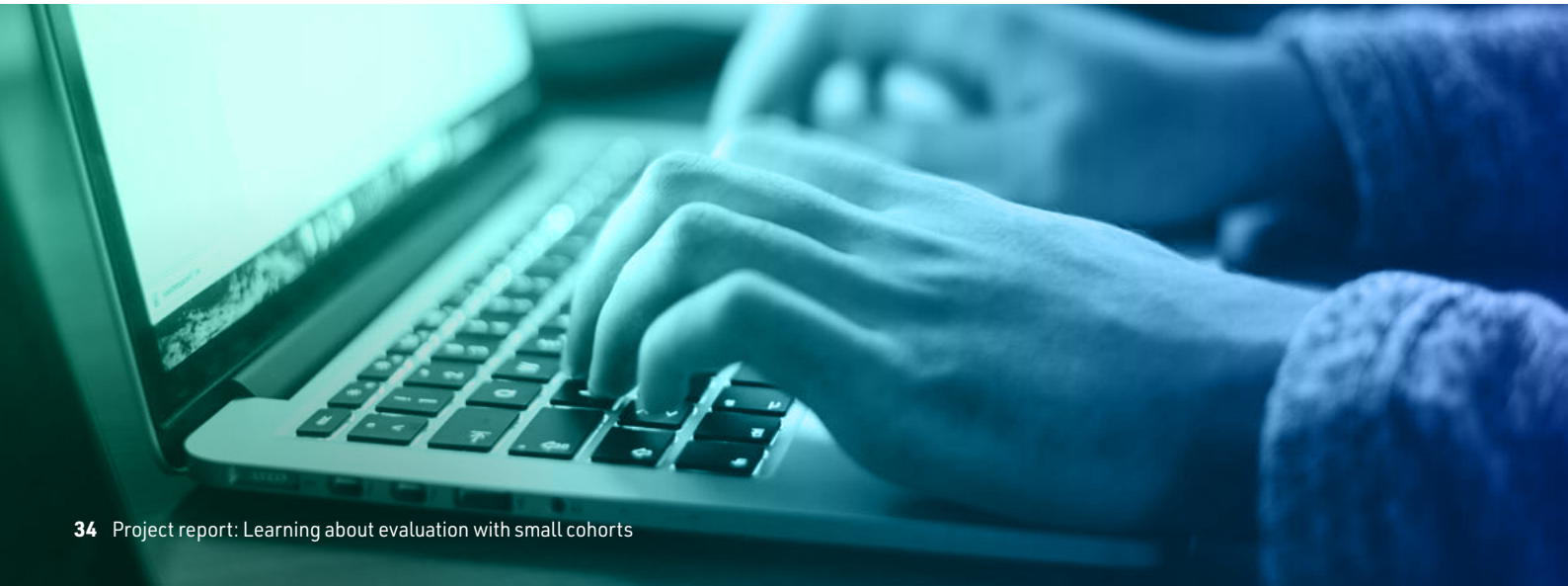
Students. As a result, the team supplemented the Transformative Evaluation data-collection approach with an additional round of qualitative data collection..

- The inductive approach that is integral to Transformative Evaluation prioritises the participant's voice and perspectives. The project team noted that, while this approach brought positive outcomes, it could limit the framing of the target intervention and reduce opportunities to consider structural or institutional factors.
- The team also observed that the story-collection approach, which relies on the story collector to document participant stories, could result in a reductive data-capture process or risk introducing biases.
- Unlike some of the other small-cohort approaches, Transformative Evaluation does not require or encourage any consideration of alternative or external influences on programme outcomes. The team noted this as a potential limitation of the methodology.

QUALITATIVE COMPARATIVE ANALYSIS

- The project team from the University of Leeds' Lifelong Learning Centre concluded that a Qualitative Comparative Analysis approach is useful for identifying patterns in the relationship between potential contextual factors and programme outcomes across multiple cases (participants). To be effective, however, an in-depth knowledge of each participant case is required.
- On a practical note, the team observed that some of the current software designed to support Qualitative Comparative Analysis is outdated and that the use of the statistical package R is often more effective. This can, however, represent a steep learning curve for researchers not already familiar with R.
- The team reflected that the strict selection criteria for the outreach programme they were evaluating limited the diversity of programme and evaluation participants. This, in turn, limited the degree of variation in the key variables they were analysing. They plan to continue using a Qualitative Comparative Analysis approach in future years in the hope that this will help them build a data set with greater diversity and increase the strength of their findings.
- As with Transformative Evaluation, the project team felt that Qualitative Comparative Analysis led them to focus on variables at the level of individual participants, and this limited their opportunity to test and consider structural, institutional and other macro- and meso-level factors.
- In terms of the methodological drawbacks of this approach, the project team suggested that because there was no requirement to involve programme stakeholders in a 'consult and challenge' process, their opportunities to test and strengthen their initial assumptions about how the programme worked were limited. Indeed, they observed that the initial Theory of Change was built on the perspectives of a small team of internal stakeholders and that without challenge it risked becoming self-referential.

Most of the project teams concluded that the initial process of building a Theory of Change, by paying close attention to the operation of their programme, was useful and productive. The increasing focus on change mechanisms as the pilot projects progressed often revealed the complexity of interventions in this space and, in many cases, required the project teams to review and reorientate their approach to evaluation, as their understanding of the programme developed.



TASO

Transforming Access and Student Outcomes in Higher Education

Evidence Quarter
Floor 4, Albany House
94-96 Petty France
London SW1H 9EA

info@taso.org.uk
taso.org.uk

TASO is an independent charity that aims to improve lives through evidence-based practice in higher education (HE). We support HE professionals through research, toolkits and evaluation guidance on what works best to eliminate equality gaps. We inform practitioners of the best available evidence and produce new evidence on the most effective approaches. TASO is an affiliate 'What Works' centre and is part of the UK Government's What Works Movement.