Enhancing Widening Participation Evaluation through the development of a self-assessment tool for practitioners: Learning from the Standards of Evaluation Practice (Phase 2) project 2017-2019

Anna Mountford-Zimdars, University of Exeter

Joanne Moore, University of Exeter

Richard Shiner, Office for Students

Email: a.mountford-zimdars@exeter.ac.uk

Abstract The ultimate goal of widening participation work, that unites policy makers, practitioners and academics, is to enhance outcomes for disadvantaged students in HE access, success and outcomes. This article presents the Office for Students (OfS) commissioned project *'Understanding effective evaluation of the impact of outreach interventions on access to higher education: Phase Two'*. This project sought to create a step-change in robustness of evidence used and evaluation practices. Our 2017-19 project explored evaluation practices among nine partner organisations drawn from higher education providers (HEPs) and third sector partners. The initial aim was to pilot test the *Standards of Evidence* developed during Phase One research, to share practices that work and highlight examples of best practice. As the project developed, a further outcome of the project was envisaged through ongoing discussions between the OfS, academics, HEPs and third-sector parties: a self-assessment tool for evaluation practitioners. This tool provides a framework and guidance which allows practitioners to map their own evaluation approaches. Through using prompts, the five dimensions of the tool highlight strengths and weaknesses of evaluation within five domains (strategic context, programme design, evaluation design, evaluation design and learning). This new tool is in effect a five point framework setting guidance about good evaluation practice. It was rolled out to all HEPs in spring 2019 when providers were invited to return the

completed tool and their reflection as part of their Access and Participation planning.

Key words : Framework; action research; collaborative research; co-creation; evaluation dimensions.

Introduction

As in other countries, UK education providers are attempting to create wider participation in higher education (HE) for traditionally under-represented social groups. A common approach to widening participation to higher education involves outreach work in relatively deprived communities and schools, for example, to provide enrichment activities to encourage young people from groups currently under-represented in higher education to apply to university.

The Office for Students (OfS) is the independent regulator of higher education in England. It was established by the Higher Education and Research Act 2017, coming into existence on 1 January 2018. It merged the Higher Education Funding Council for England and the Office for Fair Access, and formally inherited their responsibilities, while 'working in the interests of students and prospective students' and having 'a wider remit ... taking charge of the granting of degree awarding powers and university title.' The OfS largely inherited HEFCE's funding responsibilities and OFFA's responsibility for promoting fair access to higher education.

The OfS's policy tools to widen participation include setting national targets to reduce specific gaps in participation, and agreeing Access and participation plans (APPs) which set out how higher education providers will improve equality of opportunity for underrepresented groups to access, succeed in and progress from higher education. As part of it's remit for supporting all students to access and succeed in education, regardless of background, the OfS expects higher education providers (HEPs) to rigorously evaluate the impact of the significant widening participation resources which go to HE outreach. Strategic sector-wide support for this agenda included the OfS (and its predecessors) commissioning a range of research and practitioner projects.

OFFA had commissioned a project to develop guidance on best practice of evaluation methods. This first phase of the project was to investigate how universities and colleges currently evaluate outreach schemes to provide guidance on 'best practice' evaluation methods, and practical suggestions on how institutions make their outreach schemes more effective. The first phase developed a set of evidence standards and OFFA and then OfS wanted to test if these standards could be used to improve the quality of evaluation practice. The ultimate aim is to improve the effectiveness of outreach interventions because over £150m is being spent each year and there is currently a lack of evidence about what is most effective.

In 2017, the OfS commissioned a phase 2 of this initial project: *“Understanding effective evaluation of the impact of outreach interventions on access to higher education: Phase Two”*. The aim was to highlight examples of best practice in action and to share learning across the sector. The tender was awarded to a collaborative partnership using an practitioner-academic co-research design. The project involved academics, third-sector partners, Higher Education Providers (HEPs) and OfS, led by the Centre for Social Mobility at the University of Exeter. The project ran from September 2017 to January 2019. The focus was on the opportunities for learning around the operationalisation of the *Standards of Evaluation Practice* (OFFA, 2016). The work resulted in a self-assessment tool for evaluation practice for HEPs, which was published by OfS alongside their Access and Participation Plan guidance. This article draws out learning and implications for widening participation practitioners and evaluators.

Context

While Outreach has been a signficant feature of widening participaiton initiatives in England over the last 20 years evaluating its impact has been challenging, not least because it is often difficult to isolate the effect of individual interventions in the highly complex decision process involved in going to higher education.

Gorard et al’s (2006) review of widening participation research evidence identified gaps in the evidence of effectiveness of outreach, in particular the complete lack of trial based evaluation studies. A follow-up funding council commissioned review of widening participation research commissioned identified the need for robust evaluation processes to be built into outreach activities from the beginning that focus on targeting and measuring change (Moore et al. 2013). However, the availability of methodologically sound evidence was barely improved for the next major review of WP outreach practice, which studied evidence in England and on other Continents, to conclude that there are few common approaches, and few sector level frameworks to draw on (Bowes et al, 2014).

The lack of a systematic evidence base to pinpoint which initiatives are most successful has persisted, with the Sutton Trust identifying the need for increased emphasis on adequate spending on evaluation, and utilisation of the research expertise within HE institutions to strengthen evaluation methodologies including undertaking robust research trials using experimental research designs (Sutton Trust, 2015, Torgerson et ak, 2014).

There has been increasing focus and expectation on providers of outreach to raise the standard of their approaches to evaluating impact. In its 2015-20 strategic plan, the then Office for Fair Access (OFFA) asked institutions for evidence-led approaches to improving performance across the whole student lifecycle. This commitment continued with the transition to The Office for Students (OfS). For example, the Access and Participation Plans guidance for 2018 (OfS 2018) asked providers to evaluate their activities robustly to make sure they continue to be effective, and the guidance asks institutions to deliver activities that are: *"demonstrably effective, strategically focused, evidence- led and that clearly address the ambitions set out in your plan"* (OfS 2018, p. 12). This requirement includes providers showing the mechanisms they have developed to enable the outcomes of evaluation to influence practice, both within the organisation and across the sector (p. 24). From the regulator's perspective, they want to see that a provider is able to demonstrate that their Access and Participation plans are driving improvements in outcomes for students – this shows that money is being spent responsib

The remit of the OfS as a regulator includes an ability to require clear and impactful Access and Participation plans. Furthermore, under its section 35, OfS has a remit to identify and share good practice. Providing evaluation resources is one way of doing this so that providers have access to guidance and toolkits on how to improve practice.ly and with impact. Better evaluation is thus one of the mechanisms to improve practice in widening participation outreach, to ensure efficient targeting of resources, and ultimately, to support more disadvantaged students to enjoy HE access, success and post-graduation progress. However, OfS has been mindful to not only ask for enhanced evaluation and demonstrating impact but has simultaneously increased their support offerings. For example, the Access and Participation Plan advice on good practice includes practical suggestions such as building evaluation into activities from the start; demonstrating that the evaluation methods are focused on impact in terms of demonstrable changes in behaviour (such as improved access, continuation, attainment, progression to postgraduate study and graduate employment) rather than, for example, solely gathering opinions from students (OfS, 2018).

One of the predecessors to OFS, the Office for Fair Access (OFFA), commissioned in 2016 a project The *Standards of Evidence* Phase 1 project led by researchers at the University of Warwick, which resulted in reference framework of three types of impact evaluation work (Crawford et al. 2017). The aim of the framework was to facilitate robust and rigorous impact evaluation by providing a shared reference framework across HEPs. This work reported in 2017 and created a typology of evidence of three categories of evaluation:

I. The HEI can provide a narrative to motivate its selection of outreach activities in the context of a coherent outreach strategy;

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

3. The HEI has implemented an evaluation methodology which provides evidence of a causal effect of an intervention.

The aim of *the Standards of Evidence Phase 2* project discussed in the present article was specifically to take forward the previous work on the *Standards of Evaluation* with a brief to pilot test how the standards would be received by evaluation practitioner and to practically support effective evaluation of the impact of outreach interventions on access to higher education. The involvement of a range of partners with differing levels of evaluation knowledge and implementation strategies served the purpose of understanding the impact of evaluation approaches in relation to the *Standards of Evaluation Practice.*

Research Methods

An action research approach was adopted, designed to test how the *Standards of Evaluation Practice* were being applied in different contexts with a view to elucidating good practice that could be shared. By having academics and practitioners jointly researching, they can undertake ‘learning by doing’ with deep reflection combining inquiry with action as a means of *“stimulating and supporting change and as a way of assessing the impact of that change”* (Burns, 2007, p 11). Action research is an ‘orientation to inquiry’ rather than a methodology as such, since it is *“based in different relationships, and it has different ways of conceiving knowledge and its relation to practice”* (Reason, 2003, p 106). The approach exhibited the core characteristics of action research identified by Greenwood and Levin (1998), namely: context bound and addressing real-life problems; participants and researchers contributed to knowledge through collaborative communication processes in which all participants’ contributions were taken seriously; treating the diversity of experience and capacities within the local group as an opportunity for the enrichment of the research-action process; involving reflections on action which led to the construction of new meanings (Greenwood and Levin, 1998, p 93). Ethical approval for the research was granted by the University of Exeter.

Table 1 provides an overview of the partners, the type of outreach intervention they brought to the partnership and the range of impact measures represented in individual projects. The reason for the involvement of HEPs and third-sector organisations was firstly to test the *Standards of Evaluation Practice* for different organisational contexts and for different outreach activities to help establish the evidence on what works in different geographic and institutional settings.

Table 1: Overview of project partners

Partner Institution or Organisation	Nature of Intervention	Type(s) of evaluation measures used
Royal Northern College of Music Small Specialist Institution	Target groups working alongside practitioners and RNCM students in musical theatre project	Skills development (music, drama) Transferable skills development Applications to HE
University of Liverpool* Russell Group HEI	Range of activities with increasing focus over time on preparation for university life	Completers HE applications HE destinations Performance as a Liverpool student
Loughborough University** Pre-1992 HEI	Sports intervention involving coaching by current students and celebration activities	Understanding of entry requirements Enrolments in HE
Coachbright Third sector	Coaching to help pupils improve their attitude, work ethic and academic performance	Academic Grades Confidence Expectations University Access
Brightside* Third sector	Access to high quality information and personal support including mentoring	Behavioural Outcomes: Self-efficacy; Hope; Growth mindset; Coping Capital Outcomes: Human capital; Social capital
The Access Project Third sector	Intensive, long-term intervention aiming at access to the top third most selective universities.	Student achievement Y11 GCSE attainment Y13 students' university outcomes vs a control group
The Sutton Trust Third sector	UK Summer school programme.	Application to HE Acceptance in HE Enrolment in HE Completion in HE
NCOP – First Steps Southwest* Post-1992 HEI	Pilot intervention to support year 13 students from low progression neighbourhoods	Application to HE Acceptance in HE Enrolment in HE
Selected to road-test material: University of Exeter* Russell Group HEI	Supported access project which offers a two year programme of activities	Applications to HE, acceptances and enrolments (data not matched at individual level)

member of the Higher Education Access Tracker (HEAT) s; **member of East Midlands Widening Participation Research and Evaluation Partnership (EMWPREP). Both Heat and WMWPREP help WP practitioners monitor and evaluate outreach delivery by sharing practice, tools and resources; and providing tracking young people over time.

The action research undertaken with partners was rooted in the “plan, act, observe, reflect” sequence elaborated by the ‘Kolb cycle’ (Kolb, 1984). The partners were each assigned one member of the research team to provide an ongoing point of contact and conduit for promoting action and observation of the implications of developments in evaluation practices. Researchers also played a role in drawing out findings and offering support to partners where appropriate to taking forward the objectives.

To start, partners came together and we mapped the existing evaluation approaches to the three types outlined in the standards of evaluation for outreach. Out of our nine partners, one third-sector partner was undertaking a Type 3 evaluation. Two partners were at the early stages of setting up evaluation and were working towards a Type 1 approach in the first instance. The remaining five partners were working within Type 2 evaluation considerations. Once the ‘starting point’ for working with the standards was established, interactions between researchers and each partner were established in order to test operationalizing improvements on the ground to ratchet up evaluation practices against the standards framework.

Innovations to partners’ evaluation approaches were agreed and operationalised locally. Various examples, of local innovations in outreach evaluation took place during the project period, and included the following themes: identification and development of skills required for evaluation and matching to partner staff; putting in place collaborative evaluation between academics and practitioners; establishing systematic evaluation review cycles; linking to students as co-creators of outreach; conceptualisation/testing of theories of change underpinning interventions; improvements to data processes and analysis. The amount of, and nature of, the contact between the researchers and the practitioner staff members in the varied in each case according to the differing partner circumstances and the stage of the outreach evaluation work. The work with partners afforded the researchers insights into the institutional/organisational challenges involved in embedding evaluation into normal routines of professional work for widening participation; learning how institutions can incorporate data analytics; exploration of transferability and scalability issues.

All the partners met together three times during the project as well as having individual contact with the researchers outside these meetings. All work with partners took place from autumn 2017 and throughout 2018. These meeting were used as the basis for group reflection on the implications of the evaluation practices they had undertaken and to improve evaluation practice against the frameworks of the standards of evaluation.

In addition, quarterly steering group meetings provided an opportunity for expert input in shaping the work and a strategic discussion of the direction of travel. Through the iterative process of working with partners, the OfS and the steering group, it was decided that an additional useful project outcome would be a way for HEPs to gain knowledge of the rigour of their approaches to evaluation mapped against a framework of evaluation dimensions. This led to the development of a self-assessment tool for evaluation HEPs could use for understanding, mapping and enhancing their practices. This tool was piloted in a series of two workshops in autumn 2018, these workshops were attended by over 10 providers selected to present the variation of HEPs in the sector, i.e. including large and small institutions and established evaluators as well as those newer to evaluation. These workshops fed back into the development of the final self-assessment tool for evaluation.

Findings

Even amongst the small group of cases included in the project, a key conclusion centred on the challenge of applying the same quality standard uniformly to different kinds of research setting.

Shifting the framework to horizontal types

Working with our project partners, we found that having a shared evaluation framework was considered useful by practitioners and that they saw the potential for it to offer a reference point and shared language that would facilitate discussions around evaluation and enable practice enhancement. However, project partners also felt that the hierarchical ordering of levels was considered less useful. This feeling was echoed in the discussion in the steering group organized by OfS. In particular, partners and the steering group were concerned that the ordering would result in different providers aiming for a level 3 evaluation (quasi-experiments or randomized controlled trials). This seemed like an unrealistic aim in light of the methodological requirements for undertaking such work well and the resource and intervention contexts of many providers. In one case evaluation team had recourse to well developed methodological standards for how to conduct a randomised controlled trial, the case shows the importance of understanding the delivery context and processes involved in the provision of any outreach activity to understanding what works (as well as thinking about the outcomes for the participants involved). This example forewarns of the challenges of undertaking RCTs in the real word. As a result, there is a risk that many RCT in the context of outreach are not reaching the high standards of e.g. medical randomised trials since participant engagement and control conditions cannot be so rigorously maintained. There were reports from our partners about using RCTs for e.g. measuring the impact of text messaging – this is an example where RCTs work well as a research design, but only allow asking a relatively simple research question. The literature highlights that RCTs are usually high cost, only enable slow learning, and often deliver inconclusive results (Chambers, 2017). Weaker experimental designs have been shown to be more likely to spuriously find that an intervention is effective because the intervention is being confounded with other influences on the outcome. Particular dangers are noted with interventions in situations that are complex, multi-dimensional, uncontrollable, unpredictable, and idiosyncratically variable (Chambers, 2017, p.65). Even where the internal validity of an RCT is stronger, the results may have only weak external validity (Frieden, 2017). Moreover, in general the results of RCTs provide insufficient data to be able to say 'what works' (Deaton and Cartwright, 2015). Therefore decision between applying an RCT and other methods (including mixed-methods) should not be taken lightly, since the latter might have more chance of generating useable results.

Moving from the hierarchical types to horizontal types allowed partners to discuss the stage of their evaluation journey on equal footing based on a shared understanding that the rigour of evidence was key across all dimensions.

Further discussions with partners, OfS and the steering group then allowed us to re-frame the three levels of evaluation work as non-hierarchical 'types'. The key point here is that RCTs remain one way of gaining causal evidence and they provide type 3 evidence. However, the types recognise that an RCT approach is not always appropriate or possible. The final model developed by this project then uses three types of evaluation which generate different types of evidence: (1) the narrative of evaluation; (2) empirical enquiry; and (3) consideration of causal claims. Table 2 details the three types:

Table 2: The Standards of Evaluation framework in three non-hierarchical types

Type	Type 1	Type 2	Type 3
Name	Narrative	Empirical Enquiry	Causality
Description	The impact evaluation provides a narrative or a coherent theory of change to motivate its selection of activities	The impact evaluation collects data on impact and reports evidence that those receiving an intervention have better	The impact evaluation methodology provides evidence of a causal effect of an intervention

	in the context of a coherent strategy	outcomes, though does not establish any direct causal effect	
Evidence	Evidence of impact elsewhere and/or in the research literature on access and participation activity effectiveness or from your existing evaluation results	Quantitative and/or qualitative evidence of a pre/post intervention change or a difference compared to what might otherwise have happened	Quantitative and/or qualitative evidence of a pre/post treatment change on participants relative to an appropriate control or comparison group who did not take part in the intervention
Claims practitioners can make	We have a coherent explanation of what we do and why Our claims are research-based	We can demonstrate that our interventions are associated with beneficial results.	We believe our intervention causes improvement and can demonstrate the difference using a control or comparison group

In framing the standards as non-hierarchical, the framework can focus on the quality of the evidence created in the evaluation and robustness of evaluation strategy and appropriateness to the activities practitioners are undertaking. It is possible to make different judgements about the impact of access and participation activities from the different types, so long as the evidence from the evaluation is robust, high-quality evidence. While the types are on a continuum in the sense that undertaking a good 'Type 2' evaluation would be based on having a theory of change or logic model (a type 1 evaluation), the emphasis on the non-hierarchical ordering is to do evaluation well rather than racing for the highest 'type'.

A thinking framework for practitioners: developing the self-assessment tool

The Cases illustrated a range of helpful practices that strengthen decisions on evaluation practice and support appropriate prioritisation and resourcing of evaluation, the application of appropriate methodologies and an environment whereby learning benefits for local and national policy can result. For example, in the case of a small specialist provider, making improvements to evaluation quality depended on the ability to conduct robust research with limited funding and simultaneously produce reliable baseline findings for future evaluations. Identifying the skills for evaluation, and matching to the existing skillset of staff and responsibilities assigned to ensure effective evaluation was crucial here. On the other hand, for a large Russell Group institution, the main challenge involved planning cycles of evaluation activity that build year-on-year through consecutive rounds of professional reflection, evaluation activity and programme review. Making progress on evaluation quality in this context involved embedding systematic focused and purposeful evaluation to inform decision-making. In another case, identifying measures to capture the benefits of interventions emerged as the main issue: signalling a process of learning to make sure that the indicators capture well-defined changes and are relevant to the objectives. Alternatively, for another provider the key weakness was in terms of the implementation of the data collection methods, and a fifth was challenged by coordination across a wide group of stakeholders in a partnership. Here the challenges were practical in nature, including difficulties in ensuring strategic buy-in from all parties involved, establishing clear communication channels, and managing attrition due to drop-out.

The original Phase 2 project plan had intended to develop examples of best practice as a key way of sharing evaluation expertise across the HE sector. In addition, the iterative process of discussions between researchers and partner organization and especially the peer-dialogues among partners during the discussion of the Standards of Evaluation framework highlighted the complexity of good and best practice in different contexts. Simultaneously, it became clear that challenges and concerns were shared among partners. They considered a particularly useful aspect of the project to be learning they could share among each other. This included highlighting areas for practice or policy development that especially those newer to evaluation might not have initially considered. But more experienced evaluators also benefitted from considering what the next step in achieving rigour would be from their current starting point. In these discussions lay the seed of the idea for a further ‘thinking framework’ that would support evaluation practitioners in holistically considering their practice and detailed aspects of evaluation.

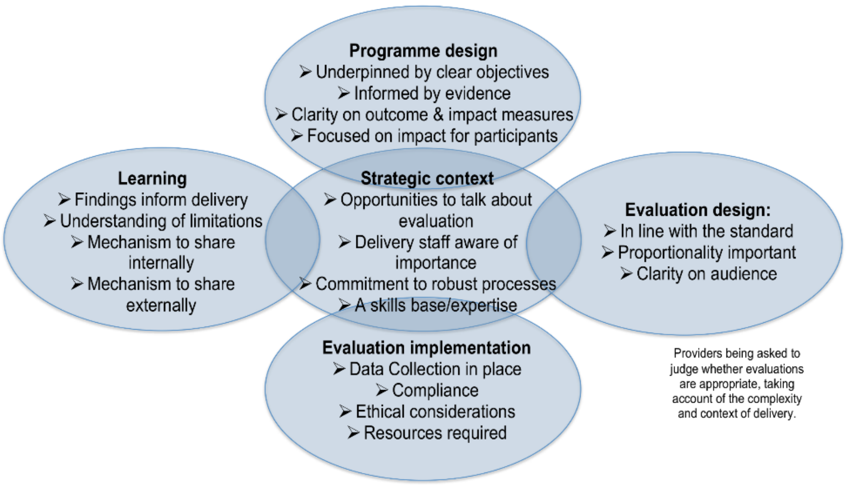
Iterative discussions then continued between the research team, the OfS, the steering group and again with the project partners. Consensus emerged that practitioners would find it useful for additional support material to be available alongside the Standards framework. Simultaneously, the OfS as the HE sector regulator, was keen to enhance its ability to access details about the rigour and level of evidence of evaluation practices mapped against the various dimensions of evaluation. This would add further value to the categorization of HEPs within the Standards framework of the three types of evaluation practices. In light of the shift to a non-hierarchical typology, the emphasis was also now on quality within each of the three types.

These thinking processes then led to the development of an initial framework of ‘evaluation dimensions’ that the project team designed based on two types of discussions: an academic and policy interrogation of the literature (including e.g.

BERA 2018, Haynes et al. 2012, EEF (undated)) and generic evaluation frameworks and ongoing discussions with partners of what was relevant and important in their HE widening participation evaluation practices.

A thinking framework with a range of dimensions was emerging. Further discussion with the steering group and guidance from OfS then created a working version of a framework that practitioners could use to map their own practices. The format was increasingly confirmed as a self-assessment tool that would have a scoring / quantitative element where HEPs would give themselves marks regarding how well developed their practice in a particular area was. Each area included the types of evidence HEPs might consider in supporting their own mark, in addition, there was space for a narrative to explain how judgement on the marks had been made. Areas of evaluation practice were grouped within five dimensions as illustrated in Figure 1.

Figure 1: The five dimensions of the Self-assessment tool



Specifically, the research identified five dimensions that evaluators need to ask about the relationship of the evaluation to the programme context:

Dimension 1: Strategic context
At strategic level it is important that roles and responsibilities in terms of agreeing, implementing, managing and learning from evaluation will be in place from the start, alongside formal structures to ensure that the delivery of outreach develops in a way that responds to evidence about what works best in their particular context to maximise the outcomes and impact of the access and participation investment. Key questions include: To what extent is an evaluation

culture supported and prioritised, including as part of a coherent programme of evaluation across different activities? Are opportunities for staff to enhance their evaluation skills and understanding?

Dimension 2: Programme design
The underpinning theories of change for interventions matter to the quality of the evaluation not only because this can help in identifying the planned outcomes of an intervention, but also because the theory of change identifies the comprehensiveness with which the formulation of the intervention has considered the precise processes and mechanisms involved in bringing about the desired outcomes. Key questions for this dimension include: drawing out the programme theory with respect to the intervention; ensuring the programme design and choice of outcome measures are underpinned by and informed by the existing evidence; and building in evaluation at the design stage of interventions (rather than as an after-thought).

Dimension: 3: Evaluation design
The robustness of evaluation design is clearly central to achieving high standards. Overall, the evaluation design factors need to ensure that the evaluation has internal validity and is capable of demonstrating the direction of association between the intervention and the outcome. However, judgment on the quality of the method may depend on what you see as the main aim of the evaluation – for example whether to test the effect of an outreach intervention on an outcome or to refine the programme theory underpinning the approach? Narrative and empirical approaches, rather than experimentation, can prove of value in generating hypotheses about the context-mechanism-outcome relationships (although are obviously of less use in testing a causal hypotheses). The availability of existing evidence of the impact of different activities may affect which type of evaluation approach is most suitable: the highest standards of evaluation are needed where there is currently little or no convincing evidence that the activity is effective whilst activities that have already been proven to work may not require such rigorous testing, bearing in mind that there is currently a dearth of rigorous high quality evidence on the impact of access and participation programmes. Key questions for this dimension therefore include: whether the research approach matches the purpose of the evaluation; whether the evaluation is proportionate to the activity and appropriate for the purpose of measuring the impact and the aims of the stakeholders; and making sure evaluators are aware of the strength and weaknesses of the approach taken to generating evidence of impact.

Dimension 4: Evaluation implementation
The credibility of the evaluation is often determined in the execution of it, in conjunction with practical considerations such as the measures of outcome variable used, the timing of the evaluation activities, and adequacy of sampling (the later potentially depending on the scale of the provision). Findings that might

be potentially highly relevant to policy are not useful if the methods of implementation are dangerously flawed, however, there is a sense that widening participation evaluators are working with operational and methodological constraints, which need to be assessed and understood. Key considerations include: Is the research technically well executed? Are the outcome measures and data collection tools reliable? To what extent are reliable data collection mechanisms in place? Are the resources for the evaluation appropriate to the task?

Dimension 5: Learning from evaluation
As well as the relevance of the evaluation to the aims, in seeking to drive up the contribution of evaluation to outreach policy and practice, the spotlight of interest also turns to the quality of the transmission of the results of evaluation. For example, by considering the ways in which evaluations are communicated to potential internal and external audiences. Key questions for this dimension included: the extent to which evaluation is presented in a way that can be appraised and used by others; where mechanisms are in place to ensure that the evaluation findings and results are used to inform practice through reflection, sharing, dissemination, translation into adjustments or innovation in practice.

The new self-assessment tool was road-tested in two events attended by 9 HEPs representing various sizes, and experiences (including those who recently attained degree awarding powers) in Autumn 2018.

This feedback and further discussions and guidance from OfS resulted in fine-tuning and tweeking the final self-assessment tool and accompanying guidance documents. Table 3 then provides a sample overview of how the detailed dimensions within an area of activity and how HEPs would use it.

Table 3: Excerpt from Dimension 4 of the finalised self-assessment tool

Dimension 4: Evaluation implementation This dimension involves thinking about how you put in place the measures and tools you are using to evaluate and other implementation factors such as the reliability of data collection mechanisms and application of resources.			**Does it exist?** Response (choose from list)	**How is it demonstrated?**	**Notes** (use this space to explain your answers and the scores you have given yourself)
Data collection	Expected	Have you identified how you will access the data required to measure outcomes and impacts?		You can point to reliable and robust data sources used to collect data on the outcomes (changes) you are making and the impact (the difference this makes for higher education access and participation). This could include the application of qualitative or quantitative research methods to collect new data or the or use of existing data sources where relevant (e.g. higher education progression data). Examples could include having a schedule in your evaluation frameworks which sets out when and how data will be collected.	
	Commended	*Do you work in partnership with other stakeholders (e.g. schools, data providers) to maximise evaluation data and results?*	SELECT Select the response which best describes your organisation from the drop-down list.	*Examples could include data sharing protocols being put in place.*	
	Expected	Does your approach to data comply with the requirements on data collection and data sharing?		For example, an audit of existing administrative and naturally occurring data used and assessment of compliance against current data protection legislation requirements and good practice.	
	Expected	Are procedures in place for addressing ethical considerations?		Use of an agreed research protocol. Approval through your institution's ethnical approval process.	

The self-assessment tool is part of a package of three outputs created as part of the present project. All material is designed to assist HEPs in continuously improving their evaluation practices and showing how their evaluation strategy is evolving: (1) Guidance on the Standards of evaluation in outreach; (2) a good practice guide to using the OfS Standards of Evaluation in outreach and (3) the above discussed self-assessment tool of the Standards of Evidence.

Document 1 lays out the standards of evidence for senior managers, decision makers and practitioners with a remit for evaluation and reporting on access and participation activities. The aim is to promote understanding of the standards of evidence and more rigorous approach to undertaking and using impact evaluation to improve the effectiveness of the investment in access and participation programmes. Document 2 is the guidance for people who already have some experience with evaluation techniques and are looking to make impact evaluations more robust and embedded. The document highlights practices that can strengthen the impact evaluation of outreach and offers case studies and signposting to further resources. This document also contains the academic and policy references consulted as part of the project, these are embedded thematically within the document at the points practitioners might wish to access them. All documents are available from the OfS website (OfS 2019a, OfS 2019b, OfS 2019c).

Conclusion

The ultimate goal of widening participation work, that unites policy makers, practitioners and academics, is to enhance outcomes for disadvantaged students in HE access, success and outcomes. OFFA and OfS, in commissioning this research, sought to create a step-change in robustness of evidence used and evaluation practices. The academic researchers undertaking this work were motivated by the possibility of translating academic knowledge into an actionable and useable resource for practitioners. The partner institutions and practitioners wished to share their challenges and success in evaluation and learn from each other and everyone involved in the project how to enhance their practices and the evidence and rigour of their work. The steering group members generously donated their time and expertise to maximise the usefulness of the work in building evaluation capacity and in being a meaningful resource for evaluation practitioners in HE as well as a tool for OfS as the sector regulator.

The action research method used for this project was to highly relevant as a the method seeks to explore the 'real life' challenges to operationalizing objectives relating to improving standards of widening participation evaluation practice, considering the highly varied, complex and stratified nature of the contextual factors involved. The cases underline the diversity of research practice, the constraints to delivering evaluation, activities and issues of access to both formal and informal quality assurance procedures. Understanding the underlying structural issues was crucial before outreach providers got to a position firstly to improve their evaluation approaches and secondly to apply any learning from evaluation. Whilst the project initially set out to operationalize a set of externally defined standards, in context, the method took account of the need for a flexible development process with a focus on ongoing improvement over time. The approach highlights the criticalness of establishing the starting point and creating the base from which the existing practices can then be challenged and providers stretched to improve incrementally. As an outcome of the project and the ongoing discussions and collaborations between everyone involved in the project, the actual project objective shifted from capturing best practices in evaluation to a focus on providing a framework that would enable good practice to be developed and provide a shared reference point for practice as well as shared language for talking about dimensions and common themes in widening participation evaluation.

The main challenges for the project were the need to work within the time and resource constraints of partners, and within the project timeline - which although relatively long term was still limited in respect of the time frames usually involved in getting meaningful outcomes/results from new outreach evaluations in terms of HE progression. There were also challenges in terms of negotiating General Data Protection Regulation (GPDR) compliant data sharing licenses between the consent holders and the academic researchers. Specifically, some initially planned data analyses activities proved not to be possible due to problems in sharing of student outcomes data.

The close engagement OfS maintained of the project allowed the final outputs to be a useable tool. OfS decided that the self-assessment tool and accompanying guidance would go out to all HEP in 2019 and that HEPs would be invited to return the completed tool and their narratives about their state of evaluation practice as part of the Access and Participation Plan approval process. While OfS cannot mandate the use of the tool, they have made it an expectation that HEPs undertake some form of self assessment when developing their evaluation strategies and the tool has been provided to help them do that. Many providers have chosen to submit their completed tools to the OfS.

Outreach teams do not always apply evaluation resources and expertise strategically and systematic mechanisms for reflection and review are not always in place, and together with methodological and practical challenges, this undermines achievement of standards in evaluation. The guidance and self-assessment tool aims to support this. Further future analysis of the completed self-assessment tools will illuminate to what extent the tool is able to capture activities and highlight areas where institutions and practitioners can enhance their practices and increase their capabilities for increasing the evidence and rigour of their evaluation practices.

The self-assessment tool itself is now part of the cycle of continuous evaluation and improvement of practice. Now that it has been implanted for the Access and Participation plans for 2019, its impact will need to be evaluated and reflected upon. If continued, feedback and reflections will lead to further development of the tool. OfS is planning to host a reunion and reflection time for everyone involved in the research to offer this space for reflection on the past and future of the self-assessment tool for evaluation practice created during this project. This ongoing work will hopefully maximise the utility and impact of the tool and support ongoing discussions about enhancing evaluation practices in widening participation evaluations to benefit prospective and current students in higher education and build a rigorous evidence base to inform OfS policy and regulation in the future.

The development of the self-assessment tool co-incided with the development of the invitation to tender for the 'knowledge and impact exchange', launched in summer 2019 as Transforming Access and Student Outcomes (TASO). This new body is tasked with enhancing evidence-based practice through collating and sharing knowledge of 'what works'.It is hoped that as providers use the self-assessment tool to identify areas to improve evaluation practice, this will result in better quality evidence that can be submitted to TASO' .

Funding
This article is based on work commissioned by what was the Office for Fair Access (OFFA) in 2017 and completed under the auspices of the newly created Office for Students (OfS) in January 2019.

Acknowledgements:
We would like to acknowledge our project partners, our project steering group and support at the University of Exeter. We would like to thank our project partners for developing this project and sharing their knowledge: Royal Northern

College of Music, University of Liverpool, Loughborough University, Coachbright, Brightside, The Access Project, The Sutton Trust, NCOP – First Steps Southwest. Paul Wakeling (University of York) effectively chaired the steering group which also include Andrew Berwick (The Access Project); Lindsey Bowes (CFE); Lee Elliot Major (The Sutton Trust); Peter Riley (Manchester Metropolitan University); Jon Beard (University of Cambridge), Neil Harrison (then University of West of England) and Alison Matthews (University of Oxford). At the Office for Students, Richard Shiner, Head of Evaluation, Alex Wardrop and Cassie Agbehenu. At the University of Exeter, Debra Myhill played a key role in assembling the project team and project partners and developing the research plan. Pallavi Banerjee supported the work with project partners in the early stages. Nicola Sinclair and Samantha Dyer as Heads of Widening Participation and Evaluation Manager respectively, piloted the emerging self-assessment tool and hosted a South-West wide dissemination and practice sharing event.

References

British Educational Research Association [BERA] (2018) Ethical Guidelines for Educational Research, fourth edition, London: BERA

Bowes, L., Thomas, L. Peck, L. and Nathwani, T. (2013) International Research on the Effectiveness of Widening Participation. Bristol: HEFCE.

Burns, D. (2007), Systemic action research: A strategy for whole system change. Bristol University Press, Policy Press.

Chambers, R. (2017) Can We Know Better?: Reflections for Development, Rugby: Practical Action Publishing.

Crawford, C., Dytham, S. Naylor, R. (2017) The Evaluation of the Impact of Outreach Proposed Standards of Evaluation Practice and Associated Guidance, Office for Fair Access

Deaton, A. and Cartwright, N. (2018) Understanding and misunderstanding randomized controlled trials, Social Science & Medicine, Volume 210, August 2018, Pages 2-21

Educational Endowment Foundation [EEF], (undated) DIY Evaluation Guide https://educationendowmentfoun dation.org.uk/tools/diyguide/getting-started/

Frieden, T., R. (2017) Evidence for health decision making—beyond randomized, controlled trials, New England Journal of Medicine, 377 (2017), pp. 465-475.

Gorard, S., Smith, E., May, H., Thomas, L., Adnett, N. and Slack, K. (2006) Review of widening participation research: addressing the barriers to participation in higher education. Bristol: HEFCE.

Greenwood, D.J. and M. Levin (1998). Introduction to action research: social research for social change. Thousand Oaks, CA: Sage.

Haynes, L., Service, O., Goldacre, B. and Torgerson, D. (2012) Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. London: Cabinet Office - Behavioural Insights Team.

Moore, J, Sanders, J, & Higham, L, (2013) Literature review of research into widening participation to higher education. Bristol: HEFCE.

Office for Students (OfS) (2018) 'Regulatory Advice 6: Good practice advice on the preparation of access and participation plans for 2019-

20'https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/access-and-participation-plans/

Office for Students (OfS) (2019a) Access and Participation Standards of Evidence, available at https://www.officeforstudents.org.uk/publications/standards-of-evidence-and-evaluating-impact-of-outreach/

Office for Students (OfS) (2019b) Using standards of evidence to evaluate impact of outreach, available at https://www.officeforstudents.org.uk/publications/standards-of-evidence-and-evaluating-impact-of-outreach/

Office for Students (OfS) (2019c) Evaluation self-assessment tool, available at https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/evaluation-and-effective-practice/standards-of-evidence-and-evaluation-self-assessment-tool/evaluation-self-assessment-tool/

Reason, P. (2003), Pragmatist philosophy and action research. Action Research Volume 1(1): 103–123. Thousand Oaks, CA: Sage.

Sutton Trust (2015), Evaluating Access, Research Brief Edition 9: December 2015, London: Sutton Trust.

Torgerson, C., Gascoine, L., Heaps, C., Menzies, V. and Younger, K. (2014), Higher Education access: Evidence of effectiveness of university access strategies and approaches: A report to The Sutton Trust, Durham: Durham University.