Evaluative Practice

EDITION 03

AUTUMN 2024



PUSHING THE BOUNDARIES

UK Evaluation Society

evaluation.org.uk

Contents

A note from the President 03
Kirstine Szifris

Executive Director 04
Nick Posford

Editorial 05
Murray Saunder

Celebrating 30 Years of Evaluation: Reflecting on the Past, Shaping the Future 07
Roo Cleland UKES

Learning Together: Evaluation as participatory sense-making 09
Emma McDowell

Better Evaluation Using Evidential Pluralism 12
Michael Wilde, Jo Jones, Alex Trofimov, Jon Williamson

Policy Evaluation: A Critical Friend or Foe? 15
Charlie Heywood-Heath

What is a method? I'm asking for a friend (of Wittgenstein's) 17
Barbara Befani

What should we measure in programme evaluations: Insights from psychology 20
Marina Bazhydai

Putting Children and Young People at the heart of evaluations 24
Jessica Ozan and Jamie Rowland

Would you like to contribute to Evaluative Practice? Check out notes for contributors via our website: ***evaluation.org.uk***

A note from the President
Kirstine Szifris

As we come towards the end of 2024, I am also coming to the end of my first year as President of the Society. The year started with recruiting our permanent Executive Director, Nick Posford, who now leads a small but growing staff team in supporting our activities across a range of areas. This has been an essential step towards becoming a self-sustaining organisation that is well resourced to ensure that our members are well represented and given plenty of reason to remain engaged in the organisation.



KIRSTINE SZIFRIS – UK EVALUATION SOCIETY PRESIDENT

This year marks our 30th and we began our celebrations in Cardiff with a wonderful birthday event. With contributions from all sectors at the world's first social science research park, who kindly hosted us, it was an opportunity for members old and new to discuss where Evaluation should go in the next few years.

As the Society's President, my focus is on what we do now to both build on the successes of the last 30 years and create the foundations for the next 30. As we look to the future of evaluation, UKES is primed to make a significant impact by providing support, training and advocacy. We are committed to leading and contributing to meaningful discussions, tackling challenges that lie ahead, and seizing opportunities to shape the future of evaluation.

With a strong government emphasis on evaluation through the Evaluation Academy of the UK government and the forthcoming Centre of Expertise in the Scottish government, the Society has an important role to play in ensuring evaluation is better resourced and that new and innovative approaches are part of the mix. In early 2025, we will build on the success of the Cardiff conference with two more events. The first, a special discussion with the Foreign Office, will explore whether diplomacy can be evaluated. We will then be in Edinburgh, for an event led by the Scottish Government's Centre of Expertise on Appraisal and Evaluation of Spending Programmes, where we will hear from key speakers including the Chief Economist and the Chief Social Researcher.

In April, we held elections for the first time in the society's history and successfully appointed five new council members. In the coming months, we hope to recruit five more members to join the council. Our aim is to strengthen and diversify the skills and experience of the council as we enter a new era as a registered charity.

These changes mean we are better positioned to grow our existing connections with the voluntary, public, and private sectors, enabling us to collaborate with like-minded organisations, in the UK and abroad. From emerging technologies like AI to the transformative power of big data, we aim to listen, learn, and play a central role in shaping conversations about the future of evaluation. As evaluation practices evolve globally, it's vital for UKES to serve as a guiding light – supporting our members and the wider evaluation community in navigating these changes to be a powerful, unified voice.

I'll end by saying how excited I am about our annual conference returning in May 2025, in Glasgow, focussing on the most important ingredient of evaluation – high-quality data. Keynote speakers will be announced soon, and our call for abstracts remains open until early in the new year. I look forward to welcoming you to what promises to be the UK's largest annual gathering of evaluators and if you haven't booked your ticket yet, there is still time to secure one at early bird rates – and remember, members enjoy significant discounts too!

Nick Posford
Executive Director



NICK POSFORD – UK EVALUATION SOCIETY EXECUTIVE DIRECTOR

It's exhilarating and a real privilege to lead a team of very dedicated staff in the UK Evaluation Society's transition to a charity. We have focussed on improving all the elements which make for a pleasant member experience – a new website, a new membership database, and a new role of Marketing and Communications Manager, introduced to ensure everyone who can benefit knows about our services. This role also supports the Society's strategic communications, with our election manifesto highlighting the role that evaluation plays in effective government policy.

Following approval of the new Articles of Association at the Extraordinary General Meeting (EGM) in July, our registered charity status is imminent. This move will have a positive impact on the organisation and introduce extra benefits for our members and for the wider evaluation community – including no longer having to charge VAT on training and conference tickets.

Further, we have been added to HMRC's list of approved professional organisations and learned societies. This means that individual members who pay subscriptions can now claim tax relief and serves to further establish the profession of evaluation for the community.

To ensure we deliver expert education in evaluation, we've commissioned a comprehensive review of our training. This ensures we meet the needs of everyone—from those just beginning their careers to future leaders in the field. Members now enjoy priority booking for training courses, allowing them to secure places on popular sessions that often sell out. With the support of the London & SE Evaluation Network, we funded bursary places for students this year. Building on this model, we aim to widen access to training and professional development opportunities in 2025. Members also receive our in-house publication, Evaluative Practice, and benefit from discounted access to the leading Evaluation journal.

To celebrate our 30th anniversary, we opened up our webinars to everyone, while maintaining member-exclusive benefits such as priority access to recordings and slides. Free webinar topics ranged from arts-based data collection methods, the Ethnographic Explorer, and evidential pluralism to insights into working in the evaluation industry, common evaluation flaws, the Most Significant Change technique, and an introduction to realist evaluation.

Our collaborations have spanned a wide range of organisations and experts. We worked with the Global Development Network, Youth Futures Foundation, philosophers from the University of Kent, the Young Foundation, AND Implementation (Canada), and many others. We're closing the year with a special webinar on AI, delivered in partnership with the Social Research Association and the Academy of Social Sciences. Looking ahead, we'll collaborate with the Policy and Evaluation Research Unit (PERU) at Manchester Metropolitan University on an event dedicated to Relational Evaluation.

We are also looking at how we make the most of income opportunities, alongside supporting the evaluation community to recruit and retain top-performing staff and giving all evaluators the insights and network connections they need to excel in their role. Ultimately, we believe in a world of limited resources and complex challenges, understanding the impact and outcome of policies, programmes and interventions is critical. Robust evaluation gives us the clear, reliable information we need to make smart decisions, find the best solutions and help organisations and policymakers make better choices to benefit people and communities.
That is why we continue to drive innovation and excellence, helping evaluators at all levels and stages in their careers achieve their goals and contribute to meaningful, positive change.

Murray Saunders
Editorial



MURRAY SAUNDERS – UK EVALUATION SOCIETY COUNCIL MEMBER

We are delighted to present the Autumn 2024 edition of Evaluative Practice and to provide an opportunity to continue the celebration of 30 years of the UKES by updates and messages from Kirstine Szifris the UKES president and Nick Posford UKES Executive Director. They both position the UKES at this exciting moment and give an indication of the direction of travel of the society with the benefits of charitable status and the return of our annual conference, along with a host of training and other networking and learning events.

I am grateful for the contribution of Roo Cleland who was our on-the-spot reporter at the recent anniversary conference at Cardiff University's Social Science Park. She offers a flavour of the occasion for those who could not attend with speakers and round table exchanges from both new faces and those 'veterans' of the society who have witnessed some profound changes in the evaluation environment and in the fortunes of the UKES over the last 30 years.

The contributions to this edition are diverse but all adopt an expansive tone pointing to the way the boundaries of evaluative practice are expanding, becoming more inclusive and adaptive. All, in different ways, eschew rigidity and suggest the adoption of a 'low fidelity' stance on practice while maintaining rigour and thoughtfulness. Expanding beyond a mono cultural approach to who does evaluation, what might count as a methodological orthodoxy and even what amounts to persuasive and authoritative evidence, we hope the contributions excite argument and encourage debate. This view is thoroughly expressed in the article by Emma McDowell which emphasises the power of collaborative sense making in evaluative practice which, she argues, not only enhances the way evaluations provide shared knowledge resources for social action but also increases the power and insight evaluations might bring.

The article on evidential pluralism from a multi-disciplinary team which includes Jo Jones, Michael Wilde, Alex Trofimov and Jon Williamson offers some cogent examples from the pandemic to suggest that there is an imperative to consider more inclusive approaches to evidence-based evaluation, especially when focusing on the increasingly complex issues which characterise the evaluative environment. They point out that there is a dislocation between this environment and guidance (or lack of guidance) on how synthesising different kinds of evidence might yield better informed evaluations of effectiveness.

Charlie Heywood-Heath argues for a repositioning of policy evaluation in the public mind from an emphasis on accountability and value for money and external judgement to one of a 'critical friend'. This is, in large part, a responsibility of all those stakeholders involved in an evaluation process but in particular he points to the way evaluations are designed and conceived. Much more emphasis, he suggests, should be given to the way evaluation enters the 'use domain'. More time, earlier involvement and an eye on the usability of evaluative outputs are among his propositions.

Barbara Befani's article also challenges received wisdom and is a mischievous prod at the mono cultural approach to methodological purism. In effect she argues for less high-fidelity approaches to methodological choices, in other words, a critique of the situation in which either you take on board the full edifice of concepts and procedures associated with an approach or you "didn't do it properly". High fidelity should give way to low fidelity where an evaluation might be 'informed' by aspects of an approach but not slavishly adhere to every element of, what is often, a rather abstract set of principles, procedures and stages. That said, she also questions metaphors like 'bricolage' which, she says, seem more concerned with distinguishing this approach from others, rather than positively asserting a different set of possibities. Fuzziness wickedness and permeability are all part of her lexicon.

There is a change of gear with Marina Bazhydai's piece in which she analyses the way a disciplinary culture of research and inquiry, in the case she cites psychology contributed to her thinking about and the design of an evaluation. She provides some fascinating insights into the dominant disciplinary thinking associated with psychological research and the internal debates which thrive in a vibrant disciplinary culture. The culture of experimentalism, comparison and control is predominant in her disciplinary universe and she presents a real world (in Jim Rugh's sense, see the Spring 2024 Issue of Evaluative Practice) example in which a control group comparison was made in the evaluation of a creative arts intervention which challenged expectations and caused her to suggest that more 'holistic' methods might be useful.

Jessica Ozan and Jamie Rowland give us what amounts to a manifesto on the way young people might be involved in evaluation. It argues we move from a 'nice to have' to a 'must have' stance on participatory approaches to evaluation involving young people in their evaluative sphere. This position is dissected with an analytical model of the way in which young people are commonly involved in evaluation from youth-initiated engagement to consultation. They identify some important benefits of a series of thorough youth engagements which go well beyond tokenism to challenge received practice.

We hope that these examples of boundary expansion stimulate discussion and possibly ruffle a few feathers. They are offered in the spirit of discussion and debate and we would love to hear reactions or comment.



Celebrating 30 Years of Evaluation: Reflecting on the Past, Shaping the Future

ROO CLELAND
UKES



This year marks a proud milestone for the UK Evaluation Society (UKES)—three decades of promoting the practice, theory, and impact of evaluation across the UK. To celebrate, UKES hosted a special anniversary conference at Cardiff University's Social Science Park, gathering members of the evaluation community for a day of reflection and forward-thinking.

The theme of the conference, Evaluation: Past, Present, and Future, provided a framework for discussions on how evaluation has advanced and what the future holds. The day brought together a diverse cross-section of the evaluation community, including from the voluntary sector, private industry, academia, and the Welsh government.

Professor Chris Taylor, Academic Director of Cardiff University Social Science Research Park (sbarc|SPARK), set a welcoming tone for the day, "This conference is not only a celebration of what the UK Evaluation Society has achieved over the past three decades, but also a chance to reflect on the road ahead." The morning sessions featured presentations from speakers who shared perspectives through the lens of their respective area of expertise. It was fascinating to hear speakers' views and the common threads across the mix of sectors.

The setting at Cardiff University's Social Science Park was perfect for encouraging cross- sector conversations. During lunch, lively discussion could be heard, highlighting the enthusiasm for cross-sector working and ideas for collaboration. Sally Cupitt, Evaluation and Impact Consultant, observed, "Being here brings a shared sense of purpose—there's real momentum in this space for future collaborations."

Feedback from attendees was overwhelmingly positive, with many praising the event's relevance to their own work and how energising it felt to be among peers. "UKES is creating a vision for evaluation that's both ambitious and achievable," said Nick Posford, the Society's Executive Director. "It's about making sure that evaluation remains a powerful tool for creating change, no matter the field."



Professor Steve Martin, Professor of Public Policy and Director of the Wales Centre for Public Policy, reflected on the progress made, saying, "We've seen remarkable growth in evaluation's influence over the past thirty years, but our work is far from over." Similarly, Professor Chris Fox, Professor of Evaluation and Policy Analysis at Manchester Metropolitan University, emphasised the need for a broader range of evaluators, "The profession is evolving, and our practices must evolve with it. There needs to be more roles for people with lived experience and more professionals engaging and undertaking evaluations."

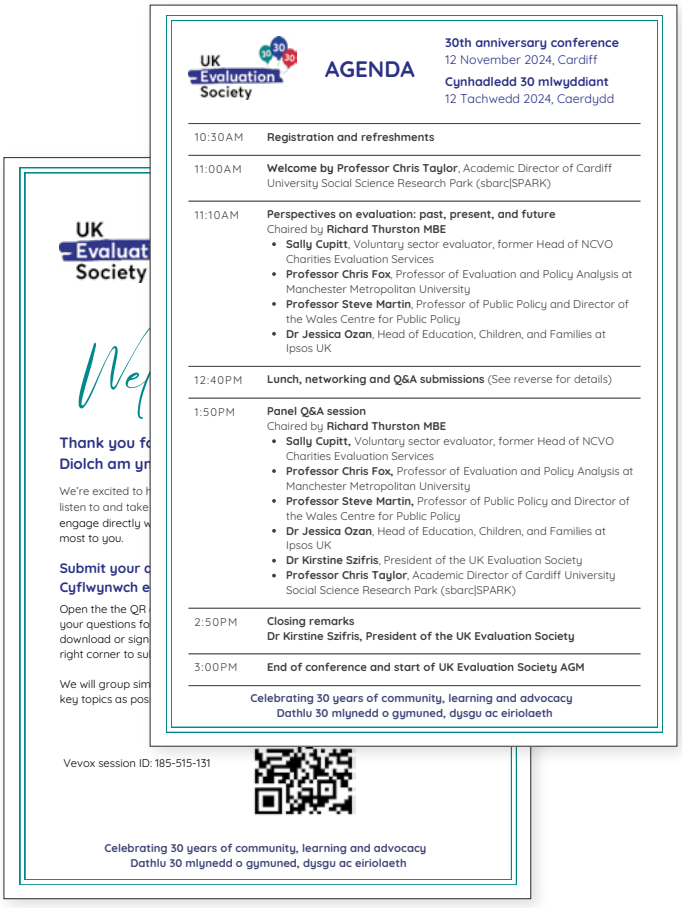
In the afternoon, a lively Q&A panel session opened the floor to questions from attendees.
Richard Thurston MBE, who chaired the event, led an engaging lineup of respected voices, including, Sally Cupitt (Evaluation and Impact Consultant), Professor Chris Fox (Professor of Evaluation and Policy Analysis at Manchester Metropolitan University), Professor Steve Martin (Professor of Public Policy and Director of the Wales Centre for Public Policy), Dr Jessica Ozan (Head of Education, Children, and Families at Ipsos UK), Dr Kirstine Szifris (UKES President), and Professor Chris Taylor (Academic Director of Cardiff University Social Science Research Park).

A hot topic of conversation was the use of A.I in evaluation; particularly in relation to what the evaluation community can do to develop the right guardrails, guidance, and ethics around its use. Dr Szifris was the first of the panel to respond, "There are two things we need to consider, first the evaluation of A.I, and then the use of A.I. in evaluation—and both issues need to be separated out." Dr Szifris talked about A.I. already being firmly on the agenda for UKES, with an upcoming free webinar presented by Professor Oliver Hauser, Will AI Change the World? Evaluating AI in the public and private sector, and the potential for a 1-day conference on the topic in late 2025.

One lasting outcome of the conference is UKES's 30th anniversary letter, a collaborative document that will outline reforms essential to the evaluation community. The letter will serve as the basis of a report aimed at funders, producers, and users of evaluation across sectors. "The report will ensure the insights from today reach stakeholders who can influence change and make a difference," noted Dr Szifris. The report will be developed following a significant consultation process across the evaluation community over the next few months.

In closing, Dr Szifris spoke about the steps we can take now to set up the evaluation community for the future, saying, "The key themes of the day have been around opportunity, collaboration, values, and the need to understand the present, to build a future."

Looking ahead, UKES is excited to see how these insights will shape its strategic direction, focusing on advancing evaluation and continuing to champion its importance. As we celebrate this milestone, we thank everyone who joined us for this special event—a reminder of our shared commitment to a future in which evaluation remains central to making a positive impact on people and communities across the UK and beyond.



Learning Together: Evaluation as participatory sense-making

EMMA MCDOWELL
LECTURER IN CULTURAL AND CREATIVE INDUSTRIES, SCHOOL OF PERFORMANCE AND CULTURAL INDUSTRIES, RESEARCH ASSOCIATE, CENTRE FOR CULTURAL VALUE, UNIVERSITY OF LEEDS, UK



Evaluation can be an effective business or management tool used to justify our value and impact to stakeholders. However, as Elizabeth Cook highlighted in the last issue, evaluation is much more than that; it is a profoundly social practice "woven into the fabric of interconnectedness with our world" (p.21). This is not least because evaluation activity is often a collaborative endeavour, but arguably because it also requires interpreting, capturing and making value judgements on the actions, experiences and impacts of others.

In this article, I'd like to take the opportunity to explore and share some personal reflections on the different dynamics of learning at the heart of much of the design and delivery of evaluation activity. Starting with this understanding of evaluation as social practice and drawing from theories of enactive cognition and participatory sense-making, I hope to emphasise how effective evaluation practice can be, in fact, much more than sharing our best practices and learning from stories of success but rather about fostering a dynamic, responsive engagement with the spaces and knowledge we create together.

The role of participatory sense-making

Learning from one another is challenging. We often talk of knowledge "transfer" or processes of "dissemination". Both concepts are founded on the notion that something discrete is moved from one context to another. But what do we lose when we only package and distribute knowledge in this way? Is learning just a question of finding and absorbing information of sufficient 'quality'? And is sharing this learning with others just a question of beaming this information directly into their brains? The idea of 'enaction' and researchers in the field of embodied cognition, among others, suggests not. Our brains are not blank slates, or 'tabula rasa', primed and ready to absorb information in neatly, pre-packaged forms and reconstitute this into learning. In their book 'Linguistic Bodies: The Continuity between Life and Language', enactive researchers Di Paolo, Cuffari and De Jaegher describe how cognition and understanding inherently involve the activity of others in processes of 'participatory sense-making' (p.100). They draw attention to the dynamic processes of incorporation and incarnation, describing how, as social agents, we continually incorporate the actions or meanings of others, be it their perspectives, attitudes, voices, gestures, movements, personalities or ways of life, and then incarnate them in our actions with the world around us (p.240). Broadly speaking, this continual, dynamic and complex set of processes constitutes our 'being in the world'.

By framing our cognition through these processes of participatory sense-making, we assert that we do not just passively absorb information from others. Rather, we try it on, incarnate it in interactions with others, continuously adapting and reforming it as we go. This is not just a process of mere replication: what we understand is changed, morphed, and incarnated anew, crafted and formed uniquely by us. That is why there is so much vulnerability to getting it wrong or feeling inauthentic. When things don't make sense to us on a fundamental level, we don't learn. It's like we have to experience it in some way for ourselves. I know this from my own experience. I often feel like I learn things 'the hard way'! No matter how much I read or how many different people I talk to, there is no substitute for trying, doing and failing to take on board the specific, contextual and unique lessons that go on to inform future actions.

The power of failure

It perhaps comes as no surprise that experiential learning and learning from failure are areas of growing interest for researchers and higher education practitioners, for example in this recent blogpost on co-production methodologies of activist research projects (Walsh, 2024). There have been significant developments in the arts and cultural sector, too. The FailSpace project, created by Leila Jancovich and David Stevenson, aimed to explore how learning from failure should be an "integral" part of the art-making, management, and evaluation processes constitute an integral part of the art-making, management, and evaluation processes (Jancovitch and Stevenson 2021). When we share familiar professional or personal contexts, learning from shared failures can, paradoxically perhaps, be an enjoyable and enriching collective journey. Other times, we need to reach out beyond the familiar and work harder across our differences, unlearning and (re)learning as we go.

The social dynamics of learning

Digging into the intersubjective dynamics at play when we consider how we learn can be a useful reflective tool for working across different contexts. To orientate my thinking, I found myself developing three loose categories to place a point of emphasis on these dynamics. These are not intended to be mutually exclusive, nor is the list exhaustive. In this sense they are more akin to facets, like that of a gemstone evoked in sociologist Jennifer Mason's Facet Methodology (2011). To illustrate each facet, I include an example of a recent project linked to evaluation in the arts and cultural sector as part of my role at the Centre for Cultural Value. These may (or may not!) resonate with your own work in this area.

Learning from one another: unearthing relevant insights

There are myriad ways in which we hope others will learn from our work and what we learn through our evaluations. We produce publications, reports, or other outputs that synthesise key messages for different readerships, create executive summaries, and consider effective recommendations to inform future work. It's crucial we pay attention to what has already been done and achieved in our field, not least because we can't afford the resources to continue to reinvent the wheel, but also because learning directly from the work of others can help us to deepen our understanding of our own practice. Unfortunately, as I have argued previously (McDowell 2024), this learning potential often lies untapped, as evaluation reports gather dust, unread on shelves, or remain entirely inaccessible to others who would otherwise benefit from them. Given recent conversations I've had with evaluators in other sectors, this is not an issue that is completely unique to the arts, culture and heritage sector.

That was the thinking behind the development of the recently launched Evaluation Learning Space, supported by the Esmée Fairbairn Foundation, which shares insights and learning from the evaluation outputs from five UK Cities and Capitals of Culture from 1990 to 2021. Topics include the people and processes behind evaluations, the methodologies used, what constituted the economic impact of these mega-events and the role of evaluation within cultural policy-making at the local and regional levels. People have appreciated the synthesis of key information, but also the signposting and orientation these resources provide, allowing them to navigate the relevant information from numerous reports more effectively.

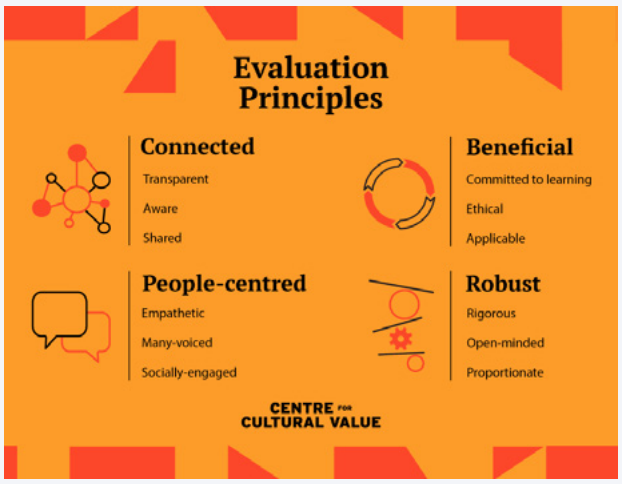
Learning with or alongside one another: sharing contexts of practice

This loose category describes spaces in which we are encouraged and able to regularly share our own personal experiences of learning alongside one another in a more iterative way. This might be over the course of a number of months or even years of our careers and incorporate both formal and informal activities. Here we depend more on shared challenges and similar roles and can result in strong professional and personal networks of practice, both within and across different sectors. Within a particular shared context of practice, narratives of learning are shared easily and are more likely to be applicable to others working in the field.

For example, the Centre for Cultural Value published a number of learning case studies authored by a wide range of arts practitioners working across diverse UK contexts. While case studies are often seen to be instructive as narratives of 'best practice', these case studies were instead designed to open out the processes of learning themselves. Topics ranged from articulating artistic value in arts-in-health research, audience and community engagement within film and performance festivals, developing effective processes for meeting access needs of artists, managing multiple agendas of stakeholders and working in partnership with funders, managing risk when staging site-specific theatre, and artistic leadership as community co-creation. However, both the value and the limits of the learning presented in these case studies do rely on the level of potential applicability to the contexts in which you work. The more aligned the context of the case study is to your own, the more resonant the learning will be within that specific context of practice (McDowell, 2022). Nevertheless, within those contexts of learning, the impact can feel profound.

Learning together: as a collective

Learning as a collective force places a point of emphasis on the construction of a shared research or evaluation endeavour. Similar to collaborative research projects, such as those practitioner-academic research projects supported by the Centre for Cultural Value's Collaborate fund, the shared inquiry pulls team members along a journey of learning which is more interdependent and reliant on others. While these collaborations are often "messy and challenging", the fundamental aspects of these types of endeavours include the building of a solid foundation of mutual understanding of the problem, "embracing the uncertainty inherent in collaborative inquiry" and significant investment in the collaboration process itself (Harrop 2024). As a result of the more unified and purposeful focus of a project, these types of endeavours are particularly well suited to tackling broad and ingrained problems within a sector – for example, improving standards and ways of doing things or developing new tools and frameworks.



Evaluation Principles Source: https://www.culturalvalue.org.uk/our-work/evaluation/evaluation-principles/

An example of this type of learning endeavour was perhaps the process behind the creation of the Evaluation Principles in 2021, which produced a framework to capture and shape how evaluation is carried out and used in the cultural sector. The set of 12 principles [pictured] was co-produced by a working group of over 40 representatives from across the sector, who covered a range of roles and perspectives on evaluation, including those who do it, those who use it and whose work is evaluated. These principles have gone on to be applied by a number of arts, cultural, and heritage practitioners and organisations over the past 4 years, supporting them in setting priorities, engaging the right people, and using appropriate methods to understand the holistic impact of their work.

The socially-rooted, dynamic nature of knowledge creation

The enactive understanding of how we make meaning through and with one another as an embodied form of cognition and being in the world shifts our understanding of learning away from simply a transfer of information from person to person to one where we are interdependent and active in our understandings of ourselves and our own experiences. Personally, approaching learning in this way as an active, embodied practice, rather than just the synthesis, dissemination and absorption of information, feels like a bit of a relief. In an age where evaluation reports abound and demands on our time are constant and unrelenting, fully appreciating the social dynamics of learning feels like we're putting ourselves back in the picture a bit more as professionals.

By considering what matters to us in our particular, situated contexts, we can more effectively share and identify opportunities to learn from, with, and through others to create real-world impact. Instead of knowledge being considered valuable simply because it has been packaged by an institution or given a rubber stamp of approval by an academic journal committee, we can refocus our energy on finding spaces for engaged, meaningful learning that suit our own unique contexts, and be empowered to choose who we learn from and when. This, in turn, informs how we understand and present the experiences of others, which is particularly important in informing our commitments to ethical and inclusive evaluative practice.

After all, evaluation is not just a business process; it's a collaborative journey of shared sense-making, where learning from and with others, and as a collective, shapes not only what we know but also how we engage with this knowledge to bring about real-world change.

Contact

e.l.mcdowell@leeds.ac.uk
https://orcid.org/0000-0003-3775-1417

You can read more about the work of the Centre for Cultural Value, including further information on the learning case studies, the Evaluation Learning Space, Collaborate and the Evaluation Principles as explored in the article, at https://www.culturalvalue.org.uk. Thanks to the Lancaster Evaluation Group for inviting me earlier this year to present some initial reflections on spaces for collective learning explored in this article.

References

Cook, E. 2024. Understanding evaluation as a social practice: Cultivating evaluative mindsets. UK Evaluation Society, Evaluative Practice Edition 02 Spring 2024.

Cuffari, E.C.; Di Paolo, E. and H. De Jaegher. (2018). Linguistic Bodies: The Continuity between Life and Language. MIT Press.

FailSpace Project. (2020). [Online] https://failspaceproject.co.uk/

Harrop, L. (2024). Research collaborations are messy and challenging – here's why we need them more than ever. Centre for Cultural Value. [Online]. https://www.culturalvalue.org.uk/why-we-need-research-collaborations/

Jancovich, L., & Stevenson, D. (2021). Failure seems to be the hardest word to say. *International Journal of Cultural Policy*, 27(7), 967–981. https://doi.org/10.1080/10286632.2021.1879798

Mason, J. (2011). Facet methodology: the case for an inventive research orientation. Methodological Innovations Online. 6(3), pp.75-92. https://doi.org/10.4256/mio.2011.0

McDowell, E. (2021). How to… write a learning case study. Centre for Cultural Value. [Online] https://www.culturehive.co.uk/CVIresources/how-to-write-a-learning-case-study/

McDowell, E. (2024). Creating space for collective learning. Arts Professional. [Online]. https://www.artsprofessional.co.uk/magazine/article/creating-space-collective-learning

Walsh, A. (2024). Learning from 'failure' in co-production: Mismatched expectations and power differentials. Engagement & Partnerships, Leeds Social Sciences Institute. [Online] https://lssi.leeds.ac.uk/engagement-partnerships/coproduction-at-university-of-leeds/co-production-blogs/learning-from-failure/

Better Evaluation Using Evidential Pluralism

FROM LEFT TO RIGHT:
MICHAEL WILDE, JO JONES,
ALEX TROFIMOV, JON WILLIAMSON
CENTRE FOR REASONING,
UNIVERSITY OF KENT






"Our central argument is that for some aspects of the pandemic, especially those characterised by a combination of complexity (multiple variables interacting dynamically with a high degree of uncertainty), urgency (decisions needed in days not years) and threat (the consequences of not acting could be catastrophic), mechanistic evidence has been mission-critical and RCTs difficult or impossible. Thousands of lives were likely lost as a result of what was incorrectly claimed to be an "evidence-based" approach – dismissing or downgrading mechanistic evidence, overvaluing findings from poorly designed or irrelevant RCTs, and advocating for inaction where RCT evidence was lacking. The pandemic is an epistemic opportunity for the EBM movement to come to better understand, debate and embrace EBM+."
(Greenhalgh et al., 2022, p. 253)

The quote left claims (i) that strict adherence to a so-called 'evidence-based' approach during the Covid-19 pandemic is likely to have had severe consequences, and (ii) that adequately addressing complex, critical issues requires a more inclusive approach to evidence. In this piece, we offer such an inclusive approach based on the principles of Evidential Pluralism, a recent philosophical account of causal enquiry.

Current Evaluation Approaches

According to current orthodoxy, experimental studies, especially randomized controlled trials (RCTs), and systematic reviews and meta-analyses of RCTs, are the gold standard of evidence. Other kinds of evidence are usually either given less evidential weight or ignored entirely. Beginning in the 1990's with the evidence-based medicine (EBM) movement (Sackett et al., 1996), evidence-based practice was a positive development that aimed to provide rigorous, systematic evaluation procedures to ensure the safety and efficacy of medical interventions. The primacy of RCTs is understandable in this context of development. RCTs are suitable for answering relatively simple, focused questions regarding population level associations, and can provide good evidence of the effectiveness of pharmaceutical interventions (Greenhalgh et al., 2022). However, RCTs are often inadequate for evaluating more complex problems and interventions, including many problems that arise in public health and public policy.

Dissatisfaction with an orthodox approach to evidence-based policy led to the development of theory-based evaluation, which focuses on mechanisms of implementation (Pawson and Tilley, 1997). The primary aim of theory-based evaluation is to assess how an intervention works in a particular context rather than whether an intervention works (Pawson et al., 2005). Understanding whether an intervention works, however, is often a key concern of policy makers. As a result, theory-based evaluation tends to be seen as a second-best alternative to orthodox evaluation, to be undertaken only when an orthodox evidence-

based evaluation is not feasible. For example, the UK Government's Magenta Book says, 'If none of these [experimental or quasi-experimental] methods seem appropriate, consider Theory-based methods' (HM Treasury, 2020, p. 47).

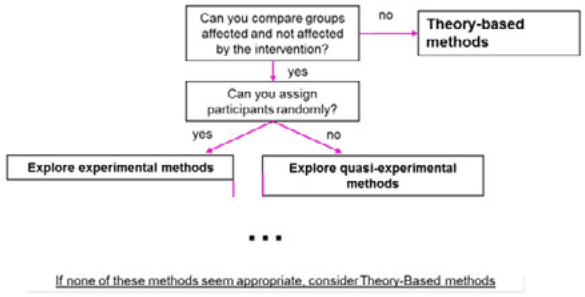


Figure 1: Flowchart on 'Selecting experimental and quasi-experimental methods' from the UK Government's Magenta Book (2020, p. 47).

The time has come for a new approach to evidence-based evaluation. On the one hand, the limitations of orthodox evidence-based practice highlight the need for a more inclusive approach to evidence. On the other hand, although theory-based evaluation can offer important insights into how an intervention works, it does not fully address the question of whether an intervention works. What is needed is an alternative approach to evidence-based evaluation that can tell us what works by systematically incorporating the full range of relevant evidence. Evidential Pluralism offers such an approach.

Evidential Pluralism

Evidential Pluralism is a philosophical account of causal enquiry. Evidential Pluralism includes the following two principles, illustrated in Figure 2:

Object Pluralism: Establishing that A is a cause of B requires establishing (i) a correlation between A and B and (ii) a mechanism connecting A and B that can account for the extent of the identified correlation.

Study Pluralism: Evaluating a causal claim normally requires assessing both association studies and mechanistic studies, where available.

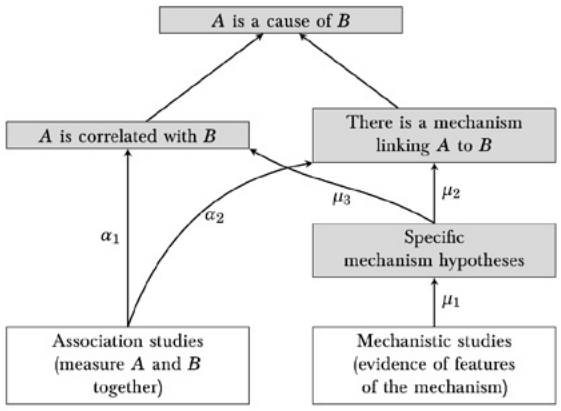


Figure 2: Evidential Pluralism

Motivation for object pluralism comes from the truism that correlation is not causation. A correlation between A and B might be due to A causing B. However, it might be due to any of a large number of alternative explanations, such as confounding, bias, reverse causation or chance. If a correlation between A and B really is due to A causing B, then there must be some mechanism of action by which A produces B. Therefore, establishing causation requires establishing both a correlation and an underlying mechanism that can account for the correlation.

Study pluralism follows from object pluralism. Since establishing causation requires establishing both correlation and mechanism, evaluating a causal claim will normally require assessing both association studies and mechanistic studies. Association studies, according to Evidential Pluralism, test for the presence of a correlation. This is done by repeatedly measuring instances of A and B to determine whether A and B are probabilistically dependent, conditional on potential confounders. Mechanistic studies, according to Evidential Pluralism, include all studies that assess the presence of features of the hypothesised mechanisms linking A and B. Such features might include mediating variables, entities or activities involved in a mechanism, or ways in which these components are organised.

By requiring an assessment of both correlation and mechanisms, Evidential Pluralism provides a framework for integrating diverse kinds of evidence when evaluating whether an intervention works. Assessing correlation requires assessing quantitative evidence from experimental or observational studies. If these studies identify a robust correlation, this can increase confidence that there is some underlying mechanism. However, they will rarely be sufficient on their own to establish that a complex intervention is effective. Evidence of key features of the proposed mechanism of action can increase confidence in effectiveness. On the other hand, if these features are found to be absent, or if key features of counteracting mechanisms are found to be present, this can undermine confidence in effectiveness.

Evidential Pluralism motivates a more inclusive approach to evidence evaluation in medicine, called EBM+ (Parkkinen et al., 2018), policy, called EBP+ (Shan and Williamson, 2023), and law, called EBL+ (Trofimov and Williamson, under review).

Covid-19 Face Masks Case Study

As Greenhalgh et al., 2022 suggest, a narrow focus on experimental studies may be partly to blame for the continued controversy and uncertainty concerning the effectiveness of Covid-19 public health interventions, including public face mask mandates. To illustrate the need for and benefits of an Evidential Pluralism evaluation, we conducted a proof-of-concept evaluation of Covid-19 face mask mandates (Trofimov and Williamson, under review). This evaluation did not include a full systematic literature review and therefore the conclusions should be taken as tentative. Nevertheless, this case study provides a good illustration of the need for and benefits of an Evidential Pluralism evaluation. Here, we provide a summary of that proof-of-concept evaluation.

An Evidential Pluralism evaluation begins by specifying the claims of interest. In this case, they are:

The causal claim: a legal requirement to wear a face mask in public reduces the prevalence of Covid-19 infections and thereby reduces the number of hospitalizations and deaths.

The correlation claim: a legal requirement to wear a face mask in public is negatively correlated with symptomatic Covid-19 infections, conditional on potential confounders.

A plausible mechanism hypothesis: a legal requirement to wear a face mask in public increases the use of face masks which in turn reduces the prevalence of Covid-19 which reduces the prevalence of symptomatic infections and thereby the number of hospitalizations and deaths (Figure 3).

A plausible counteracting mechanism hypothesis: a legal requirement to wear a face mask in public will decrease compliance with other public health interventions, such as social distancing. This, in turn, would result in an increase in the number of symptomatic infections compared to the number that would have occurred if the legal requirement to wear a face mask in public had not been implemented (Figure 3).

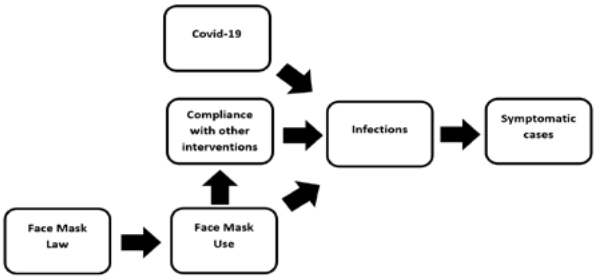


Figure 3: Mechanism Hypothesis.

The next stage is to search for and assess association studies, to enable an evaluation of the correlation claim. The search found a range of association studies that, taken together, detect a robust correlation across contexts. On the basis of that evidence, Trofimov and Williamson conclude that the correlation claim is established and that confidence is increased in the existence of a mechanism that accounts for this correlation. Although each of the association studies controlled for several confounders, some residual confounding may persist. The association studies therefore do not suffice to fully establish causation and a mechanistic evaluation was deemed necessary because it could further increase confidence in the causal claim.

A literature search revealed that each stage of the mechanism hypothesis is supported by a range of studies and that there is evidence against the hypothesised counteracting mechanism. On the strength of the evidence considered, Trofimov and Williamson conclude that the mechanism hypothesis is confirmed but not fully established, and that the counteracting mechanism hypothesis is arguably false.

Overall, the evaluation found strong evidence of both correlation and mechanisms. Association studies on their own establish the correlation claim and the mechanistic studies reinforced this conclusion. Neither association studies on their own nor mechanistic studies on their own fully establish the existence of a mechanism that can account for the correlation.

When association studies and mechanistic studies are considered together, however, confidence in the existence of such a mechanism is increased to such an extent that it is unlikely that further evidence would undermine it. Therefore, on the basis of the mutually supporting evidence of correlation and mechanism, the mechanistic claim was deemed to be established. Since both the correlation claim and the mechanistic claim are established, Evidential Pluralism implies that the causal claim is established.

By taking account of a broad range of evidence, Trofimov and Williamson were able to reach a robust, positive conclusion regarding the effectiveness of Covid-19 face mask mandates. Given the continued controversy and uncertainty surrounding Covid-19 face mask mandates, this is significant. For a more detailed explanation and further examples, please see our introductory guide (Jones, Trofimov, Wilde and Williamson, 2024).

Conclusion

As the pandemic has highlighted, there is a pressing need for a more inclusive approach to evidence-based evaluation, especially for complex issues. While this is becoming more widely recognised, official guidance and practice does not offer a framework for systematically integrating different kinds of evidence to yield better informed evaluations of effectiveness. Evidential Pluralism offers such a framework and thereby promises to improve evidence-based evaluation.

References

Greenhalgh, T., Fisman, D., Cane, D., J., Oliver, M. and Macintyre, C. R., (2022), 'Adapt or Die: How the Pandemic made the shift from EBM to EBM+ more Urgent', *BMJ Evidence-Based Medicine*, 27: 253-260.

HM Treasury (2011), *Magenta Book: Central Government Guidance on Evaluation*, HM Treasury and Evaluation Task Force, London, 2020 edition. Available at: The Magenta Book - GOV.UK (www.gov.uk)

Jones, J., Trofimov, A., Wilde, M. and Williamson, J., (2024) 'Integrating Diverse Evidence Using Evidential Pluralism: Introductory Guide'. Available at: IDE-EP.pdf (jonwilliamson.uk)

Parkkinen, V-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M. P., Norell, C., Russo, F., Shaw, B. and Williamson, J., (2018), *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*, Switzerland: Springer.

Pawson, R., Greenhalgh, T., Harvey, G and Walshe, K., (2005), 'Realist Review – A New method of Systematic Review Designed for Complex Policy Interventions', *Journal of Health Services Research and Policy*, 10: 21-34.

Pawson, R. and Tilley, N., (1997), *Realistic Evaluation*, London: Sage.

Sackett, D. L., Rosenberg, W. M., C., Gray, J. A. M., Haynes, R. B. and Richardson, W. S., (1996), 'Evidence Based Medicine: What It Is and What It Isn't', BMJ, 312: 71.

Shan, Y. and Williamson, J., (2023), *Evidential Pluralism is the Social Sciences*, Abingdon: Routledge.

Trofimov, A. and Williamson, J., (under review), 'Applying Evidential Pluralism to Evidence-Based Law: EBL+'.

Policy Evaluation: A Critical Friend or Foe?

CHARLIE HEYWOOD-HEATH
SENIOR ECONOMIST
LICHFIELDS UK



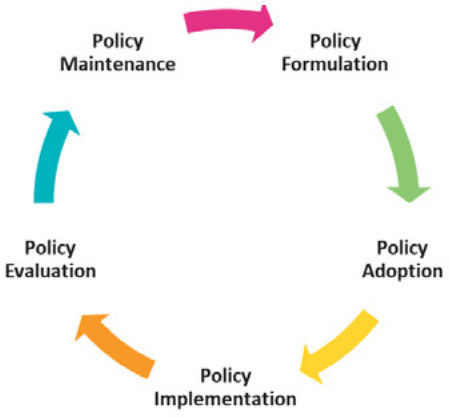
Policy evaluation in its current form has been growing as a practice since the 1960s and 70s when Governments and funders became concerned with an evidence-based approach to ensuring policies were achieving what they set out to deliver. Following the post-war interventionist boom from Governments, evaluation was both a way of supporting the use of public funds but also a means for removing it where the use was deemed inefficient and ineffective.

As policy evaluation became more institutionalised within bodies such as the European Union, the World Bank and national Government, it became an integral part of the policy cycle. Anyone who was involved in the delivery of major European Funds, such as European Regional Development Fund or European Structural Investment Fund, will know that there has been a strong emphasis on evaluating the relevance, efficiency and overall effectiveness of public intervention over the last 20+ years.

Today, policy evaluation has become a fundamental aspect and requirement for most public interventions across different levels of government. However, due to this, there is a risk that the practice has become a dirty word, sometimes regarded by public bodies as a burden rather than a benefit. In this article, I will make the case for reimagining policy evaluation as a critical friend with the design and delivery of policy rather than a critical foe set out to trip up or expose weaknesses.

As set out by Jacobs (2024)[1], evaluation arose at a similar time as auditing within public management and has often had to fight for its own distinct role within managing and reviewing public policies and programmes. Due to this, evaluation has sometimes become too much of a pejorative term. In Government and the public sector, evaluators are often brought in externally due to perceptions of objectivity, funding requirements and because of capacity constraints within the public sector. The use of outsiders can often lead to an 'us versus them' dynamic, given those evaluating the intervention have not necessarily been involved in the blood, sweat and tears of delivery. Thus, rather than seeing evaluation as trying to get the best out of a project, it is often seen as a source of criticism rather than critique.

In part this is due to its place within the policy cycle. Whilst most practitioners believe in embedding evaluation within the design of policy from the outset, it often falls at the end. The Harvard Catalyst speaks of 'policy maintenance' following evaluation.[2] Despite this, evaluators may not be involved in the maintenance or implementation of recommendations, nor may they re-evaluate whether the 'maintenance' has addressed or improved the outcomes. This in part results in evaluation being seen as a tick-box exercise, a mandatory requirement within funding agreements rather than a fundamental part of understanding and shaping policy delivery.



The design of policy evaluations has also led to a negative perception of its use within a project. Ultimately, many evaluations are concerned with assessing the Value for Money generated from an intervention. Whilst the inputs used to calculate the monetary value generated from an intervention can vary depending on what's being evaluated, the question is largely a binary one: for every pound of investment, how much extra benefit is generated? A low or negative Value for Money can often be seen as the albatross around a policy or programme's neck; a determinant of whether it has a future outside the life cycle of the current funding regime.

Due to this, delivery teams can end up feeling hostile towards evaluators whose judgement may ultimately determine the future of their project. This raises the question of how we ensure that policy evaluation is regarded as a critical friend within a policy. There is no straightforward solution to this but there are a few decisions which can move us in the right direction.

Firstly, and one that is commonly agreed by practitioners and those writing about evaluation, is that we need to involve policy evaluators within the early stages of policy interventions. By doing so, evaluation becomes part of a policy rather than something separate. It also reduces surprises when the outcome of an evaluation doesn't reflect the perceptions of those who delivered it. This includes the design of policy objectives to ensure future accountability and measurement but should also extend to the conversations with funders and policy officials to ensure all parties are engaged in talking the same language. Whilst some policies can be designed or framed as part of a linear cycle with a clear beginning, middle and end, the delivery process is often iterative and requires continued collaboration between all parties.

Secondly, the language and end output of evaluation needs a re-think. This is a question of the usability as well as use of an evaluation. It is widely noted that most policy evaluations are seldom read or engaged with, even by those who have commissioned them, regardless of whether they are policy experts or not. Research can often be dry, overly technical and frankly boring. Within policy evaluation, this maintains the air of exclusivity within the profession. Studies can be shrouded in discussion of methodology and technicality, often due to a desire to uphold the legitimacy and pseudo-scientific status of an evaluation. Whilst there is a clear need to maintain methodological rigour, outputs need to be more accessible and engaging. Evaluation isn't just a research exercise but is a tool for telling a policy's story and providing lessons learned for the future. If we cannot understand or engage with the evaluation, we cannot learn from it.

Lastly, policy evaluation needs to be given time. Whilst it should be embedded throughout a policy's life cycle, the end or final evaluation needs time. Too often the UK has been mired with short-termism. Political pressures and disruption have led to policies designed, delivered and their future decided without proper time for materialisation. In the same vein, policy evaluation has often been conducted too soon before impacts have properly materialised. Combined, there is limited opportunity for follow-up evaluations due to the fast-moving pace of many public departments and teams. Where a policy or programme requires time, at a strategic level we need to appreciate that long-term evaluations will not only lead to more fruitful findings but will also produce better accuracy and assessment of overall benefits and impacts.

In Labour's first Budget in 14 years, the Chancellor announced an Office for Value for Money (OVfM) which will scrutinise public investment and seek to ensure efficiency, maximise effectiveness and minimise wastefulness in public policy.[3] As public policies and programmes enter a new era of scrutiny at a time of fiscal prudence and challenging economic circumstances, there is a renewed opportunity to ensure evaluators are at the centre of supporting the delivery of effective and impactful public policy.

This article may be preaching to the converted. Those of us involved in policy evaluation will largely agree that evaluation is important and is ultimately concerned with the overall efficacy of an intervention. We are not there to criticise or catch-out projects but to ensure that decision-making is underpinned by robust and reliable evidence. However, we do not exist in a vacuum but co-exist amongst those responsible for delivery, design and funding. By promoting greater collaboration, communication and credibility, we can make evaluation a friend of the policy process.

References

[1] Jacob, S. (2024). Navigating the Challenges of Policy Evaluation. Canadian Public Administration, 67.

[2] Harvard Catalyst (2024) Policy Life Cycle

[3] HM Treasury (2024) Autumn Budget.

What is a method? I'm asking for a friend (of Wittgenstein's)

BARBARA BEFANI
RESEARCH DIRECTOR OF THE CENTRE FOR EVALUATION IN THE NATIONAL CENTRE FOR SOCIAL RESEARCH.



The word "method" has sometimes scary connotations, evoking imposition, linearity, and lack of freedom. This has generated a "post-methods" movement in education[1] and has attracted the interest of evaluators, as well. But what exactly is a method? And how do we differentiate between methods?

At a basic or practical level, many of us associate "method" with "data collection technique": a method is interviews, focus groups, desk reviews or a particular type of survey. A higher level, more abstract set of principles or procedures is also often associated, if not with the word "method", certainly with the word "approach", "methodology", or "design". [2] Contribution Analysis, Process Tracing, QCA, Realist Evaluation, RCTs, Difference-in-Difference, and so on, would fall under this classification.

Who gets to define what a specific method is, exactly, and draw implications from it? One might think of a seminal textbook, an article, or a series of papers, that have introduced the method to the evaluation community. The very first paper where Contribution Analysis was mentioned was published more than 25 years ago.[3] Since then, the same author has written many other papers on the same topic, adding new elements, clarifications, and concepts (like causal necessity and sufficiency). The last contribution I was able to find dates 2020,[4] or 21 years after the first. Many other authors have also contributed to enrich the literature on Contribution Analysis. What is Contribution Analysis, then? What John Mayne described in his first paper? His last one? Or what other authors have added?

QCA was first introduced in 1987.[5] Since then, many variants of QCA have been introduced over the years, and current "QCA software" includes several additional procedures that were not contemplated in the original book nor MS-DOS (!) software. In addition, QCA requires a very specific way of "calibrating" data into dichotomous or fuzzy conditions even before using any software. What is QCA then? The procedure that was first presented in 1987? Those presented later? Any of their parts? The combination of all? What about the preparatory work preceding the software-assisted synthesis, which must meet specific requirements to be run with QCA software? Is it QCA? Pre-QCA? Almost-QCA?

Process Tracing is another interesting example. One might argue that the key defining feature of PT is the evidence test, commonly known by the four metaphors (Hoop Test, Smoking Gun, Doubly Decisive, Straw In The Wind). But let's assume that our comments on the strength of evidence include considerations on its conclusiveness and direction without mentioning the metaphors. Is it still Process Tracing? One of the prominent authors of the PT literature recommends a quite specific way of formulating the causal mechanism for PT. But let's say that we want to test the credibility of a differently-looking claim or of some interesting parts of a complex system, and not only we use the PT metaphors for that, but we go all the way to formally estimating Bayesian likelihoods and confidence levels. Can anyone say we haven't done Process Tracing? You should start seeing where I'm getting at with this, but allow me one last example: Realist Evaluation. The author we all agree introduced Realism to evaluators, Ray Pawson, after discussing Realism as an ontology, proceeded to give us practical indications on how to use the obscure thing in practice, and came up with the so called "realist interview". Now, I haven't seen a huge amount of evaluations claiming to be "realist" which have confessed to doing the latter. Does that mean these were not "realist evaluations"?

Methods as (fuzzy) sets

Wittgenstein introduced the idea of "family resemblance" in a posthumously published book in 1953,[6] to essentially argue that humans use language in a fuzzy way, and even if it's unlikely he had public policy evaluation in mind when he wrote that, the same principle can be applied to evaluation methods. *A single evaluation method / approach is not an immutable and clear-cut object. It is made of parts: principles, epistemological grounding, practical suggestions to apply the method in practice, sometimes specific procedures of data collection and analysis.*

To what extent can we then legitimately claim to have applied a method? Under what circumstances can commissioners, peer reviewers, or stakeholders say "that was not really Contribution Analysis", "that was a poor man's version of Process Tracing", that was a "high-quality Realist Evaluation", and so on?

After Wittgenstein, let's look at Gary Goertz and his work on Social Science Concepts.[7] Like a concept, a method can be thought of as a set, or collection of elements. All methods I can think of are presented with at least an epistemological basis AND a series of more technical procedures or suggestions.[8]

When and under what conditions can we then claim to have legitimately applied a methodology?

Following Goertz and a set-theory framing of methods, we could be either on the loose end of its definition (the least restrictive version) and use the logical union (or disjunction) of a set of its key elements (exhibit 1), or use the most restrictive definition, or the combination (or logical conjunction) of its key features (exhibit 2).

1: METHOD = A + B + C + D? [A or B or C or D]

2: METHOD = A * B * C * D? [A and B and C and D]

In practice, this would mean that, in the first case, we'd be able to say that we've "applied the method X" if we've used at least one of its defining features. In the second case, at the other end of the spectrum, we'd only be able to claim that we've applied method X if we used ALL of its defining features.

Experts and peer reviewers play a key role in answering the above questions. But at the moment, there is confusion and uncertainty around how this is handled, and in practice it boils down to the peer reviewer's knowledge and experience.

Beyond bricolage: the continuous, reasoned, and transparent (re)definition of methods

The concept of *bricolage* has recently gained popularity.[9] In one of the most basic meanings of the term, that a school child would understand,[10] bricolage is a combination of different materials, pieces from different types of paper, that have been torn from material or paper sheets that were previously whole or intact. It has gained popularity as a metaphor for the combination of different ideas, principles, elements, and tools, normally associated to different methods.

There are at least two problems with this metaphor. First, bricolage is an art or craft, and does not require a reasoned justification of why these elements have been combined together. Why they have been selected, why others have been left out, and whether they are coherent for a specific design, rather than in conflict. Second, bricolage implicitly assumes that 1) a method can be legitimately thought of as a whole, unitary object, from which parts are taken; and that 2) we clearly know what that object is made of, and can meaningfully differentiate it from other objects.

These assumptions and requirements are problematic. First, art draws on human emotion rather than stimulating reason and as such can be exploited by power. Transferring this principle to evaluation design risks downplaying the role of evaluation quality and quality-assurance, or disanchoring the latter from rational principles.

Furthermore, taking a "piece" of something well defined in advance works less well as a metaphor when methods are already fundamentally seen as fuzzy sets of loosely linked ideas, elements, and principles. Every time we design a specific evaluation, we somehow define what one or more methods are, by picking and choosing various bits from them, depending on what is needed to meet the evaluation aims, answer the evaluation questions, and make the evaluation useful for the stakeholders.

The evaluator's role is to interpret the commissioner's and stakeholders' needs and use methods as appropriate for that purpose and situation. This will require framing methodologies: for example, explaining why a particular type of interview was used, why we used certain elements of Process Tracing but not others, and why we picked some ideas from QCA but not others. The moment of definition, or explaining why we made the choices we made, and left out the parts of methods we left out, is essential for peer reviewers to transparently establish whether the choices made are defensible.

It's important that the idea of bricolage does not lead to uncritical acceptance of poorly transparent choices influenced by power dynamics.[11] In governance, the use of art is historically not highly reputable: it is autocracies that favour it, filling the public space with dictators' portraits and suggestive imagery. Historically, it is absolute monarchs and the Catholic Church that have commissioned the vast majority of human-made beauty. The preservation and diffusion of democracy requires rationality and the kind of discussions that increase the chances of mutual understanding.

More generally, umbrella terms such as "bricolage" end up encompassing such a variety of evaluation practices as to lose utility and meaning in the long-term, like "mixed methods", "theory of change" or "theory-based evaluation"; and in my view, "contribution analysis", as well, given the diversity of applications it counts under its name, by now. In many ways, these are *residual* terms created to *negate* a prevailing or hegemonic practice, the practice one wants to differentiate from: mixed methods means, neither exclusively quantitative nor exclusively qualitative; theory-based means we're not just doing quantitative data analysis and monitoring indicators; but are developing and testing a theory which is expressed with a qualitative narrative (and will likely require qualitative data). The selling point of Contribution Analysis is historically, that it is NOT attribution. Umbrella terms are always born in protest against something, they are helpful to identify something that "is not". Bricolage means, we're not selecting one method and applying it from start to finish, in all its articulations and elements; we're picking and choosing elements of it, and combining it with other bits and bobs from other methods. Again, what the term is clearest at conveying, is what it is not.

I can understand wanting to break barriers and boundaries when terms are seen as straitjackets and limit our freedom. It is understandable to go through a phase where we give all the rubble a single name. But moving forward, the many rocks and pebbles deserve more attention and can be put together again under looser, temporary structures that will fit our conjunctural requirements. Knowing they will be disassembled again and built together again in a different shape, at the next evaluation.

Taking heed from Wittgenstein and Goertz, we should stop seeing words and concepts as rigid, iron-made boxes. We should see them as tools, prompts, starting points for discussion. In evaluation methodology, for policy design, and to improve communication among evaluators, it's important to draw on a variety of terms to convey at least a fraction of the variety of our choices. Furthermore, the meaning of these terms should be continuously discussed and clarified, rather than taken for granted.

It is up to stakeholders, experts, peer reviewers, commissioners, and others engaged in the evaluation process to discuss what a method is and what specific instance of its conceptual plasticity and fuzziness we need for a specific evaluation. It is up to the facilitators of this process, drawing on the increasingly widespread information, education, and methodological awareness, to ensure that such a debate takes place on an equal basis rather than on uneven power dynamics. Drawing on as much of an abundance and variety of terms as we can, provided we stop seeing those terms as watertight compartments.

References

[1] Golzar, J., & Fuller, C. (2020). Teacher identity formation through classroom practices in the post-method era: A systematic review. Cogent Education, 7(1). https://doi.org/10.1080/2331186X.2020.1853304

[2] Vaessen, Jos, Sebastian Lemire, and Barbara Befani. 2020. Evaluation of International Development Interventions: An Overview of Approaches and Methods. Independent Evaluation Group, MethodsSourceBook.pdf (worldbankgroup.org)

[3] Mayne, J. 1999. Addressing Attribution through Contribution Analysis: Using Performance MeasuresSensibly, discussion paper, Office of the Auditor General of Canada

[4] Mayne, J. 2020. Contribution Analysis and the Long-term Perspective: Challenges and Opportunities. In K. Forss, I. Lindkvist, M. McGillivray (eds.) Long Term Perspectives in Evaluation: Increasing Relevance and Utility

[5] C. Ragin. 1987. The comparative method: Moving beyond qualitative and quantitative strategies, University of California Press.

[6] Philosophical Investigations

[7] G. Goertz (2005) Social Science Concepts: A User's Guide. Princeton University Press.

[8] Realist Evaluation has the realist ontology and the realist interview. QCA has multiple-conjunctural causality and the algorithms / software. Process Tracing has the Bayesian logic and the empirical analysis of evidence (four types: account, pattern, sequence, and signature: see D. Beach, R.B. Pedersen (2013) "Process-Tracing Methods Foundations and Guidelines" University of Michigan Press DOI: https://doi.org/10.3998/mpub.10072208). Even Contribution Analysis, despite its comparatively loose guidelines, has the contribution story, and the six steps.

[9] Aston, T. and Apgar, M. (2022) The Art and Craft of Bricolage in Evaluation, CDI Practice Paper 24, Brighton: Institute of Development Studies, DOI: 10.19088/IDS.2022.068

[10] and that is my first personal encounter with the term when I was in primary school

[11] The idea that "Deciding which parts of which methods to include or exclude is a deeply political process" (Aston & Apgar 2022) is scary and, no matter how realistically it can describe a particular situation, it must be fought and rejected on a moral basis.

What should we measure in programme evaluations: Insights from psychology

MARINA BAZHYDAI
LANCASTER UNIVERSITY
PSYCHOLOGY DEPARTMENT



Our society benefits from many programmes and services created and implemented by various organisations. Programmes can range from large scale, government funded, such as Arts Council England, to smaller scale locally delivered ones, such as a well-being support group at a town's community centre. While a choice of an approach to programme evaluation is a question of ongoing debate (see for example Mertens and Wilson 2018), the value of conducting evaluations as such is not to be under-estimated: evaluations allow documenting the programme's value and may help secure funding or influence policy changes. Research literature documents the results of such programme evaluations, such as a workshop in an art museum aimed at enhancing children's emotion and creativity skills (Ebert et al., 2015), a healthy eating and movement preschool programme (Hardy et al., 2010), a new national curriculum in schools (Baer, 2003), or a parent-infant interaction training (Choi & Rowe, 2021). What are such programme evaluations capturing – and are the design and measures used appropriate? Using an example from a recent programme evaluation study, I offer some reflections from the field of psychological science, specifically its cognitive, developmental, and social domains.

It is only natural that organisations would be interested in knowing if their programmes have had an impact – 'worked' in some beneficial ways or 'delivered' as intended. With limited resources, a common approach to this would be to ask participants to answer a short list of questions at the end of the programme. For example, at the end of a 6-week educational programme in schools, teachers and pupils involved might be asked whether they liked the activities, felt engaged, or gained new knowledge, and also invited to leave any comments. The results of such surveys are then summarised to demonstrate impact (e.g., 90% of participants felt engaged by the activities) and open-ended questions' responses described to illustrate the positive effects. Such evaluations, despite their simplicity, may offer valuable insights for organisations and stakeholders.

In contrast, in the field of psychology, we operate with 'psychological constructs. In the context of programme evaluation, our focus would be on psychological processes: does the individual change as a result of partaking in the programme, in terms of psychological states, cognitive skills, or learning outcomes. Related to the example above, instead of simply asking if participants liked the activities with a single question, a psychologist would instead need to reliably capture concepts such as "curiosity", "wonder" "enjoyment of learning", or "attitudes towards learning". As for evidence of learning, our goal would be to assess it with reliable measures.

Thus, instead of coming up with ad-hoc questions, we would give participants either standardised assessments (e.g., a reading ability test, or a creativity task) or psychometrically validated self-report scales, typically consisting of several related questions aiming to tap into the same psychological construct. For example, take the idea of a 'growth mindset', which is a belief that individual abilities such as intellect or creativity, rather than being determined, can be enhanced or decreased and are thus malleable. This is typically captured with the Growth Mindset Scale (Dweck et al., 1995). To claim that this particular set of questions constitutes a 'validated' scale, researchers must undertake a rigorous mixed-methods process, establishing its reliability, appropriateness, internal and external consistency, convergent and divergent validity, among other statistical and disciplinary methodological indicators.

Measurement is a hotly debated issue in psychological research as some argue that ad-hoc, often single item, measures that have not undergone rigorous validation constitute questionable research practices threatening research integrity (Flake & Fried, 2020). Such harmful "measurement schmeasurement" attitude in turn raises doubts about the validity of study's conclusions. Others, however, argue that ad-hoc questions may serve the purpose just as well and should not be completely dismissed. Using the example of the Growth Mindset Scale, a recent paper specifically compared it to a single item, ultimately demonstrating that both are acceptable (Rammstedt et al., 2024).

In light of the above considerations, in order to evaluate the impact of a programme on an individual, the researcher would first need to carefully consider and select target constructs that are likely to show a change as a result of the programme. Among other methodological challenges are to ensure that measures are appropriate (e.g., for young children, caregivers, special populations, etc.), reliably capture the construct of interest, and are feasible to be administered in the setting or time allowed. Furthermore, the 'golden standard' to strive for would be a randomised controlled trial design, where the programme is treated as an 'intervention', with the pre- and post-intervention battery of appropriate measures, and a matched control group that is not receiving the intervention. The scores are then compared both given the time that had passed (e.g., 6 weeks) and the group assignment (e.g., receiving or not receiving the intervention). Thus, the selected measures of psychological constructs also need to be amenable to being presented twice to the same participants, and groups need to be well matched.

Next, I present a study to illustrate some of these challenges and discuss potential solutions and implications of taking different measurement approaches in programme evaluation.

My research team collaborated with two local community organisations, More Music in Morecambe and Ludus Dance in Lancaster, UK, that developed an award-winning 6-week school-based creative arts programme, entitled "The Alien Needs Our Help"[1]. Its aim was to help educate children about health and well-being, in particular keeping safe from diseases and germs. This was accomplished through songs, dance moves, and putting on a school performance. Importantly, children were given opportunities to not only learn about health-related practices and information, but also teach an 'alien'– an educator playing the role of a naïve learner – about staying healthy. Thus, children acted as both learners and teachers themselves, and two-way information transfer occurred through the creative arts media.

When faced with a question - what did children gain following engagement in the programme? – our research team turned our focus away from directly assessing the programme's postulated goals (health-related information gain) towards psychological processes that might have been impacted. Given the unique mode of programme administration through the creative arts, we posited that children's perceptions of their ability to be curious, teach others, and be creative may have been impacted by the programme's activities. We worked closely with the programme's creators and educators delivering it in schools to propose our research ideas and gain their feedback, making this study participatory.

We tested 165 primary school children in the UK Year 6 classrooms, employing a longitudinal design, with two pre- and post-intervention testing sessions in the schools. This was a 2 (group: experimental (77 children) vs. passive control (86 children)) x 2 (time: time 1 vs. time 2; pre-intervention vs post-intervention in experimental group or 6 weeks apart for control group) mixed design, with group as the between-participants and time as the within-participants factor. This was, therefore, an RCT design as we did this study both with children who were due to take part in this programme through their school and those who do not, to understand the effect of the programme, if any, on children's understanding of learning, teaching, and creativity. We collected the data in person, in children's classrooms, and each child worked through the questions independently and could ask questions from the researcher present. It took about 20-30 minutes to complete at each time point.

As we aimed to detect changes in children's attitudes, beliefs and behaviours following the target creative arts educational intervention, the measures we selected assessed children's curiosity, agentic engagement, and enjoyment of learning, their own understanding of learning, creativity and teaching, as well as specific learning outcomes based on the programme's content on health and well-being vocabulary and causal understanding. Whilst some scales included previously validated measures for the reasons mentioned above (e.g., Children's Academic Orientation Scale (Ruzek et al., 2020); Children's Images of and Attitudes towards Curiosity questionnaire (Post & Walma van der Molen, 2019); Agentic Engagement (Reeve & Tseng, 2011); Embedded Interest Scale (Ainley & Ainley 2011)), other questions were developed specifically for this study and thus did not constitute psychologically validated scales (see Figure 1). This was because there were no measures that corresponded closely to the programme's content and delivery and the constructs, we were keen to capture. In addition to quantitative measures, we also included open-ended questions to consider qualitative data as well.

We expected that children who took part in the educational intervention would be more likely to report higher curiosity and enjoyment of learning, think of themselves as someone curious, creative, and who can confidently share knowledge with other people, and report more flexible and broad reflections on one's own capacity for learning and teaching others (for example, learning by teaching, or teaching using creative approaches). Finally, we expected that the intervention will have a positive effect on children's knowledge of and interest in health-related information.

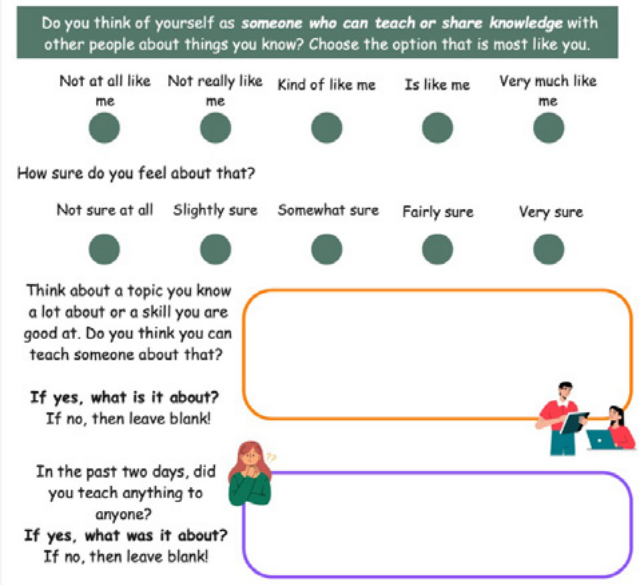
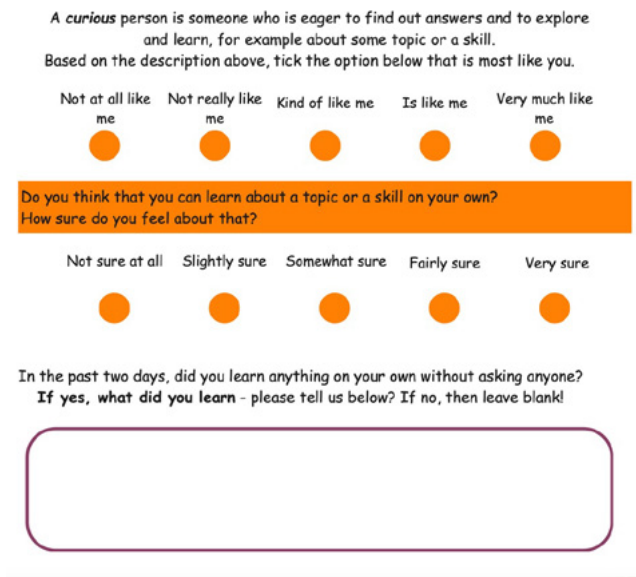
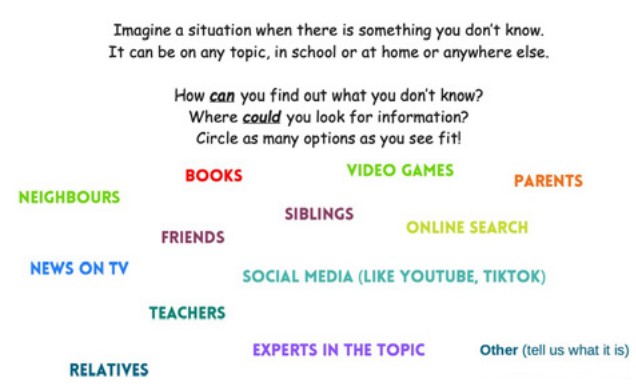


Figure 1: Examples of questions used
1a: *Perception of own ability to share knowledge with others:*



1b: *Perception of one's ability to be curious and learn independently*



1c: *Propensity to consider different sources of information*

Having collected the data and subjected it to rigorous statistical analyses, we found that children's responses to most measures we included did not yield statistically significant results. This suggests that the educational intervention did not lead to the hypothesized changes in the psychological constructs we tested between children who took part in the educational programme and children who did not above and beyond the mere passage of time. Interestingly, and crucially here, before we had our control group responses, our preliminary results from the data obtained only from the group who took part in the programme showed that children were generally broader and more flexible about where they can look for information, motivations for learning, ways they can learn, ways they can share what they have learned with someone else, and considerations for teaching. In addition, children's both self-reported and objective knowledge of health-related vocabulary had improved. However, these findings were dampened once our control group data was collected and analysed – as in comparison to those, we ultimately found no reliable differences.

We were therefore bound to conclude that we could not detect an effect of the educational programme on the psychological constructs and learning outcomes selected for this study. However, does it mean that the programme was not effective, useful or interesting to children, or that it did not achieve its goals? We are not so keen to jump to such conclusions, and here is why.

First, it is possible that despite our best efforts to tap into the true value of programme, even if validated scales do not exist or are not necessarily best suited, the actual measures we chose to capture children's perceptions of themselves as creative individuals who are capable of communicating what they have learned to others were not ideal. While consulting with programme leads, we did not have an opportunity to consult with teachers or children themselves before collecting the data. It is possible that children did not perceive the questions in the same way we intended, and a focus group or an interview approach would have produced different results. It is also possible that measures did not capture the construct of interest, despite many of them being psychometrically validated, or that the constructs themselves were not suitable in this context. Other questions abound: did children engage with the study correctly and sufficiently to contribute to accurate results; were there cognitive or social biases present (such as social desirability); did they pay attention or were they bored? Because the measures were the same before and after, were children responding genuinely? Furthermore, all measures aimed at capturing group-level changes, rather than looking at individual children. What if some, even a small percentage of children in our sample, have been positively impacted so substantially that the programme was worth it even if it was not the case for the full cohort? These considerations are not to be dismissed.

Despite all these potential limitations, the results stemming from the RCT design are not to be dismissed so easily either. Accounting for both the effect of time, as well as the programme experience itself, is an important

feature of such design. Had we not collected the data from the control group, we could have easily (and mistakenly) concluded that the changes we did observe were true effects of the programme, and celebrated its positive outcomes. However, our rigorous analysis prevents us from doing so.

Of note, our study, by design, evaluated instrumental benefits of the programme, which has been previously shown hard to demonstrate. In particular in arts engagement, the intrinsic (art for arts' sake), rather than instrumental (transfer effects on non-arts cognitive abilities, e.g., evident through academic achievement) value has been emphasised in arguments for inclusion of the arts into curriculum (e.g., Winner & Cooper, 2000; Winner et al., 2013). Instead, we could have taken a different stance, looking broadly at children's development in cognitive and social domains through their engagement with creative arts (Goldstein et al., 2017).

What could we possibly extract further from the research we conducted? We also collected, but have not yet analysed, children's responses to open-ended questions probing their examples of learning something out of curiosity, teaching someone, and doing something creative. Such data is extremely valuable in its own right, regardless of the programme evaluation as such as children's responses to our open-ended questions were very insightful and varied. For example, in response to a question about what they learned independently, the responses included: "I learned how to play a song on the violin"; "How to script a game and how to do a front flip on grass"; "An efficient way on how to focus"; "How to tie a knot called the x knot". In response to a question about they did that was creative: "I made a shirt on this drawing app called pro create"; "My own sharp wooden weapon"; "I wrote a short story about a fire"; "Stickman movie". When asked about teaching someone about what they know: "I can teach people about how to move on a stage when doing acting / drama"; "Mainly about robotics and engineering"; "Teaching 2 one about the human body digestive system"; "I can teach people French since I speak French". What we can do now is analyse if these responses are different in some ways in the intervention vs control groups.

In sum, we reasoned that looking at whether the creative arts intervention is a useful educational practice contributing to children's self-perceived curiosity, creativity and teaching could inform our conceptual understanding of children's information-seeking and information-giving. Our lack of statistically significant results precludes us from seeing the programme-level benefits along these concepts. What we found using methods of psychological science turned out to not be very useful for the organisation, especially if we have been commissioned to conduct a programme evaluation. At the same time, we also cannot conclude that children are not benefitting from it at all. This highlights the need for more holistic approaches to programme evaluation.

Contact

m.bazhydai@lancaster.ac.uk

References

[1] See https://ludusdance.org/schools/the-alien-needs-our-help/

Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology*, 36(1), 4-12.

Baer, J. (2003). The impact of the Core Knowledge Curriculum on creativity. *Creativity Research Journal*, 15(2–3), 297–300.

Bazhydai, M., Silverstein, P., Parise, E., & Westermann, G. (2020). Two-year-old children preferentially transmit simple actions but not pedagogically demonstrated actions. *Developmental Science*, 23(5), e12941.

Choi, B., & Rowe, M. L. (2021). A parent gesture intervention as a means to increase parent declarative pointing and child vocabulary. *Infancy*, 26(5), 735-744.

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry*, 6(4), 267-285.

Ebert, M., Hoffmann, J. D., Ivcevic, Z., Phan, C., & Brackett, M. A. (2015). Teaching emotion and creativity skills through art: A workshop for children. *The International Journal of Creativity & Problem Solving*, 25(2), 23-35.

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465.

Goldstein, T. R., Lerner, M. D., & Winner, E. (2017). The arts as a venue for developmental science: Realizing a latent opportunity. *Child Development*, 88(5), 1505-1512.

Hardy, L. L., King, L., Kelly, B., Farrell, L., & Howlett, S. (2010). Munch and Move: evaluation of a preschool healthy eating and movement skill program. *International Journal of Behavioral Nutrition and Physical Activity*, 7, 1-11.

Karadağ, D., Bazhydai, M., & Westermann, G. (2024). Toddlers do not preferentially transmit generalizable information to others. *Developmental Science*, e13479. https://doi.org/10.1111/desc.13479

Mertens, D.M. and Wilson, A.T., 2018. *Program evaluation theory and practice*. Guilford Publications

Post, T., & Walma van der Molen, J. H. (2019). Development and validation of a questionnaire to measure primary school children's images of and attitudes towards curiosity (the CIAC questionnaire). *Motivation and Emotion*, 43, 159-178.

Qiu, F. W., Park, J., Vite, A., Patall, E., & Moll, H. (2024). Children's Selective Teaching and Informing: A Meta-Analysis. *Developmental Science*, e13576.

Rammstedt, B., Grüning, D. J., & Lechner, C. M. (2024). Measuring growth mindset: Validation of a three-item and a single-item scale in adolescents and adults. *European Journal of Psychological Assessment*, 40(1), 84.

Reeve, J., & Tseng, C. M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257-267.

Ruzek, E., Jirout, J., Schenke, K., Vitiello, V., Whittaker, J. V., & Pianta, R. (2020). Using self report surveys to measure PreK children's academic orientations: A psychometric evaluation. *Early Childhood Research Quarterly*, 50, 55-66.

Winner, E., Goldstein, T. R., & Vincent-Lancrin, S. (2013). Art for art's sake. *The impact of arts education*. Educational Research and Innovation, OECD Publishing.

Winner, E., & Cooper, M. (2000). Mute those claims: No evidence (yet) for a causal link between arts study and academic achievement. *Journal of Aesthetic Education*, 34(3/4), 11-75.

Putting Children and Young People at the heart of evaluations

JESSICA OZAN: HEAD OF EDUCATION, CHILDREN, AND FAMILIES AT IPSOS UK & COUNCIL MEMBER FOR THE UK EVALUATION SOCIETY
JAMIE ROWLAND: YOUTH FUTURES FOUNDATION





A Shift in Practice and Culture

The landscape of evaluation is changing. Gone are the days when children and young people were simply seen as subjects of research, passively observed and analysed. Today, there is a growing recognition that children and young people have valuable insights and perspectives that can contribute significantly to both the design and delivery of evaluation projects. This shift in practice and culture is driven by a number of factors, grounded in rights-based approaches and the UN Convention on the Rights of the Child and progressing through to a growing body of evidence of innovative methods at times commissioned by key players such as the What Work Centres.

This shift is reflected in the infrastructure of organisations such as the Youth Endowment Fund (YEF) and the Youth Futures Foundation (YFF), which have both established their own Youth Advisory Boards to inform evaluation work. These boards provide young people with a platform to share their views on evaluation and to help shape the organisations' work. There is also a growing openness on the part of evaluation commissioners, with government departments such as DfE, DCMS, and the Home Office contracting evaluations that include youth advisory groups and peer research. However, building the road to high quality participatory approaches is still at an early stage, and much more needs to be done to ensure that children and young people are meaningfully involved in evaluation projects, especially those from marginalised backgrounds.

Some of Jamie's most recent work includes a review providing an overview of current youth participatory approaches in research and evaluation funded by the YFF (Rowland et al., 2024).

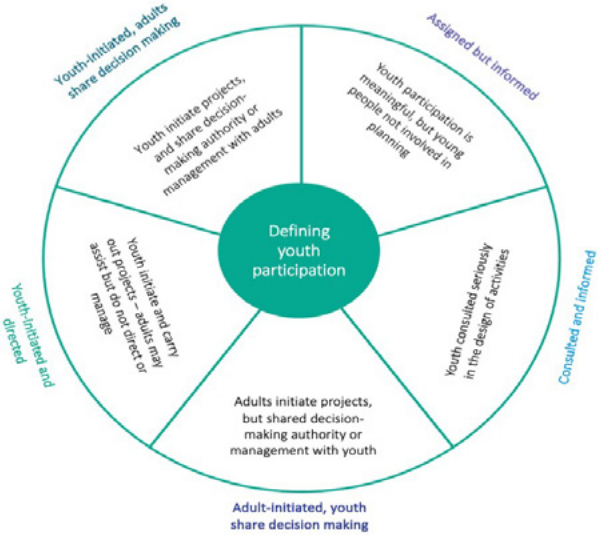


Figure 1 conceptualises the different ways young people's voice can be involved in evaluations.

The review identifies drivers of promising practice that can enable meaningful and ethical youth participation. Importantly, the review consolidates the fragmented evidence base to support informed implementation of participatory approaches. Some key considerations include:

Participation throughout the evaluation process

Ideally, children and young people should be involved at all stages of the evaluation process, from planning to data collection, analysis and reporting. This means working alongside young people on the evaluation design, methods, and interpretation of findings and challenges current practice which often engages with young people at touch points which are most convenient to research teams. Engagement in the early phases of design (i.e., proposal stage) can be particularly challenging as funding and ethical approval are not necessarily in place. It is therefore important to be honest and transparent about the remit of children and young people's role in evaluations. Often, young people may not be able to shape primary outcome measures established by commissioners on a Randomised Control Trial, but they may be able to inform secondary measures, how data can be best collected and shape discussion guides for the Implementation and Process Evaluation side of the study. This is an important point as it shows that children and young people's participation is a principle that transcends methodological divides. Where possible, providing opportunities for them to give feedback throughout the evaluation is highly beneficial. They can also have an active role in the dissemination of findings. This can range from designing child-friendly summaries, to presenting findings to key stakeholders or at conferences, and even co-authoring academic papers on the topic. The diversification of research outputs is likely to result in wider engagement in learnings, further democratising the world of research and evaluation.

Shifting power balance

Creating safe, supportive and empowering environments where children and young people feel comfortable sharing their views is key. This requires building trust, ensuring they feel respected and valued, and providing spaces for them to express opinions, even critical ones. This requires a shift in power dynamics that can often be challenging due to the assumed positionality of adult researchers, who typically hold the power due to both experience and age. This is why it is important to carefully consider the type of activities proposed and the settings in which they take place. Above all, participatory approaches need to be underpinned by reflexive, trauma-informed practice by adult researchers to surface the nuanced issues which may be present within typical evaluation design.

Appropriate methods

Using age-appropriate, engaging methods tailored to the developmental stage of the children and young people involved is essential. Creative, accessible approaches are important for enabling participation. Specifically, diversifying methods ensures that young people from all backgrounds are able to be heard through approaches that feel empowering to them. Traditional research methods may replicate young people's experiences of school assessments subsequently undermining the aim to realise youth ownership of projects. For those young people who may have had experience of youth justice issues, the use of Dictaphones and the directness of one-to-one interviews may also inadvertently re-affirm power dynamics that do little to create space for young people to articulate their experiences freely.

Feedback loops

Providing accessible feedback to children and young people on the evaluation findings, explaining the implications and next steps in ways they can understand is an important factor that is often overlooked. Children and young people are keen to understand how they contributed to a project, even when it is solely as participants. This may be as simple as circulating a child friendly summary of findings. It can also encompass implication workshops where young people work with evaluators to consider key findings, their implication for the programme or policy, and co-design recommendations. This approach is particularly useful as a process to validate findings and ensure recommendations are grounded and meaningful. It does however require finding a delicate balance between scientific rigour, policy priorities and constraints, and children's perspectives. This is not always a simple task.

Recognising and developing expertise

It is important to recognise from the onset that children and young people are involved as experts in their own lives. They bring a perspective that is different to that of the evaluation team. It is therefore imperative to offer recognition, compensation or incentives for young people's time and expertise. While there are differing views on payment, it is important to tangibly value their contributions. Furthermore, participatory approaches also present an opportunity for young people to develop key skills relating to research and evaluation that may serve them throughout their lives. To realise this more consideration should be taken about what skills young people can practically gain from involvement in evaluation projects.

Benefits

When done well, emerging evidence suggests that engaging children and young people in evaluation can yield several key benefits:

1. They provide valuable perspectives as experts that often differ from adults, shaped by their unique experiences. For example, children and young people's understanding of growing up in a digital world, with both access to knowledge and exposure to potential harm, will be very different to the experience of those responsible for evaluation design.

2. They can help improve evaluation tools and approaches to be more engaging, accessible and relevant to their needs and experiences. Young people's input strengthens the quality of data collection resulting in an evidence base that can more effectively promote systems/programme change.

3. Involving young people in data analysis and sense-making improves data quality and interpretation. Their perspectives help contextualise findings in light of their lived experiences. This provides an important balance to the subjective interpretation of adult researchers.

4. Ultimately, youth engagement contributes to the development of more effective programs and services that better meet young people's needs. Their involvement provides critical insights on what works and what doesn't.

There is a lack of evidence supporting the impact participation may have on children and young people. Anecdotal evidence suggests that they feel empowered and may gain new skills such as leadership ones. More needs to be done to ensure that we can speak with conviction on the benefits realised through participatory approaches. Specifically, there is significant space for participatory approaches to be monitored and evaluated for impact to bring greater rigor to the anecdotal evidence that is typically cited when considering its positive effect.

Conclusion

In conclusion, involving children and young people in evaluation holds the potential to ensure programs and services are meeting their needs, especially for those facing marginalisation. By engaging with young people, we gain invaluable insights to improve our approach, methods, data quality, interpretation of results and dissemination of findings.

This principle of youth engagement is now being extended to policymaking at the highest levels of government. The new Labour government has announced plans to create a National Youth Strategy designed to put the views of young people at the centre of decision-making on policies that affect them, notably through the establishment of a Youth Advisory Board. This represents a significant shift towards empowering young people as partners and leaders in shaping the policies that impact their lives. It also represents an opportunity for evaluators to shift perceptions. Children and young people's participation is likely to move from 'nice to have' to 'must have' in the coming years. In this context, it will be increasingly important to avoid tokenistic approaches and remember that evaluation design should always build on evaluation questions and context rather than preferred methodology.

Above all, participatory approaches must be designed and delivered in a safe and ethical manner. Ensuring that approaches are implemented in a highly reflexive manner and are led by the existing evidence avoids the potential for unintended harm. As such, participatory approaches should be developed and delivered in a measured and pragmatic way and evaluators should acknowledge where they may be positioned in the wider movement of participation, considering available learnings to ensure that youth voice is amplified through empowering approaches.

References

Rowland, J., Wills, E., Ott, E. (2024). *Youth Participatory Research: A Review of Reviews and Practice Guidance.* Centre for Evidence and Implementation.

Further reading

Dixon, J., Ward, J., &Blower, S. (2019). "They sat and actually listened to what we think about the care system": the use of participation, consultation, peer research and co-production to raise the voices of young people in and leaving care in England. Child Care in Practice 25(1), 6-21.

Richards-Schuster, K & Plachta Elliott, S. (2019). A practice matrix for involving young people in evaluation: Possibilities and considerations. American Journal of Evaluation.


UK Evaluation Society
EVALUATION
data.in.focus.
driving evaluation excellence
Join us for the biggest UK evaluation conference of the year!
19 May online
21 & 22 May in Glasgow
Book now
Early bird tickets available for a limited time!
2025
evaluation.org.uk

UK Evaluation Society

Discover more:
evaluation.org.uk

Disclaimer
The contents of this magazine do not necessarily represent the views of the UK Evaluation Society. Articles, diagrams and photographs are published with the assumption that contributors have obtained the necessary copyright and have read and agreed to the Guide for Contributors.

© Copyright UK Evaluation Society 2024.
All rights are reserved. Except where indicated, no part of this magazine may be reproduced or utilised in any form or by any means, electronic or mechanical, including photocopying, recording, or by information storage and retrieval systems, without permission in written form from the UK Evaluation Society. Designed and produced by Borderless Consutlancy Ltd.