Research Evaluation, 2025, **34**, rvaf033
https://doi.org/10.1093/reseval/rvaf033
Article

Failures in impact evaluation

Leila Jancovich[1,*], Ceri Pitches[1], David Stevenson[2]

[1]School of Performance and Cultural Industries, University of Leeds, Leeds LS2 9JT, United Kingdom
[2]School of Arts, Social Sciences and Management, Queen Margaret University, Edinburgh EH21 6UU, United Kingdom

*Corresponding author. School of Performance and Cultural Industries, University of Leeds, Room 1.10 Stage@Leeds, Leeds LS2 9JT, United Kingdom.
E-mail: l.jancovich@leeds.ac.uk.

Abstract

While many definitions of research impact exist, what most share is a belief in the responsibility of research, and researchers, to support positive change in wider society. But this article outlines the growing body of literature on both evaluation and impact that raises concerns with this approach. On the one hand an assumption of positive change may not only ignore the potential for negative impacts but also discourage research which is critical, exploratory or risky. The authors of this article further argue it may encourage narratives of success that mask stories of failure. This article discusses The FailSpace project, research funded by the Arts and Humanities Research Council (AHRC), which examined how evaluation might better identify, acknowledge, and learn from failures This article embodies the principles of FailSpace by reflecting on the failures, rather than successes, of this research project, regarding its intended impact based on findings of an autoethnographic evaluation of FailSpace's impact. In so doing the authors consider what might be gained from the inclusion of failure metrics in impact evaluations.

Keywords: impact; evaluation; failure.

1. Introduction

Finding ways to utilize research for the benefit of society has long been a priority for some researchers. But in several countries, such as Australia, Canada, and the UK, the 'impact' of research beyond academia, as opposed to the ways in which research contributes to academic advances across and within disciplines, is increasingly considered an essential element of all research processes, particularly those which are publicly funded.

In the UK, which is the focus for this study, reporting on impact is a requirement for research funding from UK Research and Innovation and contributes a significant percentage of the measurement for the national assessment of Higher Education institutions through the Research Excellence Framework (REF).

While many definitions of research impact exist, in essence they have been shown to broadly espouse the same core message (Penfield et al. 2014), that is, the potential and power of research to effect change in an environment beyond academia. Bayley describes impact as the 'the conversion of [new knowledge] into change in the wider world' (Bayley 2023: 13), while Reed crystalizes his philosophy of impact as 'the good that researchers can do in the world' (2016a). This idea of public good is consistent with definitions provided by research bodies in the UK. REF, for example, defines impact as '… an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia' (REF 2019: 68). While the UK Research Council describes it as '… the demonstrable contribution that excellent research makes to society and the economy' (UKRI 2023).

But concern has been raised by researchers, that measuring impact in this way may influence the type of research undertaken in the first place (e.g. Chubb and Reed 2018; Bayley 2023). It may for example, encourage research that is palatable to those whom it wants to influence, or easy to implement, over research that is critical, exploratory or risky. This may not only limit the range of research undertaken but also reduce the opportunity for radical solutions to wicked problems[1] or the systemic change that is needed for more long term and meaningful societal impact.

Furthermore, based on an examination of the FailSpace research project, which was undertaken by two of the authors of this article between 2020 and 2023, we argue that the language of benefit and contribution used by policy makers encourages measurement of positive outcomes, that ignore the learning from the complex and contingent nature of impact in which the benefits for some may have negative consequences for others. In addition, the assumption of positive change leads to assessments that rarely acknowledge when the anticipated impacts fail to materialize, so fail a second time in failing to address or learn from this.

1.1 FailSpace

The FailSpace Project, originally known as *Cultural Participation: Stories of Success, Histories of Failure*, was a 3-year exploration of the levers and barriers within the cultural sector to identifying, acknowledging, and learning from failures (Jancovich and Stevenson 2023). The research resulted in the development of an evaluation framework and associated toolkit intended to support more honest and open conversations about failures in the publicly funded cultural sector between artists, arts managers, funders and policymakers.

Receiving almost £350,000 in funding from the AHRC, the research and associated outputs have been made available open access to academics and cultural professionals. To date over one thousand professionals have taken part in FailSpace workshops and over two thousand have downloaded the FailSpace toolkit. By May 2023, the associated book (Jancovich and Stevenson 2023) had been accessed thirty-eight thousand times. As a result of this profile the researchers have been invited by their host institutions to write up an

impact case study of the work for potential inclusion in the institutional submission to REF2029.

But, the authors posit that the template for measuring impact within academia creates similar challenges for researches in acknowledging failure, to those of evaluation in the cultural sector. This paper aims to challenge this, by embodying the principles of FailSpace to reflect on both the successes and the failures of achieving the intended impact of the research project, which were articulated in the "pathway to impact" (Wilsden 2020), submitted as part of the original funding application. These included the ambitious objectives of delivering attitudinal, behavioural and policy change in both professional discourse and practice.

By sharing the failures of impact in this case study the article considers what might be gained from the inclusion of failure metrics in impact evaluations. In doing so the authors recognize the different connotations associated with failure in different academic fields. Embracing "intelligent fast failure" (Matson 1991) is advocated in some business literature as a process from which to learn. But we acknowledge that in some contexts, such as the wrong medical intervention, the consequences of failure may result in harm that cannot be so easily rectified. However, we assert that failing to acknowledge and reflect on when things go wrong, misses an opportunity for learning which may avoid repetition of the mistakes that may lead to failure. Thus this article argues that for impact measurement to be of greatest use, it must not only record positive change but also facilitate it. To do so, future approaches to evaluating research impact must explicitly seek to reflect on failures and the learning that can arise from them, alongside evidencing success and identifying 'what works'.

2. Understanding the context of research impact in the UK

Smith et al. (2020) detail the emergence and rise of the specific notion of 'research impact' in the UK from the 1980s to the present day, noting the central role of the New Labour government in 1997 in effecting a shift towards evidence-based policymaking, compounded in 2000 by the then Education and Employment Secretary David Blunkett's positivist vision of research and its uses.

Where 'previously, academic research was generally treated as of value in itself' (Hammersley 2014: 345), following the 2006 Warry review, funding applications submitted to the UK Research Council included the requirement for a 'Pathway to Impact' plan (Wilsden 2020), identifying intended benefits and beneficiaries, and reporting on how these had been achieved at the end of the research. These requirements for discrete impact plans remained in place until 2020 when UKRI discontinued them, stating that 'impact is now a core consideration throughout the grant application process' (Smith et al. 2020: 21).

Alongside this in the UK, is REF, an ex-post peer review process to assess research quality and allocate public funding at institutional level according to the results. REF has placed increasing value on the impact component—rising from a 20% weighting in REF2014, the first time impact was formally assessed, to 25% in REF2021, a weighting that will be maintained for assessment in 2029, with the component being renamed as 'Engagement and Impact' (REF 2025:10).

Early resistance to the impact agenda was expressed through concern that the prioritization of impact would lead to an instrumentalized view of research that limited curiosity-driven endeavours (e.g. Chubb and Reed 2018; Smith et al. 2020; Kidd, Chubb and Forstenzer 2021). Such concerns seem well founded as funding allocated to address contemporary societal challenge initiatives has increased, over pots responsive to researcher interests. The UKRI Challenge Fund, for example, is investing £2.6 billion of public money with £3 billion in matched funding from the private sector, for projects that are directly aligned to the government's industrial strategy, while AHRC's responsive mode of funding has stagnated.

Persuasive arguments have been made (e.g. Chubb and Watermeyer 2017; Derrick 2019) for how this emphasis on rewarding research that delivers positive societal benefits is changing the behaviour of researchers, changing how research is conducted, and how knowledge is created. The pressure on researchers to 'deliver the goods' (Chubb and Watermeyer 2017: 2361) is resulting in what some perceive as the marketization of research, as well as a tendency to favour short-term, easier to measure impacts over longer-term, more complex impacts that are less easily evidenced through simple metrics (Bandola-Gill 2019). Furthermore, it has long been recognized (e.g. Spaapen and van Drooge 2011; Bornmann and Marx 2014) that impacts within the disciplines of the social sciences, arts, and the humanities are generally difficult to measure. This arguably leaves these disciplines less well-placed than their counterparts in the natural, applied, and formal sciences regarding evidencing their return on investment.

Derrick (2019) further suggests the existing focus in much impact evaluation on innovation and economic development, betrays government biases towards these elements over 'more nuanced concepts of cultural impact' (387). Economic impacts are thus potentially being valorized over other kinds of value to the public.

2.1 The UK's approach to the assessment and evaluation of impact

In practice it has been shown that measuring impact presents a range of significant challenges (McCowan 2018; Lauronen 2020; Smith et al. 2020) as it is 'elusive, complex and context-dependent' (Smith et al. 2020: 34). The variety of methods and tools used for the ex-post assessment of impact is broad [see Reed et al. (2021) for a comprehensive discussion of these] but it has been suggested (Lauronen 2020) that the case study, with its blend of qualitative and quantitative methods is the most useful for assessing societal impacts. Case study is the method currently favoured by the REF[2] and judged by Bornmann and Marx (2014) to be the best-known national impact evaluation system in the world. Here, the approach is the expert peer-reviewed, evidence-based, narrative case study, assessed on the criteria of 'reach' and 'significance', and which in REF2029, will also include 'rigour'.

While the case study has been identified as the most widely used method for assessing societal impact (Viana-Lora and Nel-lo-Andreu 2021) it has also been critiqued for several key concerns, not least for being very costly due to the excessive labour requirement associated with its production (de Jong et al. 2014). It has been noted that there are also problems of attribution, of constructed causality, of standardization shortfalls and of the lag between the occurrence of societal impact and the original research (e.g. Gunn and Mintrom 2017; Lauronen 2020; Smith et al. 2020). As summarized by Lauronen (2020), there is often a mechanical approach to the evidencing of impact case studies where,

particularly in the arts, humanities and social sciences, the emphasis on linear feedback and an understanding of impact that can only be seen to flow *directly* from research outputs (Gunn and Mintrom 2017), provides a source of tension for less predictable impact cycles.

2.2 The potential for negative impacts

The definitions of impact outlined above assume inherently beneficial impacts from research. But Reed (2016) observes that researchers have a responsibility to anticipate and mitigate the potential harmful effects of their work. While Izzi and Murray (2019) question the extent to which negative impacts should be tolerated, Reed et al. acknowledge that both positive and negative impacts are contextual as 'benefit for one group in one place, time and culture, may be perceived as damaging the interests of others' (2021: 2). Penfield et al. (2014) muse on this complexity, highlighting both the positive and negative impacts in the example of the drug thalidomide, which, in the 1950s alleviated morning sickness in pregnant women, but resulted in birth defects in babies, thus an example of a failure that had significant and life-changing repercussions. However, despite being withdrawn in the 1960s, it has since been found to have beneficial effects in treating certain cancers and symptoms of HIV and is back in use in this context. This example illustrates the contingent nature of impact and raises questions about how and when it is measured. One attempt to mitigate the problem of negative impacts is Reed et al.'s (2025) '3i framework' which provides an analytical tool for understanding interest, influence and impact, enabling researchers to assess the potential benefits and risks of their partnership work.

But McCowan (2018) has argued that the normative definition of benefit in practise encourages researchers to identify impacts as improvements for *some*, which may not ultimately provide net good for society as a *whole*, and which training or such a framework would not remedy. Building on this idea and united by a shared concern over how research is sometimes distorted politically, Derrick et al. (2018) developed the term *grimpact*, which they argue arises, in part, because of the evaluation and auditing processes that put emphasis on only the *positive* consequences of research. But they suggest it is also a consequence of the 'ethically dubious applications/impacts' submitted by researchers (Smith et al. 2020: 49), citing the 1998 report that fraudulently linked the MMR vaccine to autism leading to widespread rejection of the vaccine in the UK and the subsequent re-emergence of measles in England. Suggesting that all researchers may be complicit in this problem, in a later paper, Derrick and Benneworth (2019) forecast the likelihood of increased instances of grimpact in academia as political and public pressure for researchers to quantify and justify the contribution of their research intensifies.

This increasing prioritization of the outcomes of research rather than the research outputs themselves also has concerning implications *within* academia. Building on criticisms of the impact agenda by Chubb and Watermeyer (2017) and Chubb and Reed (2018), Kidd, Chubb and Forstenzer (2021) highlight the damage that may be done to the 'motivations, dispositions of academics and their students' (151). Characterizing their standpoint as 'epistemic corruption' (2021: 152), Kidd, Chubb and Forstenzer argue that measuring impact may deprive (i.e. corrupt) academic endeavours, and indeed academics themselves, of their integrity. Using examples from their own interviews with academics in the UK and Australia, they show how 'epistemic vices' such as 'inaccuracy, insincerity and hyperbole' (158), lead to exaggerated claims of impact and a willingness to succumb to 'disingenuous' (160) research practices such as providing or endorsing neat accounts in funding applications of what *will be* discovered and to *what effect* before the work is even begun. Such concerns about the consequences of a culture of celebration (Jancovich and Stevenson 2023) in culture sector evaluations set the context within which the FailSpace project first examined the implications of consciously evaluating failures in order to break this cycle and inform policy learning. This built on research on failure that exists in other fields but has to date largely been absent from evaluations or discussions about impact.

3. Researching and evaluating failures

Despite the importance placed on evaluation of funded research and indeed of policy goals, to not only assess impact but also support evidence based policymaking (Sanderson 2002) the academic research on failures in the public sector is relatively sparse.[3] Hogwood and Gunn (1984) argue that evaluations must be goal-oriented, explaining how and why policies and practices succeed or fail. Yet it has been noted that while failure to achieve goals is widespread, it goes under-reported in evaluation, thereby reducing opportunities for learning (King and Crewe 2013). In practice, Andrews (2018) argues evaluations are often biased, focusing on reporting outputs and outcomes rather than measuring the extent to which they have addressed the problem they seek to address.

The evidence base on which such evaluations are built has been shown to be articulated by those controlling the narrative, balancing empirical information and emotive appeals (Taylor and Balloch 2005). This is due to the fact that policy actors are driven less by evidence and more by 'political learning' (May 1992), which focuses on building positive perception of their interventions. This in turn leads to a preference for success stories over critical reflection. The purpose of such evaluation is not learning but 'making the case' for public investment. This leads to a reluctance to acknowledge failures which in turn hinders learning and improvement (Howlett, Ramesh and Wu 2015). Consequently, policies and projects may repeat past mistakes (Howlett 2012), wasting resources on ineffective activities. The FailSpace project argued that more honest evaluations are crucial for developing policies that make significant, sustainable contributions to a more equitable cultural sector (Jancovich and Stevenson 2023). However, successes or failures are rarely clear-cut as the two are not binary but involve varying degrees of achievement (Bignell and Fortune 1984). Within this context learning from failure may be difficult due to the complexity of the multiple dimensions and interests that need to be considered in their assessment.

As shown through the literature on impact above, academia may suffer from similar challenges, in not only overlooking negative impacts but also ignoring failures to have the intended impact or deliver on some other aspect of the research proposal. This not only limits the ability to learn but leads to a situation where researchers underestimate how frequently others experience research failures and thereby

personalize a fear of what failures to generate impact indicate about themselves (Sousa and Clark 2019).

Connectedly, Caffrey (2023) highlights the relational damage that can be done by universities claiming full attribution for positive real-world change rather than as a committed partner in a respectful, equitable relationship with others. He advocates for celebration of the 'messy, imperfect processes' involved in research that enable learning from failures as well as successes. Smith et al. (2020) conclude their co-edited book with recommendations for an alternative, evidence-informed approach to research impact, including the need to 'create spaces in which valiant failures are celebrated and learned from' (203) by both funders and universities, while Lonsdale and Seruagi (2023) highlight the need to be permitted to reflect on failure as part of impact evaluation.

It is in response to such proposals for greater honesty, transparency and critical reflection that the current article has been written. In doing so, it aims to ensure that the FailSpace project embodies its own principles through reflecting on the failures regarding the desired impact of the research.

4. Methodology for this article

This article offers an autoethnographic account of the FailSpace project as a case study, through critical reflection-on-action, with a specific focus on failures to deliver impact.

Autoethnography is a qualitative research methodology that allows researchers to use their lived experience as a primary source of data to explore given phenomena. It offers a nuanced, first-person perspective that highlights the interplay between situated practice, professional context, and the research process itself (Chang 2008). It is particularly effective for examining research practice, as it provides deep, reflective insights into the processes, challenges, and transformations experienced by the researcher. This introspective process fosters a deeper understanding of how personal and professional experiences shape research approaches and outcomes (Adams, Holman Jones and Ellis 2021). Such insights can enhance the transparency and credibility of research by offering a less 'polished' presentation of its strengths and weaknesses.

Relatedly, self-reflection on practice, also known as reflective practice, is a research methodology that involves critically examining one's own actions, decisions, and experiences in a systematic manner to improve professional practice and personal understanding (Boud, Keogh and Walker 1985). This approach is widely utilized across various fields with the aim of fostering continuous, transformative learning and development through challenging and changing existing frames of reference (Mezirow 1991). Reflective practice is central to autoethnography as it involves self-examination and narrative inquiry.

This autoethnographic approach was mobilized through reflective conversations between the authors of this paper, which took place at the end of both the original research and the follow-on impact project. These conversations were informed by the reflections and opinions of people engaging with FailSpace, whose thoughts we recorded through a range of means, and at stages throughout the project to ensure that reflection was an iterative process throughout the research journey as shown in the table 1.

Using field notes from all these sources, we (Jancovich and Stevenson) undertook a critically reflective analysis of the impact of our research, employing an adapted version of our FailSpace framework (Fig. 1) to structure our discussions.

As explained in the associated research outputs of Failspace, the framework above was intended to provide a method for the cultural sector to reflect on failures in a structured way, by 'employing language that allows for a more nuanced and critically reflective account of where failures may have occurred, for whom, and to what degree' (Jancovich and Stevenson 2023: 153) and so is well suited to our own reflections on impact generation.

The framework recommends those leading projects to define from the outset what differing degrees of success and failure would look like across five facets. These should then be returned to during and after the project to reflect, with different stakeholders, on which description of success and/or failure each element of the project most resembles and relatedly, to evaluate what learning or action needs to be taken in response to this. We were unable to employ our framework at the outset, as we had not yet conceived of it, and so instead employed it as part of the reporting process on the outcomes of the original funding application.

This therefore involved comparing the newly created Failspace framework against the outcome based aims we had articulated in our original funding application which were summarized as effecting attitudinal, behavioural and policy change within the cultural sector. We retrospectively outlined on a grid (Fig. 2) what the differing degrees of success and failure might have looked like in relation to our aims across each of the five facets of the FailSpace framework. We then reflected on which of our descriptions our project most resembled as we completed the research phase and highlighted these on the grid in bold text.

Critically reflecting on the degrees of failure/success regarding the intended outcomes led to a successful application for follow-on funding from the AHRC to support the further development of impact from our work, thereby ensuring a longer-term legacy of the work through embedding the FailSpace framework within policy and practice. Once the follow on project was completed, and as part of developing a draft impact case study for REF2029, we returned to the framework. This involved reviewing the evidence of changing attitudes to failures; organizational change in professional practice; and policy change through building the possibility of failure into funding applications and evaluations. In doing so, we were also gathering evidence of where we had failed to manage each of these things to the level that we had hoped. The following section reflects on the findings from these processes and considers some of the learning from our failures that can be taken from them.

5. Failures to change attitudes and encourage openness

As stated above the FailSpace research was underpinned by existing literature on the value of talking openly about failures. This informed the approach to engaging with research participants throughout all our activities. However, as the project researchers (Jancovich and Stevenson), we failed to engage with the body of literature which considers the psychology of failure. Doing so might have prepared us for the deep-seated fear of personal failure encountered from research participants.

Despite the fact that words like cathartic, essential, and relief were often present in the feedback on our workshops, words like discomfort, shame, and even trauma were

Table 1. Failspace research journey and data collecting reviewed to inform critical reflection on Failspace's impact.

Date	Failspace research journey	Method used for research and evaluation of impact	Questions/themes explored
2020	Phase 1 of Failspace research	8 workshop with 150 people 80 interviews 3 case studies 100 postcards sharing stories of failure	Exploring 1) meanings of failures, 2) learning from sharing stories of failure 3) levers and barriers to talking about it
2021	Culmination of phase 1 of research	Toolkit and workshops to test emerging framework	How to embed learning from failure in practice
		Conference and special edition to share findings	Encourage practitioners to identify, acknowledge, and learn from failure
		Application for AHRC follow on funding to build impact	First use of our framework to acknowledge our failures in impact (Fig. 2)
2022	Phase 2—follow on activity to develop impact	8 Failspace champions trained to disseminate framework and report on impact	Written reports on how many workshops delivered for who and evidence of impacts
		Partnership with 6 funding bodies to embed framework in designated funding stream Survey sent to 300 people who had engaged in the Failspace research so far with 100 responses	Regular meetings to reflect on experiences of integrating FailSpace framework in their work questions on 1) how often they discuss failure and with whom 2) levers and barriers 3) how Failspace is changing practice
2023		Workshops delivered by research team and champions to embed framework First annual review with Champions	Evaluation of who attends and feedback on what has been learnt 90 min focus group exploring experiences of embedding FailSpace
		Repeat of survey sent to 500 but only 7 responses	Same questions as first survey
2024	End of funded research	Second annual review with champions	90 min focus group exploring how to maintain momentum without funding
		Informal conversations with independent evaluation consultants Reflections between the research team and Director of impact at host university	Email and teams meetings with a range of people using our approach to capture evidence of impact 6 monthly evidence capture and review meetings for potential impact case study

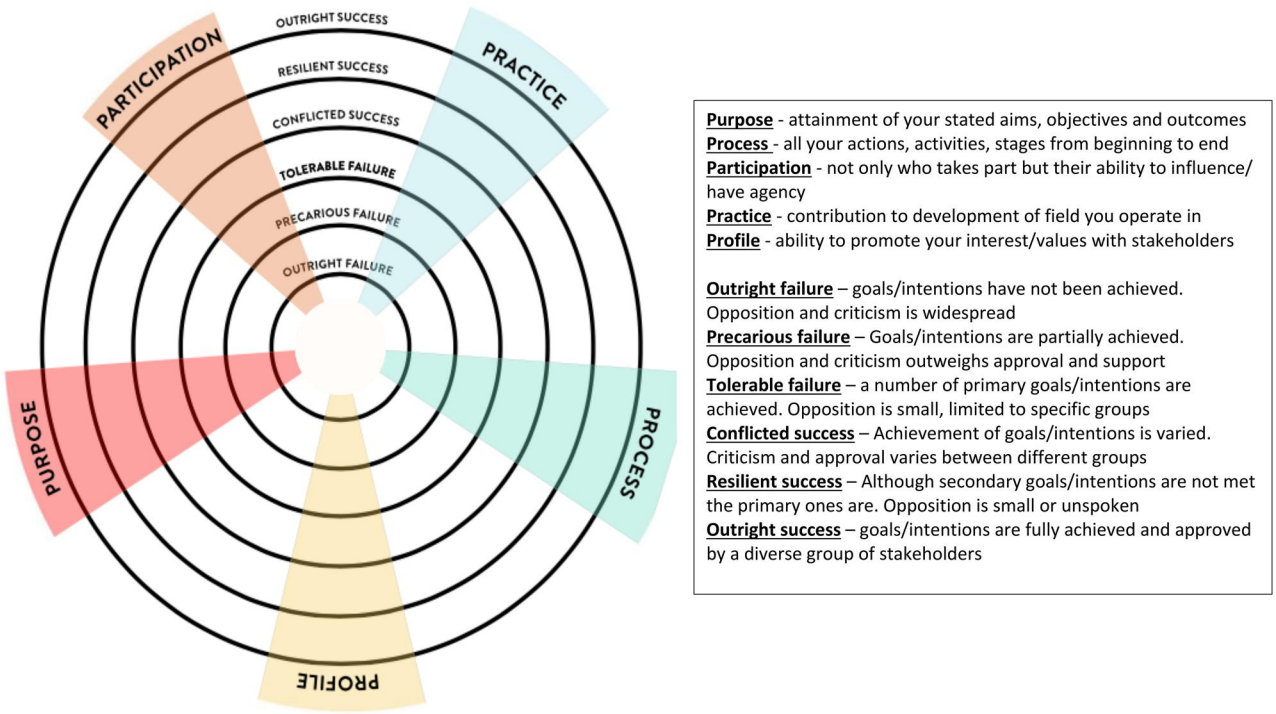

Figure 1. Wheel of Failure from the FailSpace framework (Jancovich and Stevenson 2023: 143).

regularly used by participants in response to simply talking about failure. Some even suggested the research project was dangerous in the current economic and political climate, in what many perceive to be a precarious/vulnerable sector. Even one of our Champions said talking about failure can initially 'take the conversation in a really unhelpful way … down and lacking energy'. While it was clear that many enjoyed taking part in our workshops and interviews, we

	Outright Failure	Precarious Failure	Tolerable Failure	Conflicted Success	Resilient Success	Outright Success
Profile –the reputation and the ability to promote	We fail to raise interest in our impact plans	We engage interest but fail to convince	**The programme has visibility but its impact is less clear**	The impact is talked about as both a success and failure	Key stakeholders acknowledge that our work has positive impacts	We are recognised as affecting sustainable change
Purpose - the attainment of stated aims, objectives and outcomes	We fail to get people talking about failure	**The value is acknowledged but there is no evidence people are doing it**	There is a short-term increase in people talking about failure with colleagues	More people value talking about failure but there is no evidence it is changing practice	More people value talking about failure and emerging evidence it is changing practice	Talking about failure has become normalised and is changing practice
Process - the actions, activities and stages of delivery	We fail to identify partners to work with us on impact	We identify partners but struggle to work together	**Our partners only create limited opportunities to talk about our approach**	Our partners test our approach but don't share their learning	Or partners test our approach and share their learning	Or partners test and share learning, leading to wider adoption
Practice – impact on academic practice	We do not gain any new knowledge about generating impact from research	**We have new knowledge but feel less confident in our abilities to generate impact than before**	We gain new knowledge about creating impact but do not prioritise it in future	We gain new knowledge and use this to create impact but do not share it	We gain new knowledge which we use and share	Our new knowledge on impact changes how others do impact work
Participation – level of agency partners have in shaping or evaluating the programme	**We fail to make partners feel any ownership of our framework**	Partners exhibit some ownership but we fail to adapt in response to feedback	Partners exhibit some ownership but don't show any desire to develop it for their own needs	Partners exhibit a good degree of ownership but have adapted in ways we are uncomfortable with	Partners exhibit ownership and have worked with us to embed the framework	Partners have full ownership and are confident without our continued

Figure 2. Assessment of the degrees of Failure/Success for the FailSpace project at the end of the initial research funding (in bold).

failed to put in place mechanisms to deal with those for whom a feeling of trauma may have persisted, as neither we, nor the reviewers who granted ethical approval for the research, had considered this need.

Furthermore, although the feedback on our workshops may suggest some immediate attitudinal change given the way in which, for many, their initial reticence seems to have been replaced with a recognition of the value of talking about failure, it is less easy to evidence whether this change in opinion is sustained or has resulted in any changes to their practice. The survey at the end of the initial research project, which attracted 100 responses demonstrated short term impact on those who responded; with for example sixty-eight per cent saying their involvement in the research had challenged their thinking and seventy-three per cent feeling that it had built their confidence. But a repeat survey a year after publication of the research outputs failed to achieve more than a handful of responses. Although it is not possible to know if this relates to the challenge of academics expecting people to give time to meet our need to evidence impact, one person did send an angry email saying it was unreasonable for salaried researchers to be asking unpaid participants to complete follow-on surveys. Several other people asked to be removed from the mailing list at this point. Thus, we may presume that we are failing to convince many of those who were initially positive about our work, that there is any value for them in continuing to engage with it.

In conversation with our champions there were differences of opinion about the extent to which any attitudinal change was extending beyond those who directly took part in our research or workshops to their colleagues and peers. One said 'we're hearing a bit more [about failure] within the art sector, people saying I've heard about [FailSpace]'. While another felt 'wider conversations about failure are not necessarily happening as a result […] it's not necessarily causing a culture shift beyond the niche of those interested' (Champion). The research team recognize that a weakness of both the research and follow-on activities is that we relied on the choice of individuals to participate, meaning that we were primarily working with those who were already interested in the topic and arguably open to having their opinions changed. This is likely to be even more so with those who take the time to read the research outputs themselves. So, although FailSpace may offer the language and tools to have more honest conversations about failure, its ultimate success relies upon those who engage with it to not only encourage others to do so, but to advocate for its wide spread use within the sector. The reality of a time limited project with finite resources means that any impact successes rely on our partners as much as the original research findings.

Another element of how we hoped to generate attitudinal change was through encouraging people to share stories of failure more publicly. However, in the initial research phase we committed to 'safe' spaces where participants could talk about failure anonymously. This not only failed to model our call for openness but also unwittingly encouraged a focus on personalizing blame and introspection. To address this, in the follow-on phase, we asked people to waive their anonymity and created opportunities for them to share stories of failure more widely. For example, participants were able to post stories of failure on the FailSpace website and we invited organizations to write short case studies of failure that we would arrange to have published in Arts Professional, a publication for industry

professionals. While we got a small number of posts on our website, most were apologies for what were seen as inevitable failures, rather than reflections on learning and we could get no one to agree to write something for publication. A similar problem occurred when it came to writing our monograph, where we failed to persuade any of our case study organizations in Scotland to give us permission to include them as a named organization in our publication. Time and again we heard that while the research might have helped individuals or teams acknowledge and discuss failures they were still unwilling to 'put something on paper that someday someone else could pick up and use against them' (Champion).

Despite this, we have found evidence of an increase in references to failure in different forms of media, including articles in trade and academic journals and conversations on social media, some of which cite our research outputs. Likewise, our research informed the Centre for Cultural Value's Evaluation Principles, and the How to Fail Well guide we created for them is one of their top reads and credited with driving traffic to the website. On this basis it would be tempting to argue that this interest in failure is all attributable to FailSpace. But as highlighted by Lauronen (2020) it is hard to know whether this is the case. It may equally be that interest in our research is the result of a shift already happening in the cultural sector and a long standing frustration with 'the lack of honesty in evaluations that function more as advocacy documents' (Funding partner). Either way, while the media interest in failure may demonstrate success in our research gaining a profile it is disingenuous to claim it is proof of attitudinal change. Making such claims encourages dishonesty and, as Caffrey (2023) highlighted, undermines the potential for making partnerships with others who share our interests.

The funded project ended with a national conference in Edinburgh to share the findings of FailSpace's alongside other approaches to learning from failure. The conference sold out, demonstrating a significant interest and awareness of our project. Yet on the day, it was apparent that while delegates were happy talking about failure as a topic, and were vocally positive about the research, there was less comfort in talking about what would or should change in the sector as a result as this. One respondent who talked about using the framework said "my practice itself has not changed, but the way I talk and think about it has" (survey). Of itself this comment might suggest some attitudinal change, as they are more reflective in their thinking. Conversely, it may suggest that the language of failure is being appropriated to justify the right to fail without the responsibility to change, with the use of our tools being a performative act. A core principle of our research was that while failure should be seen as something to learn from not to punish, equally the right to fail carries a responsibly to change behaviours. Yet our research may have failed to get the majority of those who engage with it to accept this responsibility.

5.1 Failing to change organizational practice

In line with the literature, the FailSpace research showed that evaluation, just like impact measurement, is strongly associated with justifying funding rather than learning to improve future delivery or support organizational change. As a result, evaluations often perpetuate narratives of success that protect the status quo. FailSpace set out to challenge this way of working and create behavioural change within the cultural sector, primarily through changing organizational practice. From the outset the research team acknowledged that embedding the FailSpace methodology in practice would also mean handing over control of how this was communicated to those working in the cultural sector.

As we developed the framework and tools to support its delivery, we worked on the basis that organizational change requires buy-in at all levels, for new ideas to be embedded. As an evaluation consultant who used our framework said 'change is a very bureaucratic process. [There are] different people running different schemes. They all have to be brought on board'. However, as a small research team, in which the PI and Co-I were only funded to work on the project 1 day a week, we were aware from the outset that a significant limitation would be our ability to 'reach' as many people as possible in as many organizations as possible. Therefore, the workshops delivered by the research team were conceived as 'train the trainer' opportunities in which we would disseminate our research and introduce people to the FailSpace tools. Our assumption was that these would give those taking part the knowledge and confidence to take the FailSpace methodology and tools into their workplace and encourage their colleagues to adopt it. But, while the majority of workshop participants who responded to the follow up survey said taking part had changed their thinking, fifty-two per cent said they only used the framework for self-reflection and hadn't brought any others into the reflection processes or introduced colleagues to our approach. While the workshops may have influenced individual practice they appear to have failed as a pathway to delivering widespread organizational change.

Throughout the period that the project was funded the research team were proactively promoting the research through various channels. During 2022, the second year of the project, this promotion resulted in regular requests for the team to deliver workshops, which we increasingly found difficult to accommodate due to the limitations on our time. The follow-on funding allowed us to train eight culture sector professionals to take on the role of freelance FailSpace Champions, to increase the capacity to disseminate our findings and embed our framework in practice. Despite the funding only allowing us to pay them for their contribution for 1 year, several of the Champions have continued to facilitate failure workshops and discussions with their own industry networks. Several have also said that working as a Champion has changed their wider practice as freelancers. But once the funding stopped and none of the researchers had dedicated time to focus on promoting FailSpace, the inquiries about workshops slowed down. The Champions must now proactively seek out opportunities and ask interested parties to pay for workshops which, during the period in which the project was funded, we were providing for free. In the current financial climate it seems particularly difficult to get organizations to pay for something that is not an essential requirement. Our hope that the FailSpace methodology would become embedded in practice is hampered by the nature of short-term funding which fails to support long-term impact. However, several evaluation consultants have built the framework into their practice. For example, Thinking Practice have combined it with the *Most Significant Change* approach, originally developed by Rick Davies. This suggests an alternative pathway to impact, which we failed to consider, was to focus more on consultants and encourage them to use our material to enhance the range of evaluation methodologies they could offer to clients.

Through reflection with Champions on the apparent failure to change organizational practice there was a support for the

view 'organisations want to use it, but they won't be the first to use it'. One Champion said they were asked 'can you get a guarantee that [other organisations] are all going to use it as well?'. Another gave the example of a member of staff at an organization they were working with who had acknowledged the potential for failures in a draft funding application but was pressured by peers to take it out. This suggests that even where some in an organization have been persuaded of the need to change ways of working, peer pressure can limit the extent to which they can affect such changes at the organizational level.

Resistance to change was most often expressed in relation to the lack of belief that funders would accept failures without associated 'punishment' for organizations and individuals who did not deliver on the outcomes that they had committed to when accepting funding. While it was regularly acknowledged that this necessitated a selective approach to evaluation it was also noted that 'success is just keep on doing the thing that makes you get the funding'. As such, the majority of our participants said that in order for them to acknowledge failures the FailSpace project would need to 'get funders to change processes' and policy makers to acknowledge their own failures and show how they were changing their approach in response. As one of our Champions noted, 'cultural change will happen when the people with power are able to accept some sort of change or adapt their practices'. We have argued in our monograph that such attitudes may be as much to do with a failure of those working in the sector to recognize their own agency bringing about change as many 'struggle to talk about failure as it related to their own work but found it rather easier to talk about funder failure' (evaluation consultant). However, these attitudes were pervasive and we acknowledged the need to persuade funders to change their funding processes to support change.

5.2 Failing to get funders to model the changes they supposedly want to see

From the outset a core aim of the research was to examine the role failure plays in policy learning. Policy makers and funders were invited to their own workshop at the start of our research and several were interviewed as the research developed. Unlike the reticence about adopting the FailSpace methodology that was expressed by those receiving funding, the majority of funders we spoke with saw the potential value it had in providing them with 'an accessible way to shift the conversation about evaluation […] to be less about accountability and impact to something more reflective' (Funding partner). They also challenged the perception common among artists and arts organizations that funders are not open to hearing about failures from those they fund. Indeed, several suggested they would be more likely to fund someone who showed they had learnt from previous failures than someone who wrote what many described as unbelievable evaluation reports.

But when pushed for examples of where they had acknowledged and learned from failures in the work that they funded, the funders we spoke with generally failed to think of any. They also failed to demonstrate how the language and structure of application and evaluation forms invited those they fund to do so. As part of our follow-on funding we therefore invited a small number of funders to work in partnership with us to change some of processes. The idea was to identify a specific funding programme for which we would use the FailSpace methodology and framework as the basis for redesigning application forms and evaluation parameters. By so doing we planned to support policy learning within that programme, as well as demonstrate the value of this approach being adopted as standard practice, not only by the funders we were working with but by others who would be keen to adopt a new form of 'best practice'.

Arts Council Wales and Creative Scotland agreed to take part, but the biggest of the national agencies invited, Arts Council England, declined to be involved. At a local level, Edinburgh City Council agreed, whereas Leeds, Bradford and Calderdale Council were contacted, but we failed to persuade them to take part. Finally, despite contacting several trusts and foundations, including the Paul Hamlyn Foundation, who were part of the initial research process and endorsed our findings in principle, only the Art Fund signed up as partners. While the partners we did get provided an opportunity to pilot how our framework might work, our failure to get more funders on board limited the potential for the wider impact we hoped for.

The partnership agreements with funders included the commitment to embed the Failspace framework into one of their funding programmes. However in two cases the contacts we were working with appeared to lack the authority or the time to instigate any changes to application forms or evaluation criteria. Those who were able to altered the language of application forms from asking what applicants will do with the expectation that they would measure whether they did it, to asking them to consider what success and failure would look like at the outset. Applicants were then asked to evaluate the degrees of success and failure across different facets at the end of their projects. Finally, and crucially, they were asked to reflect on what was learnt from this and what they would do differently in future.

The funders who did this hoped that this would provide learning for them about the effectiveness of their intervention and 'lead to improved relationships and more strategic projects'(Funding partner). We all acknowledged that for this to work it was important that Failspace replaced their previous approach, so it didn't become an additional reporting burden on funded organizations. However, feedback from funders was that applicants needed more support than anticipated in completing the new requirements. In one case this was delivered by an independent evaluation consultant and having repeated this approach for 3 years they say there are signs of growing confidence from those who they have funded before. But it remains a challenge for those who are new to funding. Likewise, while it appeared applicants embraced the nuanced reflection on what success and failure would look like at the beginning of the process, there was a tendency to still highlight successes at the end.

While we could make the case for some limited impact on adoption by funders, we have failed to deliver something that is easy to adopt, lessening the likelihood that other funders will do so unaided. Indeed, while our partners have said they are keen to continue, recognizing that 'slow is osmosis' (Funding partner) they have failed to persuade their wider organization to adopt FailSpace as their 'preferred evaluation framework' (Funding partner).

6. Reflecting on failures

Reflecting on the findings above, what it demonstrates is a complex picture in which we may recognize successes and failures in all aspects of our work to generate impact from our work.

	Outright Failure	Precarious Failure	Tolerable Failure	Conflicted Success	Resilient Success	Outright Success
Profile –the reputation and the ability to promote	We fail to raise interest in our impact plans	We engage interest but fail to convince	**The programme has visibility but its impact is less clear**	The impact is talked about as both a success and failure	***Key stakeholders acknowledge that our work has positive impacts***	We are recognised as affecting sustainable change
Purpose - the attainment of stated aims, objectives and outcomes	We fail to get people talking about failure	**The value is acknowledged but there is no evidence people are doing it**	There is a short-term increase in people talking about failure with colleagues	***More people value talking about failure but there is no evidence it is changing practice***	More people value talking about failure and emerging evidence it is changing practice	Talking about failure has become normalised and is changing practice
Process - the actions, activities and stages of delivery	We fail to identify partners to work with us on impact	We identify partners but struggle to work together	**Our partners only create limited opportunities to talk about our approach**	***Our partners test our approach but don't share their learning***	Or partners test our approach and share their learning	Or partners test and share learning, leading to wider adoption
Practice – impact on academic practice	We do not gain any new knowledge about generating impact from research	**We have new knowledge but feel less confident in our abilities to generate impact than before**	We gain new knowledge about creating impact but do not prioritise it in future	We gain new knowledge and use this to create impact but do not share it	***We gain new knowledge which we use and share***	Our new knowledge on impact changes how others do impact work
Participation – level of agency partners have in shaping or evaluating the programme	**We fail to make partners feel any ownership of our framework**	Partners exhibit some ownership but we fail to adapt in response to feedback	***Partners exhibit some ownership but don't show any desire to develop it for their own needs***	Partners exhibit a good degree of ownership but have adapted in ways we are uncomfortable with	Partners exhibit ownership and have worked with us to embed the framework	Partners have full ownership and are confident without our continued

Figure 3. Researchers review of degrees of Failure/Success for the FailSpace project at the end of the Follow-On Funding for Impact Generation in italics (bold text was the assessment of the researchers at the end of the original research project).

But rather than disguising this complexity through evaluating only positive impacts we argue that it is nuancing the success and failures that allows us to generate an understanding of the true nature of our impact, and learn some of the reasons for those failures.

We therefore return to the grid we shared in our methodology (Fig. 2), which highlighted our successes and failures in generating impact at the end of our research phase in bold, to consider what has changed as a result of the impact activities undertaken with follow on funding. These changes are shown in the text in italics in Fig. 3.

In terms of the research profile we could argue that it is now a resilient success, as many stakeholders have acknowledged the importance of our work and its impact on their thinking. This would clearly provide the basis of a strong narrative for an impact case study, especially if those stakeholders were prepared to share examples of how this has led to behavioural change. Yet it may equally be a precarious failure as we have failed to convince most funders and cultural organizations to adopt our approach as standard practice. Acknowledging this is more likely to lead us to consider what more we could do, and what resources might be needed, if we, or others, wished to build the longer-term impact of the FailSpace project.

Reflecting on the purpose of our work, we have outlined the uncertainty of both ourselves and our collaborators about whether any increased discussion about failure within the cultural sector can be attributed to us or, even if it is, whether this will be sustained. Even where talking about failure may have been normalized we have less evidence that it is changing practice. Our failure to agree how much we have achieved our purpose on our own grid demonstrates the problem with measurements of impact based on narrative examples, that may not reflect a wider picture.

In terms of process and participation, it is easy to see that more funding has allowed for more partnerships, and this has provided some opportunities to share learning. But like current input/outcome approaches to measuring impact, we question whether on its own this tells us anything other than the value of funding to allow such activities to take place. Our critical reflections have shown that, despite any successes here, the process of embedding our approach needed more work than we were funded to deliver, resulting in the sector relying on us as researchers to facilitate their discussion about failure, rather than taking ownership of the methodology that underpins our work or the tools that we provided. Indeed, the very successes of our tools in regard to their profile may add to the failures of embedding practice, as for many, completing the tools at one of our events became an end in itself rather than part of a process for learning and change.

Finally, in terms of practice, we learnt much through the failures we had regarding achieving all we had committed to when applying for our follow-on funding from the AHRC. Yet despite this we have done little to share this learning with others in the HE sector. This article is a direct result of that reflection, and in publishing our failures, we feel we have moved our project from a conflicted to resilient success in this area.

7. Conclusion

We wrote this article to ensure we were modelling our own framework of failure when reflecting on the impact of our research. Opting to do so is not without its risks as our

institutions decide whether to include our work as part of an impact case study for REF2029 and thus ask us to provide a strong narrative, supported by demonstrable evidence, as to the positive impact our research has had. As it stands, the template for REF impact case studies does not provide space to acknowledge failures alongside successes.

We have demonstrated how academia, like the cultural sector, which was the focus of our original research, shares a tendency towards overpromising what can be achieved with relatively small pots of money supporting time-limited projects. Firstly, this encourages research on those things that are easy to change and not those things which need to be changed. Secondly, it encourages a tendency to 'prove' the existence of impact rather than critically reflect on the extent of the impact and the failures which are likely to co-exist alongside success. While it is important to know 'what works', it is perhaps more important to know what doesn't work, in order to learn from this and adapt future activities to mitigate against the factors that caused failures previously. We also argue that being prepared to acknowledge how failures have limited the nature and/or scope of research impact adds credibility to claims made about successes, rather than undermining them, and should be seen as a core component of research integrity. Furthermore, greater openness about research and impact-related failures would arguably build confidence amongst researchers, in a sector rife with imposter syndrome that may, in part, be exacerbated by only reading about successes.

This article demonstrates that many of the problems with evaluation in the cultural sector are applicable to the impact agenda in HE. In using our framework to reflect on our own impact failures we have shown that our methodology and approach can therefore be transferred from cultural policy to academia. As such, we end with a call to all those engaged in impact activities or measurement to employ the same principles of identifying, acknowledging and learning from failures. Doing so will aid social learning (May 1992) within the HE sector about the degree to which any given time-limited research can contribute to changes that benefit society, and the difficulties in doing so. Specifically, we call on UKRI and the REF2029 team to make one small change now and alter their impact reporting mechanisms so that it does not focus only on the successful impacts. Instead, case study authors should also be asked to reflect on the impact failures of their work, for in doing so we will collectively gather insights into the contingent nature of generating benefit and public good through academic research.

Funding

This work was supported by the Arts and Humanities Research Council [grant numbers AH/R008760/1, AH/W002337/1].

Notes

1. Wicked problems were proposed by Horst Rittel and Melville Webber in (1973) to contrast with 'tame' problems that could be solved using standard scientific techniques. The term is now commonly used in a policy context to describe problems that may not be completely resolvable because they involve competing values and interests, resulting in the need for difficult choices that cannot equally respond to opposing ideas.
2. Reed *et al.* (2021: 4–5) provide a summary of national research impact assessments around the world which highlights a number of countries (e.g. Sweden, Norway, Hong Kong) that emulate the UK's REF methodology.
3. For a fuller discussion of the literature on policy failures, and the degree to which failure has been researched and understood in different disciplines, see Chapter 3 of Jancovich and Stevenson (2023).

References

Adams, T. E., Holman Jones, S., and Ellis, C. (Eds.). (2021) *Handbook of Autoethnography* (2nd ed.). Routledge.

Andrews, M. (2018) Public Policy Failure: 'How Often?' and 'What is Failure, Anyway'? Working Papers. https://doi.org/10.1787/gov_glance-2013-en

Bandola-Gill, J. (2019) 'Between Relevance and Excellence? Research Impact Agenda and the Production of Policy Knowledge', *Science and Public Policy*, 46: 895–905.

Bayley, J. (2023) *Creating Meaningful Impact: The Essential Guide to Developing an Impact-Literate Mindset*. Bingley: Emerald Publishing Limited.

Bignell, V., and Fortune, J. (1984) *Understanding Systems Failure*. Manchester: Manchester University Press (Open University Set Book).

Bornmann, L., and Marx, W. (2014) 'How Should the Societal Impact of Research be Generated and Measured? A Proposal for a Simple and Practicable Approach to Allow Interdisciplinary Comparisons', *Scientometrics*, 98: 211–9.

Boud, D., Keogh, R., and Walker, D. (1985) *Reflection: Turning Experience into Learning*. London: Routledge.

Caffrey, T. (2023) Why We Need an Imperfect Impact Agenda. *LSE Impact Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2023/10/20/why-we-need-an-imperfect-impact-agenda/, accessed 13 Nov. 2023.

Chang, H. (2008) *Autoethnography as Method*. Walnut Creek, CA: Left Coast Press.

Chubb, J., and Reed, M. S. (2018) 'The Politics of Research Impact: Academic Perceptions of the Implications for Research Funding, Motivation and Quality', *British Politics*, 13: 295–311. https://doi.org/10.1057/s41293-018-0077-9

Chubb, J., and Watermeyer, R. (2017) 'Artifice or Integrity in the Marketization of Research Impact? Investigating the Moral Economy of (Pathways to) Impact Statements within Research Funding Proposals in the UK and Australia', *Studies in Higher Education*, 42: 2360–72. https://doi.org/10.1080/03075079.2016.1144182

de Jong, S. et al. (2014) 'Understanding Societal Impact through Productive Interactions: ICT Research as a Case', *Research Evaluation*, 23: 89–102. https://doi.org/10.1093/reseval/rvu001

Derrick, G. (2019) 'Cultural Impact of the Impact Agenda: Implications for Social Sciences and Humanities (SSH) Research', in D. Bueno et al. (eds) *Higher Education in the World 7, Humanities and Higher Education: Synergies between Science, Technology and Humanities*, pp.385–392. Barcelona: Global University Network for Innovation.

Derrick, G., and Benneworth, P. (2019) 'Grimpact: Time to Acknowledge the Dark Side of the Impact Agenda', *LSE Impact Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2019/05/28/grimpact-time-to-acknowledge-the-dark-side-of-the-impact-agenda/, accessed 16 Nov. 2023.

Derrick, G. E. et al. (2018) *Towards Characterising Negative Impact: Introducing Grimpact*. Conference paper, STI 2018 Leiden.

Gunn, A., and Mintrom, M. (2017) 'Evaluating the Non-Academic Impact of Academic Research: design Considerations', *Journal of Higher Education Policy and Management*, 39: 20–30.

Hammersley, M. (2014) 'The Perils of 'Impact' for Academic Social Science', *Contemporary Social Science*, 9: 345–55.

Hogwood, B. W., and Gunn, L. A. (1984) 'Chapter 2: Analysing Public Policy', in B. Hogwood and L.A. Gunn (eds) *Policy Analysis for the Real World*, pp.12–31. Oxford: Oxford University Press.

Howlett, M. (2012) 'The Lessons of Failure: Learning and Blame Avoidance in Public Policymaking', *International Political Science Review*, 33: 539–55. https://doi.org/10.1177/0192512112453603

Howlett, M., Ramesh, M., and Wu, X. (2015) 'Understanding the Persistence of Policy Failures: The Role of Politics, Governance and Uncertainty', *Public Policy and Administration*, 30: 209–20. https://doi.org/10.1177/0952076715593139

Izzi, V., and Murray, B. (2019) 'Do No Harm—What Development Practice Can Teach Us About Negative Impact', *LSE Impact Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2019/07/11/do-no-harm-what-development-practice-can-teach-us-about-negative-impact/, accessed 28 Nov. 2023.

Jancovich, L., and Stevenson, D. (2023) *Failures in Cultural Participation*. Cham: Palgrave Macmillan.

Kidd, I. J., Chubb, J., and Forstenzer, J. (2021) 'Epistemic Corruption and the Research Impact Agenda', *Theory and Research in Education*, 19: 148–67. https://doi.org/10.1177/14778785211029516

King, A., and Crewe, I. (2013) *The Blunders of Our Governments*. London: Oneworld Publications.

Lauronen, J. P. (2020) 'The Dilemmas and Uncertainties in Assessing the Societal Impact of Research', *Science and Public Policy*, 47: 207–18. https://doi.org/10.1093/scipol/scz059

Lonsdale, C., and Seruagi, G. (2023) 'Five Lessons to Deal With Uncertainty in Research Impact Evaluation', *LSE Impact Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2023/10/03/five-lessons-to-deal-with-uncertainty-in-research-impact-evaluation/, accessed 6 Nov. 2023.

McCowan, T. (2018) 'Five Perils of the Impact Agenda in Higher Education', *London Review of Education*, 16: 279–95. https://doi.org/10.18546/LRE.16.2.08

Matson, J. V. (1991) *How to Fail Successfully: A Bold Approach to Meting Your Goals through Intelligent Fast Failure*. Dynamo Publishers.

May, P. J. (1992) 'Policy Learning and Failure', *Journal of Public Policy*, 12: 331–54. https://doi.org/10.2307/4007550

Mezirow, J. (1991) *Transformative dimensions of adult learning*. San Francisco: Jossey-Bass.

Penfield, T. et al. (2014) 'Assessment, Evaluations, and Definitions of Research Impact: A Review', *Research Evaluation*, 23: 21–32.

Reed, M. S. (2016) *The Research Impact Handbook*. Huntley: Fast Track Impact.

Reed, M. S. et al. (2021) 'Evaluating Impact from Research: A Methodological Framework', *Research Policy*, 50: 104147. https://doi.org/10.1016/j.respol.2020.104147

Reed, M. S. et al. (2025) 'Analyzing Who is Relevant to Engage in Environmental Decision-Making Processes by Interests, Influence and Impact: The 3i Framework', *Journal of Environmental Management*, 373: 123437.

Research Excellence Framework/REF (2025) *Guidance—Section 1 Overview*. https://2029.ref.ac.uk/guidance/section-1-overview/#section-engagement-and-impact-e-amp-i-, accessed 16 May 2025.

Research Excellence Framework/REF (2019), *Guidance on Submissions*. https://www.ref.ac.uk/media/1447/ref-2019_01-guidance-on-submissions.pdf, accessed 21 Jan. 2024.

Rittel, H. W., and Webber, M. M. (1973) 'Dilemmas in a General Theory of Planning', *Policy Sciences*, 4: 155–69.

Sanderson, I. (2002) 'Evaluation, Policy Learning and Evidence-Based Policy Making', *Public Administration*, 80: 1–22. https://doi.org/10.1111/1467-9299.00292

Smith, K. E. et al. (2020) *The Impact Agenda: Controversies, Consequences and Challenges*. Bristol: Policy Press.

Sousa, B. J., and Clark, A. M. (2019) 'The Ubiquity and Invisibility of Research Failures: A Call to Share More', *International Journal of Qualitative Methods*, 18: 1–3. https://doi.org/10.1177/1609406919846977

Spaapen, J., and van Drooge, L. (2011) 'Introducing 'Productive Interactions' in Social Impact Assessment', *Research Evaluation*, 20: 211–8.

Taylor, D. and Balloch, S. (eds) (2005) *The Politics of Evaluation: Participation and Policy Implementation*. Oxford: Oxford University Press.

UKRI (2023) *Defining Impact*. https://www.ukri.org/councils/esrc/impact-toolkit-for-economic-and-social-sciences/defining-impact/, accessed 13 Nov. 2023.

Viana-Lora, A., and Nel-lo-Andreu, M. G. (2021) 'Approaching the Social Impact of Research Through a Literature Review', *International Journal of Qualitative Methods*, 20: 1–11.

Wilsden, J. (2020) 'A Requiem for Impact?', *WONKE* https://wonkhe.com/blogs/a-requiem-for-impact/, accessed 5 Sep. 2024.