



Evaluating impact from research: A methodological framework

M.S. Reed^{a,g,*}, M. Ferré^{b,f}, J. Martin-Ortega^b, R. Blanche^c, R. Lawford-Rolfe^d, M. Dallimer^b, J. Holden^e

^a Thriving Natural Capital Challenge Centre, Department of Rural Economies, Environment & Society, Scotland's Rural College (SRUC), Peter Wilson Building, Kings Buildings, West Mains Road, Edinburgh EH9 3JG

^b School of Earth & Environment, University of Leeds, Leeds, LS2 9JT, United Kingdom

^c Division of Media, Communication and Performing Arts, School of Arts, Social Sciences and Management, Queen Margaret University, Queen Margaret University Way, Musselburgh EH21 6UU, United Kingdom

^d School of Earth & Environment, University of Leeds, Leeds, LS2 9JT, United Kingdom

^e School of Geography, University of Leeds, Leeds, LS2 9JT, United Kingdom

^f Agricultural Research Centre for International Development (CIRAD), 42 Rue Scheffer, 75116 Paris, France

^g Centre for Rural Economy and Institute for Agri-Food Research and Innovation, School of Natural and Environmental Sciences, Newcastle University, Agriculture Building, Newcastle upon Tyne NE1 7RU, United Kingdom

A B S T R A C T

Background: Interest in impact evaluation has grown rapidly as research funders increasingly demand evidence that their investments lead to public benefits.

Aims: This paper analyses literature to provide a new definition of research impact and impact evaluation, develops a typology of research impact evaluation designs, and proposes a methodological framework to guide evaluations of the significance and reach of impact that can be attributed to research.

Method: An adapted Grounded Theory Analysis of research impact evaluation frameworks drawn from cross-disciplinary peer-reviewed and grey literature.

Results: Recognizing the subjective nature of impacts as they are perceived by different groups in different times, places and cultures, we define research impact evaluation as the process of assessing the significance and reach of both positive and negative effects of research.

Five types of impact evaluation design are identified encompassing a range of evaluation methods and approaches: i) experimental and statistical methods; ii) textual, oral and arts-based methods; iii) systems analysis methods; iv) indicator-based approaches; and v) evidence synthesis approaches.

Our guidance enables impact evaluation design to be tailored to the aims and context of the evaluation, for example choosing a design to establish a body of research as a necessary (e.g. a significant contributing factor amongst many) or sufficient (e.g. sole, direct) cause of impact, and choosing the most appropriate evaluation design for the type of impact being evaluated.

Conclusion: Using the proposed definitions, typology and methodological framework, researchers, funders and other stakeholders working across multiple disciplines can select a suitable evaluation design and methods to evidence the impact of research from any discipline.

1. Introduction

Interest is growing rapidly in the evaluation of non-academic benefits or “impacts” (see Section 3 for definition) arising from research, as funders and Governments around the world increasingly seek evidence of the value of their research investments to society (Edler et al., 2012; Oancea, 2019). The growth of research over the past few decades has outstripped available public funding in many countries, leading to discussions about how to get best value from research, particularly basic research which may not have immediate application (Boreman, 2012). The Global Financial Crisis of 2007/8, further intensified discussions about how to measure the quality of research and how to evaluate its societal value, to provide public research funding agencies with

evidence to justify budgetary requests to governments. The drive to evaluate the societal impact of research is exemplified by the assessment of non-academic impact by the UK’s Research Excellence Framework in 2014 and 2021 (REF; the system for assessing the quality of research in UK higher education institutions), and the growing trend to evaluate research impact at national scales around the world (Box 1).

In this paper, we refer to evaluation as the process of collecting and interpreting data to assess the significance, reach and attribution of impacts from research. We refer to evidence as the communication or “demonstration” of impact based on robust evaluation. However, defining the benefits of research is a highly subjective process, and a benefit for one group in one place, time and culture, may be perceived as damaging the interests of others (e.g. other groups, future generations or

* Corresponding author.

E-mail address: mark.reed@sruc.ac.uk (M.S. Reed).

the environment). The diversity of benefits and perceptions of benefits arising from research presents a major methodological challenge for evaluating and evidencing impact claims (as an illustration, 3709 unique impact pathways were identified from the 6679 case studies submitted to REF2014; [Grant, 2015](#)). In the face of such diversity, there can be no single process or checklist for evaluating and evidencing impact. Rather, methods need to be adapted to the unique impacts, pathways and contexts associated with research on a case-by-case basis.

There is no shortage of methods for evaluating research impact ([Alla et al., 2017](#); [Reed, 2018](#)). The challenge therefore lies in choosing the most appropriate methods in an evaluation design that is suited to a given impact and context. Guidance from the realms of evidence-based policy/practice and research-informed international development typically follows a hierarchy of methods, based implicitly on their assumed accuracy and minimization of bias (e.g. [Gertler et al., 2011](#); [HM Treasury, 2011](#); [USAID, 2011](#)). Randomised controlled trials sit at the top of this notional hierarchy, followed by quasi-experiments, mixed methods and qualitative methods. Implicit in this hierarchy is the idea that quantitative measures are superior to qualitative approaches. This hierarchy may be valid in the evaluation of some types of impact in certain contexts, for example where it is possible to isolate and evidence the sole cause (e.g. an intervention based on research) of any given effect (the impact).

However, it is increasingly clear that the relationship between research and societal impact is far more indirect, non-linear, and complex than many evaluation frameworks allow ([Bornmann, 2012](#); [UNEG, 2013](#)). Indeed, it is rare for an impact in any domain to be solely attributable to a single research project or output. More commonly, impacts arise from a body of knowledge that may include hundreds or even thousands of strands of research, some of which may stretch back several decades ([Morris et al., 2011](#)). Moreover, effects from research are often mediated by many other enabling factors (e.g. new incentives, economic volatility or changing attitudes) without which the impacts would not have been possible. Furthermore, pathways to impact (the knowledge exchange or engagement activities that facilitate impacts; [UKRI, 2018](#)), are often littered with unintended positive or negative consequences ([Alvarez et al., 2010](#)), time lags ([Morris et al., 2011](#); [Sanjari et al., 2014](#)), lack of researcher control over the implementation of recommendations ([Rau et al., 2018](#)), ethical challenges ([Sanjari et al., 2014](#)), spillover effects and knowledge creep ([Penfield et al., 2014](#)) and that makes evaluation difficult. Even when these factors are taken into account, few evaluations of research impact draw on the latest literature or are aware of the full range of evaluation options available ([Stem et al., 2005](#)).

As a result, many evaluations of research impact are not able to capture the multifaceted, complex and long-term benefits arising from research, and so can lack credibility and potentially offer few lessons to enhance future practice in research or impact domains ([Cartwright and Hardie, 2021](#); [Woolcock, 2013](#)). In response to these challenges, there have been calls for research impact evaluation to draw on mixed methods approaches ([Gaunand et al., 2015](#)), triangulating evidence from multiple sources to demonstrate rigour ([Reed, 2018](#)).

Evaluating and evidencing impact is harder for some research disciplines than others. The impact agenda aligns well with the norms and practices of some (especially more applied) disciplines and the intrinsic motivations of certain researchers, legitimising their investment of time and energy in the pursuit of impact ([Watermeyer, 2019](#)). However, there is evidence that other researchers (especially from arts, humanities and pure science disciplines), whose work may have no obvious or concrete application or immediate/obvious public interest, are concerned by expectations that their work should generate impact, and feel that their academic freedom is under threat from the increasing evaluation (and especially metricisation) of impact ([Chubb et al., 2017](#); [Bulaitis, 2017](#); [Chubb and Reed, 2018](#)). With this in mind, it is important to emphasise that rather than legitimizing a narrowing and instrumentalization of impact through evaluation, we seek to provide a holistic and adaptive

framework within which to think critically about a diverse range of impacts from research from any discipline.

In this paper we attempt to tackle some of the key challenges of evaluating and evidencing impacts arising from research. We do so by proposing a comprehensive research impact evaluation typology and methodological framework, based on an analysis of evaluation frameworks from multiple disciplines. Methodological frameworks currently available are not well adapted for application beyond the disciplines within which they were originally developed. By comparing impact evaluation frameworks from different research fields, we hope to enable researchers, funders and other stakeholders, to easily select (and where relevant integrate) the most appropriate methods for evaluating and evidencing the impact of research. Our analysis makes a theoretical contribution by providing new and universally applicable definitions of research impact and impact evaluation in a field that is dominated by discipline-specific and technocratic definitions. We make a methodological contribution by proposing the first typology of research impact evaluation designs, which we use as the basis for a wider methodological framework to guide rigorous impact evaluations in any discipline.

2. Definitions: What is research impact evaluation?

2.1. What is research impact?

A number of definitions of research impact have been developed, primarily in technical documents guiding research assessments (e.g. [Australian Research Council, 2017](#); [Research England, 2019](#)) or within narrow disciplinary contexts (e.g. [Halse and Mowbray, 2011](#); [Neiderman et al., 2015](#); [Alla et al., 2017](#)). [Alla et al. \(2017\)](#) reviewed 108 research impact definitions, noting the tendency to discuss rather than define impact, and called for greater conceptual clarity on impact (their definition was tailored specifically for use in health policy contexts). There are problems with many of the existing definitions of research impact. For example, they tend to restrict their focus to certain types of beneficiary leading to the exclusion of others (e.g. [Research England's \(2019\)](#) anthropocentric focus on "economy, society and/or culture" to the apparent exclusion of environmental impacts, non-human beneficiaries and future generations). They also typically combine definitions of impact with typologies, listing examples of types of impact as (part of) their definition (e.g. [Nutley et al. \(2007\)](#) and [Morton \(2015\)](#) define impact as changes in: "awareness, knowledge and understanding; ideas, attitudes and perceptions; and policy and practice as a result of research"). Temporal dimensions of impact are rarely considered; as [Brewer \(2011, p.256\)](#) noted, impact "varies over time and can change, positively or negatively, at the one-point snapshot whenever it is measured". It is also worth considering how the significance of past events can be revised as contexts change and the importance of an event becomes clearer, and hence evaluations of impact may always have to be considered provisional e.g. insights from the philosophy of history suggest that views of the significance of past events change repeatedly based future events, and hence historical significances can never be fixed once and for all ([Danto, 1962](#)).

The most widely used definitions rarely explicitly recognise the subjectivity associated with determining who benefits from research and how, and the extent to which research can be shown to have made a necessary or sufficient contribution towards the benefit. Impact is in the eye of the beholder; a benefit perceived by one group at one time and place may be perceived as harmful or damaging by another group at the same or another time or place. These value judgements and assumptions are implicit in most definitions of research impact, which are rarely unpacked (the word "impact" could refer to positive or negative effects of research, but the implicit focus is on benefits ([Australian Research Council, 2017](#); [Research England, 2019](#); [Samuel and Derrick, 2015](#)). A researcher aspiring to achieve one impact may discover unexpected alternative benefits or unintended negative consequences. As such, there is a normative assumption underpinning the "impact agenda" that

research should seek positive and not negative impacts. This focus on seeking positive outcomes matches perceptions of impact evaluators who were interviewed by [Samuel and Derrick \(2015\)](#) as part of the REF2014 process, which showed most viewed impact as an “outcome” that they would define as a “change” or “difference” that was conceptualised by some as the “final” outcome and by others as a series of secondary or intermediary outcomes that may ultimately lead to the final outcome. As such, our definition recognises and makes explicit this normative dimension of impact as benefit.

Finally, definitions of research impact rarely consider the nature or level of attribution between research and impact, which can vary considerably. The causal relationship between research and impact can be: i) necessary, implying that a body of research was necessary to generate the impact but could not alone have caused the impact (i.e. the research was a significant contributing factor amongst other causes but was not sufficient alone to generate the impact); or ii) sufficient, implying that a body of research alone was sufficient to generate the impact. A “body of research” could range from a body of evidence within a single project or programme to a body of work by a single researcher or group or a wider body of research by multiple authors and teams on a given topic. We distinguish between necessary and sufficient causation on the basis of literature from philosophy (e.g. [Mackie, 1974](#)), law (e.g. [Greene and Darley, 1998](#); [Braham and Van Hees, 2009](#)), and mathematics (e.g. [Pearl, 1999](#); [Tian and Pearl, 2000](#)), which has been applied in contexts as broad as epidemiology (e.g. [Parascandola and Weed, 2001](#)), genetics ([Moss, 1981](#)) and international development ([Mayne, 2012](#)).

As such, the task of any impact evaluation is to establish whether or not there is a causal relationship between research and impact, providing evidence that the research was necessary (at least) or sufficient (at best). Necessary and sufficient cause can be established in a number of ways. Counterfactual causation is demonstrated by showing that it is plausible that the research led to the impact and that the impact would not have been possible without the research. Additive causation is demonstrated by showing a dynamic relationship between research and impact variables, such that one varies with the other. Generative causation is demonstrated by showing the mechanism or process that causes the research to generate impact. Each of these types of cause and effect relationship may be demonstrated probabilistically, for example using experimental design and statistics, or through triangulation, where multiple sources of evidence are compared to infer a likely relationship ([Pawson, 2013](#)). The extent to which sufficient or necessary causation is required in any evaluation will depend on the context, with high risk or controversial claims typically requiring a higher burden of proof, for example where impact claims (such as the efficacy of a medical treatment) could lead to harm if later disproven. In these contexts, evaluations require significant research investment (for example, commissioning randomised controlled trials).

Building on these considerations, we define research impact as demonstrable and/or perceptible benefits to individuals, groups, organisations and society (including human and non-human entities in the present and future) that are causally linked (necessarily or sufficiently) to research.

2.2. What is research impact evaluation?

Although by definition (see previous section) the impact agenda focuses on benefits, it is clear that there may be a variety of perspectives that may challenge whether or not research led to unquestionably beneficial outcomes. It is therefore essential that the process of impact evaluation looks even-handedly at these different perspectives to provide researchers with formative feedback that can enable them to learn from mistakes, identify and hopefully reduce negative outcomes during the pathway to impact and build capacity for more responsible research and innovation ([Scriven, 1991](#); [Patton, 1996](#); [Joly et al., 2017](#)). If this is not possible, then an impact evaluation needs to represent the diversity

of perspectives on the outcomes of the research, whether positive or negative, based on the same ethics that govern the research process itself.

We therefore define research impact *evaluation* as the process of assessing the significance and reach (defined later in this section) of both positive and negative effects of research. Impact may be evaluated over different time horizons, at different social scales (from individuals to society), spatial scales (from local to international) and across multiple domains (including social, economic, environmental, health and well-being, and cultural). In addition to these ultimate impact domains, there are a range of intermediary domains where impacts can occur, including understanding/awareness, attitudinal change, behaviour change and decision-making, policy and capacity building (based on [Reed's \(2018\)](#) impact typology).

Our approach focuses on evaluating impact: i) on individuals and organisations (including funders) who may be engaging directly with research, who are the object of research, or are being targeted in other ways as beneficiaries of a research project; and ii) those indirectly affected by research. We are interested in how these individuals or organisations learn, think, behave and benefit (or are compromised or harmed) as a result of their engagement with research. As such, evaluation of impact must go beyond the measurement of outcomes to more nuanced assessments of tacit and implicit effects of research that may need to be accessed indirectly and evaluated in qualitative terms. Based on the definition of impact above, it is clear that impact evaluation is not only concerned with identifying ultimate, end-of-pipe impacts (e.g. economic or health and wellbeing benefits), but also the range of intermediate impacts that occur on the pathway to impact (e.g. understanding/awareness, behaviour change and policy).

Significance and reach are the two most commonly used criteria to assess impact from research (as used, for example, in the UK's Research Excellence Framework). The significance of an impact can be defined as the magnitude, or intensity of the effect of research on individuals, groups or organisations (after [Alvarez et al. \(2010\)](#) and [Research England \(2019\)](#)). The reach of an impact can be defined as the number, extent or diversity of individuals, groups or organisations that benefit from research (after [Douthwaite et al. \(2003\)](#) and [Research England \(2019\)](#)). Reach can be understood in two ways. First, scaling-out refers to an impact spreading socially (from one individual, community, organisation or interest group to another) and/or spatially (e.g. from the farm to the catchment level, or from one state or country to another). Second, scaling-up and scaling-down refer to an impact reaching a higher or lower institutional or governance level, for example, from influencing individual behaviour change and changing policy mechanisms (e.g. regulation) to influencing the policy frameworks within which those mechanisms sit. Alternatively, scaling-up could range from changing individual perceptions, to social learning (where ideas spread through social networks to become situated in Communities of Practice or social units; c.f. [Reed et al., 2010](#)). To take another example, scaling-up could range from informal changes in individual professional practice to changes in codes of conduct, professional guidance or organisational practice. These processes can operate in reverse, where impacts scale-down from higher to lower institutional or governance levels, for example evidence-based policies, operationalised through regulation, may lead to individual behaviour change. These two dimensions of reach are linked in the sense that scaling-up an impact to higher institutional levels increases the probability of more widespread adoption of ideas, practices and other changes that reach new beneficiaries at wider social or spatial scales.

3. Methods

We analysed existing theoretical and methodological frameworks for impact evaluation from a range of fields, using an adapted Grounded Theory Analysis ([Strauss and Corbin, 1997](#)) to develop robust definitions of research impact and impact evaluation and a novel methodological

framework, including a new typology of research impact evaluation designs. To do this, we started by using a narrative review of cross-disciplinary peer-reviewed literature to identify a wide range of evaluation frameworks and methods that could be used to evaluate impact from research. We also considered grey literature from the non-academic realm. Grey literature included documentation capturing the way in which governmental departments and agencies, non-governmental organizations and other organisations evaluate their own impact, and impacts more broadly within their sector, including the evaluation of actual or likely benefits as well as negative impacts (e.g. the assessment of environmental, economic or social impacts of policies as part of the policy appraisal process). Unlike systematic reviews or meta-analyses, a narrative literature review is an expert-based “best-evidence synthesis” of key literature; it does not seek to capture all literature (Baumeister and Leary, 1997). Greenhalgh et al. (2018) argue that such methods may be more appropriate than systematic approaches for reviews that aim to pursue a broad overview via expert synthesis of literature, and where it is harder to identify specific outcome measures, as is the case here.

Given the wide range of frameworks and methods that can be adapted to evaluate impact from almost every discipline, the goal was to generalize across this literature (rather than to provide an exhaustive list of frameworks and methods) to identify a comprehensive list of distinctive types of impact evaluation. We sought to illustrate the breadth of methods available to operationalise each type of evaluation and show how different approaches and methods can be used to evaluate different types of impact. Google Scholar (for peer-reviewed literature and books) and Google (for grey literature) were searched by two co-authors with the keywords “impact”, “evaluation”, “monitoring”, “research”, and “framework”, reading until theoretical saturation was reached in the categories that emerged (see adapted Grounded Theory Analysis approach below). Despite early criticism of the reliability of Google Scholar (Falagas et al., 2008), more recent analyses have shown strong correlations between citation counts in Google Scholar, Web of Science and Scopus, with Google Scholar consistently returning the highest percentage of citations across subject areas (Martin-Martin et al., 2018a), with significant coverage deficiencies in Web of Science and Scopus (Martin-Martin et al., 2018b). Subsequent to this, further searches were performed for arts-based methods, which were under-represented in the search results, using “arts and humanities” and “arts-based methods” in combination with the previous search terms. Following an adapted Grounded Theory Analysis approach (Strauss and Corbin, 1997), open coding of literature was used to identify emergent themes, continuing to read individual texts until theoretical saturation was reached for each theme. Axial coding was then used to organize themes into theoretical constructs that informed the development of the typology and methodological framework for research impact evaluation.

4. A research impact evaluation typology

The methods for evaluating impacts are as numerous and diverse as the research and impacts they seek to evaluate. There is no “gold standard” method, checklist or standard process. Rather than attempting to lay out a prescriptive methodology for impact evaluation, this section reviews different evaluation designs. We distinguish between approaches and methods for evaluation design. Table 1 identifies five different types of evaluation design from the literature, within which a range of methods (e.g. experimental) and approaches (e.g. logic model) are then nested. While the first three types of evaluation design consist of related evaluation methods, the last two consist of related approaches to impact evaluation. These approaches may draw on any of the methods covered in the first three types, but they do so in distinctive ways that provide higher order insights based on a theory-driven or systematic synthesis of insights from those methods. As with any choice of method or approach in research, this will be influenced by the ontology,

epistemology and theoretical perspective of the choice-maker (Moon and Blackman, 2014). For example, experimental and statistical evaluation designs are more likely to arise from a realist (ontology), objective (epistemology) and positivist (theoretical) perspective, whereas textual, oral and arts-based evaluation designs are more likely to arise from a relativist (ontology), subjective (epistemology) and constructivist, interpretivist or post-modern (theoretical) perspective.

Two key theoretical constructs emerged from the analysis of literature, and these are conceptualised in Fig. 1 as two continua along which research impact evaluations can be arranged or categorised:

- Evaluation designs with a **summative** focus on achieving, evidencing and claiming impacts and being accountable (referred to as external evaluation by Richards, 2008) versus a design with a more **formative** focus on ongoing monitoring, learning, adaptation and taking epistemic responsibility for the generation of impact (referred as internal evaluation by Richards (2008)).
- Evaluation designs that provide evidence that a body of research was a **necessary** (e.g. an important contributing factor) or **sufficient** (e.g. sole attribution) cause of impact (see Section 2.1).

Fig. 1 shows how the five different types of impact evaluation design that emerged from the literature (covered in the next section) were categorised in relation to these two continua, leading to the typology. Experimental and statistical methods and evidence synthesis approaches tend to be used in summative mode, and textual, oral and arts-based methods, systems analysis methods and indicator-based approaches are used in either summative or formative mode. There are evaluation designs that can help disentangle the contribution research has made towards an impact as one of a range of different factors (demonstrating that research was “necessary” to cause impact), and designs that are typically used to demonstrate sole, direct attribution between research and impact (demonstrating that research was “sufficient” to cause impact). The position of evaluation designs in Fig. 1 is approximate, and necessarily generalised (given the diversity of methods and approaches that can be used within each evaluation design) to illustrate how the different designs are typically used in practice. As such, Fig. 1 shows how the evaluation designs in the typology are arranged from more formative approaches that establish the contribution research makes as a necessary cause of impact (bottom left) to more summative approaches that establish research as a sufficient cause of impact (top right).

Each type of impact evaluation design takes a different approach to establishing attribution between research (cause) and impact (effect) (see Section 2.1 for a discussion of the different types of causality used to classify evaluation designs in Table 1). Each type gives rise to different forms of evidence, ranging from testimonials and other forms of qualitative evidence to statistical inferences and other forms of quantitative evidence. Some types of evaluation design have distinct epistemological and/or disciplinary roots (e.g. experimental or arts-based methods), but are not restricted to evaluating impacts from this sort of research (e.g. experimental methods could be used to evaluate impacts arising from arts and humanities research, and arts-based methods could be used to evaluate impacts arising from experimental research). The rest of this section reviews each type of impact evaluation in turn, considering some of the key advantages and limitations associated with each.

4.1. Experimental and statistical methods

Experimental and statistical methods for impact evaluation typically provide evidence of research as a sufficient cause of impact. This is often done by inferring counterfactual causation, based on the difference between two otherwise identical cases, one that is manipulated and the other that is controlled giving rise to evidence of cause and effect (see Table 1). Traditionally, experimental and statistical methods have dominated impact evaluation, and in many fields (e.g. in medical trials and many international development programmes) are still considered

Table 1
Typology of research impact evaluation designs.

Type of research impact evaluation	Examples of commonly used methods and approaches	Characteristics	Approach to establishing attribution between research and impact	Examples of type of evidence	Types of impact typically evaluated
Experimental and statistical methods	Statistical modelling, longitudinal analysis, econometrics, difference-in-difference method, double difference method, propensity score matching, instrumental variable, analysis of distributional effects, experimental economics	Typically used in summative mode, <i>ex ante</i> and/or <i>ex-post</i> , to infer the extent to which research is a sufficient cause of impact (often showing sole and/or direct attribution from research to impact)	Counterfactual causation based on the difference between two otherwise identical cases (cases include individuals, sites, environments/contexts), one that is manipulated and the other that is controlled giving rise to evidence of cause and effect. Additive causation may be inferred from correlation between cause (dependant variables) and effect (independent variables) or statistical difference between effect before/after or with/without an intervention (cause), controlling where possible for confounding effects, and quantifying the extent to which effects can be attributed to multiple causes	<ul style="list-style-type: none"> • Improvements in water quality based on improved regulation arising from research • Reduced morbidity and mortality amongst patients receiving new treatment based on research compared to control group • Monetary benefits arisen from a change on asset management practices in financial organisations informed by research • Optimization in the choice of policy instrument to promote a specific land management technique, informed by research • Numbers of companies, employment or new roles in the workforce • Numbers of (or profits from) new commercial products or spin-out companies • Improvements in indicators of social cohesion or social mobility, within a defined perimeter/community • Time, money, ecosystem variables, or lives saved as a result of new evidence-based practices • A significant contribution made by research to the solution of a previously intractable problem • Increase and strengthening of the number of nodes or connections in a social network following a participatory process • Understanding of how a group of actors relate to each other and act • Testimonials or statements from end users (e.g. policy makers) now applying a modelling tool • Testimonials from practitioners explaining how they gained a higher level of capability and capacity handling daily work thanks to a new guidance (improved skills, understanding, and confidence levels) • Improvements in variables that indicate the achievement of goals set by a stakeholder or other social group who co-produced research (e.g. number of community members having acquired a particular skill) • Changes of perception, awareness or attitudes of a social group as a result of engaging with research • Changes in culture, cultural discourse or appreciation and benefit from cultural artifacts and experiences • Change in pre-established indicators set at the start of a 	<ul style="list-style-type: none"> • Economic • Environmental • Social • Health and wellbeing • Policy • Other forms of decision-making and behaviour change
Systems analysis methods	Contribution analysis, knowledge mapping, Social Network Analysis, Bayesian networks, agent-based models, Dynamic System Models, influence diagrams, Participatory Systems Mapping, Bayesian Updating	Can be used in formative or summative mode, usually <i>ex-post</i> or during a pathway to impact	Additive causation based on tracing links between causes and effects along causal chains or pathways to impact	<ul style="list-style-type: none"> • Policy • Other forms of decision-making and behaviour change • Capacity building 	
Textual, oral and arts-based methods	Testimonials, ethnography, participant observation, qualitative comparative analysis, linkage and exchange model, interviews and focus groups, opinion polls and surveys, other textual analysis e.g. of focus group and interview data, participatory monitoring and evaluation, empowerment evaluation, action research and associated methods, aesthetics, oral history, storytelling, digital cultural mapping, (social) media analysis, poetry and fiction, music and dance, theatre	Used either in formative mode to enable beneficiaries to engage and shape feedback that then enhances impact, or in summative mode, <i>ex-post</i> , to assess the extent to which research contributed to impact.	Causation is inferred by building a case (sometimes generative and sometimes jointly with beneficiaries) that triangulates multiple sources of evidence to create an evidence-based, credible argument for research being a necessary cause of impact.	<ul style="list-style-type: none"> • All types 	
Indicator-based approaches	Theory of Change, Logical Framework Analysis, Payback	Indicators-based approaches use indicators to assess	Generative causation, identifying causal processes in		All types

(continued on next page)

Table 1 (continued)

Type of research impact evaluation	Examples of commonly used methods and approaches	Characteristics	Approach to establishing attribution between research and impact	Examples of type of evidence	Types of impact typically evaluated
Framework, other logic models, SIAMPI, DPSIR	progress towards anticipated impacts. Any method may then be used evaluate each indicator. These frameworks can be used in summative or formative mode, typically <i>ex ante</i> (but can be used <i>ex-post</i>), to show the extent to which research contributed towards, or was a necessary cause of, impact		chains from the generation of research to the wider impacts in the context of wider supporting or mediating factors and contexts	project that would be expected to show change as impacts occur, for example: <ul style="list-style-type: none">○ Number of farm advisors adapting their discourse to farmers, numbers of farmers taking up the conservation measure, hectares of land restored, followed by reduction in water pollution, savings to water companies and reductions in water bills○ Access to health care, number of individuals getting immunized against a particular disease, number of individuals not contracting the disease, followed by reduction in the predominance of the disease, and savings in health expenditures.	
Evidence synthesis approaches	Meta-analysis, narrative synthesis, realist-based synthesis, rapid evidence synthesis, systematic reviews	Used in summative mode, <i>ex-post</i> , to infer sole attribution or quantify the extent to which research was a sufficient cause of impacts	Causation based on the systematic aggregation and analysis of cause and effect across multiple evaluations (of any type) in different contexts	<ul style="list-style-type: none">• Time, money or lives saved as a result of new evidence-based practices• Example of an actual product, service or policy based on evidence synthesis, with evidence of benefits for those using the product/service or affected by the policy	All types

the “gold standard” (Khandker et al., 2009). This type of evaluation typically compares treatment and control groups (e.g. using a Randomised Control Trial), using statistics to analyse results (e.g. using the difference-in-difference method). Where there are large populations (of observed data), statistical methods can help identify biases and provide quantitative assessments of the likelihood that impacts occurred and are statistically related to a research intervention (Garbarino and Holland, 2009). Attribution between intervention and outcomes often rely on pre-post assessments (i.e. comparison of outcomes before versus after intervention implementation; Dimick and Ryan, 2014). New methods have emerged to cope with time-dependant trends in outcomes that are unrelated to interventions (e.g. the difference-in-difference method uses a comparison group experiencing the same trends that is not exposed to the intervention; Lance et al., 2014).

Experimental and statistical methods may be essential for high risk and/or controversial studies, however they are often costly and time-consuming to implement. As a result, less costly and time-consuming methods have been developed to evaluate impact, for example using quasi experimental designs in which space (a comparable situation or territory without the intervention) is substituted for time. Examples include the comparison-case approach or matching design (e.g. using propensity score matching) (Dickson et al., 2017). Yet, it is often difficult to find a comparable case that represents the alternative state. There are three other weaknesses associated with experimental and statistical impact evaluation (Hewlett et al., 2017). First, the potential to replicate and synthesise studies to provide reliable evidence of what works at national or system levels to inform wider policy and practice is compromised by a lack of common standards for collecting and reporting data (Victora et al., 2011). Second, quantitative, metric-based approaches to impact assessment have been criticized as oversimplifying and so providing partial and/or misleading findings (e.g. Bayley and Phipps, 2017). For example, Australia’s Engagement and Impact Framework (2017) allows higher education institutes to use up to eight quantitative indicators to assess engagement with non-academics, and

two out of the four mandatory indicators are “cash support from research end users” and “research commercialization income”. Economic indicators such as these are a crude proxy for engagement, may or may not be correlated to impacts and favour certain disciplines over others (e.g. engineering over many other sciences, and design over many other arts and humanities disciplines). Third, quantitative approaches can be used to establish correlations that may be mistaken for cause and effect without the use of additional methods to infer causality.

4.2. Systems analysis methods

Evaluation designs based on systems analysis are similar to evaluations based on Theory of Change. However, they are typically used *ex-post* to explore whether research was necessary to cause impact, by disentangling the messy complexity of impacts that occur in complex systems (compared to indicator-based approaches that are more often used in impact planning). They tend to draw on a range of qualitative and quantitative research methods to depict more complex cause-and-effect relationships. They are able to capture the complex range of other factors mediating impacts, to enable the generation of arguments that the research made a significant contribution to the impact, even if direct and sole attribution is not possible.

For example, Reed et al. (2018) used a combination of Social Network Analysis and qualitative interviews to map knowledge flows through science-policy networks to attribute policy impacts to specific research outputs. Research findings were traced as they were communicated between members of the network, identifying which findings got into policy and practice (or not) and how the research findings had been transformed as they were translated for different audiences. Working with another part of the same network, Chapman et al. (2009) used Agent-Based Modelling to understand how target stakeholders were likely to respond to different policy scenarios, to evaluate the social processes through which impacts typically occurred in the study system and guide ongoing impact generation activities (the outcomes of which

Box 1

National research impact assessments around the world.

Europe:

- Horizon Europe has the most advanced programme of impact evaluation that has been seen in any EU framework programme (Directorate-General for Research and Innovation (Directorate-General for Research and Innovation, 2018). Across Europe, Governments are incentivising the generation of impact through conditions attached to research funding and through research evaluations, which increasingly evaluate impact alongside research excellence. For example:
- **UK:** The UK's Research Excellence Framework (REF) incorporated an evaluation of impact using case studies which comprised 20% and 25% of total scores in 2014 and 2021 respectively. Impacts were evaluated in terms of their relative significance and reach, and unlike most other impact assessments elsewhere in Europe, Government funding to Higher Education Institutes was then linked to the outcome. A recent survey suggested 57% UK researchers held negative attitudes towards REF2021 (compared to 29% positive) (Weinstein et al., 2019)
 - **Netherlands:** Since 2015, Dutch Universities have had to submit 3–5 page impact narratives for each of their research units as part of their six-yearly Standard Evaluation Protocol (VSNU/KNAW/NOW, 2014)
 - **Sweden:** Since 2019, Swedish Research Council Strategic Research Centres have to submit impact case studies for evaluation (based on a template derived from the UK's REF)
 - **Italy:** Italy's Research Quality Evaluation (VQR) evaluates technology transfer activities in Italian Universities and Research Bodies (Rebora and Turri, 2013; Geuna and Piolatto, 2016) bib122
 - **Spain:** From 2019, the Spanish National Commission on the Evaluation of Research Performance (CNAI) has provided monetary incentives to researchers who submit "evidence of impact and influence" of their research "on social and economic matters" as part of their six-yearly individual research performance review (Spanish Government, 2018)
 - **Norway:** Norway's Humeval exercise (2015–2017) assesses research at the unit of research groups, and social sciences and environmental research institutes are expected to submit impact case studies based on the UK's REF model (Wróblewska, 2019)
 - **Poland:** Poland will soon follow suit with their own research impact assessment planned for 2020 (Dziennik Ustaw Rzeczypospolitej Polskiej, 2018, 2019; Wróblewska, 2017)
 - **Finland:** in 2019 the Strategic Research Council agreed a set of funding principles mandating impact assessment in all their programmes, requiring funded projects to report impacts (as well as challenges encountered), which are given to external evaluators who assess how well the programme has "solved the challenges facing society and...how efficient this funding instrument is in promoting such research", including the promotion of public debate. Evaluation reports for the first four programmes are expected towards the end of 2020

Rest of the world:

- **Hong Kong:** Hong has broadly replicated the UK's REF methodology in its 2020 Research Assessment Exercise (University Grants Committee, 2017)
- **Australia:** Engagement and Impact Assessment was introduced as part of Excellence in Research for Australia in 2018 (Australian Research Council, 2017)
- **New Zealand:** A 2019 review of the Performance Based Research Fund (PBRF) examined options, costs and benefits of introducing additional impact measures into the PBRF, and the Ministry of Business, Innovation and Employment's 2019 position paper on The Impact of Research called on research funders to articulate line-of-sight to impact in all research funds and contracts and perform impact assessment exercises, collecting impact data to common standards. It called on Universities to support researchers to plan for and generate impact and "work with MBIE towards systems that capture linkable data along the results-chain".
- **USA:** The America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science (COMPETES) Reauthorization Act of 2010 highlighted that a 'broader impacts criterion' was crucial for National Science Foundation funding and encouraged higher education and non-profit organisations to take an institutional approach to achieving societal impacts and determining their accountability (National Science Foundation, 2014; Bozeman and Youtie, 2017). In 2016, the National Science Foundation, National Institute of Health, US Department of Agriculture, and US Environmental Protection Authority developed a data repository for assessing the impact of federal research and development investment (StarMetrics, 2016)

were reported by Reed et al., 2018). Woolcott et al. (2019) built on quantitative measures of social networks to build a methodological framework based on human cultural accumulation theory, and used interviews, questionnaires and focus groups to assess how interpersonal as well as person-environment (including stored knowledge e.g. via books and internet) interactions contributed to the accumulation of memory within individuals and groups, leading to cultural change. As such, in complex systems, they argued that research impact should be

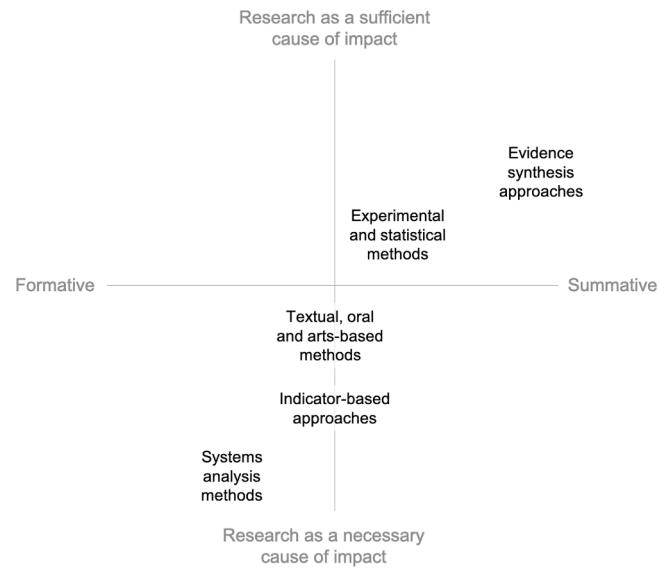


Fig. 1. Five types of impact evaluation designs categorized by the extent to which they provide summative evidence versus formative feedback and the extent to which they provide evidence of research as a sufficient (e.g. sole attribution) or necessary (e.g. a significant contributing factor amongst many) cause of impact.

seen as arising from the "cultural effects of societal interaction", rather than from individual researchers and research outputs, focussing on "research impact as 'our' rather than 'my' impact".

More broadly, systems models can provide detailed understanding of causal links from research to impacts, and are particularly useful for understanding complex, non-linear and unpredictable outcomes. As a family of methods, systems models range from highly quantitative, process-based models, to qualitative conceptual models (referred to variously as mediated modelling, conceptual modelling and participatory systems modelling). At the quantitative end of this spectrum are process-based modelling methods, which can be used to estimate impacts arising from evidence-based interventions in policy and practice. For example, Ewen et al. (2000) developed a spatially distributed process-based model of the full water cycle for integrated land and water management, integrating new techniques for modelling flow and transport of sediments and contaminants, to support decision making at the catchment scale and inform policy related to the environmental impacts of land erosion, pollution, climate change, and land use change within river basins. At the qualitative end of this spectrum, Kenter et al. (2014) used conceptual models to trace the shared social and cultural impacts of new policies based on research, considering environmental, economic and social effects alongside deeper effects on transcendental values and beliefs of affected populations. Sitting in the middle of the spectrum are Dynamic Systems Models, fuzzy cognitive mapping and Bayesian methods, which can integrate both qualitative information (e.g. a relationship between two variables of unknown direction or strength) and quantitative information (e.g. a regression equation). Although more technically challenging, Bayesian methods are particularly useful for quantifying the uncertainty arising from missing information and are able to integrate multiple complex sub-models in addition to qualitative information. By modelling beliefs elicited from relevant experts about likely causal chains between research and impact, Bayesian methods can be used to improve the clarity and precision of likely impacts as part of an *a priori* effectiveness analysis and when integrated with monitoring data assess the relative contribution made by research to impacts (e.g. Befani et al., 2017). Some evaluations, however, are based purely on qualitative data, as the next section shows.

4.3. Textual, oral and arts-based methods

Textual, oral and arts-based evaluation methods tend to build a case that research was necessary to cause impact by triangulating multiple sources of evidence to create a credible, evidence-based argument that attributes impacts to research. All of these methods can be participatory, engaging beneficiaries and other stakeholders in the evaluation itself, enabling these groups to engage and shape the evaluation, which then has the potential to further enhance impact.

Textual and oral methods have a number of key advantages for reflecting impact (Hewlett et al., 2017). Referring to arts and culture case studies in REF2014, Hewlett et al. (2017, p40) commented that, “while reach [of impact] was largely presented as a quantitative measure, a qualitative layer of information about the type of engagement it described also appeared vital. Little distinction can be made between direct and indirect beneficiaries when considering reach in purely statistical terms”. In many research settings, there are multiple lines of evidence (and lines of argument) and other factors contributing towards impact, and it can be difficult to isolate and collect data on all factors, risks, and assumptions. However, qualitative data, for example from interviews/testimonials and focus groups, can help explain and contextualise a project’s results, and create a rounded picture of the likely impacts, considering economic, political, institutional and socio-cultural factors (Dickson et al., 2017). In fact, compared to quantitative methods, qualitative methods lead in some cases to a greater depth of understanding of how and why a research project was or was not effective and how it might be adapted in future to make it more effective (Garbarino and Holland, 2009).

Analysis of textual and oral data, when combined with quantitative work as part of a case study, can furthermore help in the interpretation of quantitative data and relationships, especially in terms of inferring cause and effect. Using a mix of quantitative and qualitative methods in the impact evaluation process can enhance the validity or credibility of evaluation findings, facilitate the development of a method, extend comprehensiveness of evaluation findings, and generate new insights into evaluation findings (Bamberger, 2012). Having said this, criticisms faced by qualitative evaluations of textual and oral data include: the difficulty of generalizing from case-specific findings; the risk of excessive reliance on the opinion and perspective of the evaluator or those providing testimonials; perceived bias arising from small sample sizes where there is insufficient triangulation, and the inability to replicate or validate findings in quantitative terms; and the difficulty of obtaining standardized data allowing us to measure change over time or between groups.

Qualitative Comparative Analysis attempts to overcome some of these limitations by mixing qualitative and quantitative methods in a case-based study (Rihoux and Ragin, 2008). It is particularly useful for disentangling complex relationships where there are multiple causal factors at play. Positive and negative cases of impact to be evaluated (e.g. behaviour change versus offence caused by a public engagement event) are identified and analysed with stakeholders. The group defines a range of likely causal factors (e.g. the research versus a range of other contextual factors) which are analysed using Boolean algebra to assess the combination of causal factors most likely to lead to cases of negative or positive impacts.

Arts-based methods may be used to evaluate impacts arising from any discipline, and should not be seen as only relevant for the evaluation of impacts arising from research in the arts and humanities. Although they derive strongly from an arts and humanities context, we found creative arts methods reported across a very wide range of disciplines within social sciences, healthcare, anthropology, biodiversity and environment settings. The use of arts-based methods in particular has “grown from the desire of researchers to elicit, process and share understandings and experiences that are not readily or fully accessed through more traditional fieldwork approaches” (Greenwood, 2012:2). Research methods used in the arts and humanities aim to provide a

deeper and more nuanced understanding of human experience, meaning and values (Coates et al., 2014). As such, they are able to provide “thick” narratives of impact that highlight lived experience and meaning, and attend to contextual factors (Boydell et al., 2012). Such a constructivist approach towards building up accounts and understanding of beneficiaries’ experiences has distinct value for capturing impact. Furthermore, such approaches to impact evaluation typically infer causation by jointly building a case with beneficiaries that triangulates multiple sources of evidence (including data collected by beneficiaries) to create a credible argument for a significant contribution of the research to impact.

In resisting binary thinking (van der Vaart et al., 2018), arts-based methods have the capacity to capture meaning, implicit and ephemeral phenomena, and benefits that are difficult to express and might therefore pass unrecorded (Hewlett et al., 2017). Methods based on the arts can be particularly useful for researching implicit and tacit impacts that are difficult or impossible to conceptualise or articulate. It is well known that some types of knowledge cannot easily be conveyed through language, such as emotional, aesthetic and symbolic aspects of experience (Fraser and al Sayah, 2011; Dunn and Mellor, 2017). In these cases, arts-based research methods can add value where more traditional tools such as interviews or questionnaires fail to articulate impacts. This is particularly important when working with (often vulnerable) populations with limited verbal or written competence (van der Vaart et al., 2018); arts-based methods enable “better access to the emotional, affective, and embodied realms of life, cultivate empathy, and challenge and provoke audiences to engage with complex and difficult social issues” (Chamberlain et al., 2018).

Visual arts methods commonly used in impact evaluation include photo elicitation (Harper, 2002; also known as photo voice (Wang et al., 1998) and photo survey (Moore et al., 2008)), drawing (e.g. rich pictures from soft systems methodology; Checkland, 2000), paintings (e.g. Gillies et al., 2015) and collages (e.g. Gerstenblatt, 2013). Music, theatre and dance may be used in participatory monitoring and evaluation, for example in ethnotheatre evaluation data are translated into a play script, which is performed, offering potential for further debate and insight (Chamberlain et al., 2018). Fiction writing may be used as a method of enquiry and analysis. For example, Sundin et al. (2018) used storytelling to increase stakeholder engagement in environmental evidence synthesis (see next section) and Kenter et al. (2014) used storytelling to elicit implicit knowledge about the values people held for the natural environment in research that sought to understand the social impacts of policy.

The participatory nature of many textual, oral and arts-based evaluation methods means that people are engaged with research through an action-reflection cycle, enabling new understandings of the phenomena under study to come to light (Fraser and al Sayah, 2011), often challenging perceptions and providing fresh perspectives (Daykin et al., 2017). These methods emphasise plural perspectives from a multiplicity of voices (Coemans et al., 2015) and promote “a form of understanding that is derived or evoked through empathetic experience” (van der Vaart et al., 2018 citing Eisner 2008). In addition to understanding impact at new levels, arts-based methods in themselves provide a medium for communicating the findings of an evaluation in a powerful way (Coates et al., 2014) and are often used to support dissemination, making project reporting more engaging, accessible and relevant to those beyond professional practice and academia (Daykin et al., 2017).

Participatory evaluation methods that generate textual and oral data include transect walks (walking interviews) and matrix ranking (Chambers, 2013). Van der Vaart et al. (2018) used creative workshops about place, identity and community resilience to create an exhibition, gaining multifaceted knowledge of factors leading to impacts (van der Vaart et al., 2018). Others have used process tracing: a qualitative causal inference method where participants score and rank the importance of different possible causal factors for a given impact (Dickson et al., 2017). Role playing games are another type of participatory approach that is

often combined with art-based work, and can be used to test, for example, policy impacts arising from research. For example, Garcia et al. (2015) used role-playing games to engage ecosystem users and academics in the co-design of a board-game that represented and simulated socio-ecosystem functioning, in order to address issues regarding decision processes between stakeholders and predict policy impacts on ecosystem management.

Participatory methods that can be borrowed from anthropology and ethnography include sensory ethnography (exploring subjective experiences through interconnected senses (Crossick and Kaszynska, 2016), and 'Spirit of Place' (capturing the intrinsic values of an environment and why and how people connect to it emotionally; Chamberlain et al., 2018). Many of the methods used in the wider 'action research' tradition, seek to challenge and sometimes overturn the typical power dynamics that exist between the evaluator and those being evaluated, empowering supposed beneficiaries to set the questions for the evaluation and interpret the outcomes, rather than acting as passive research subjects to an external evaluator (van der Vaart et al., 2018).

As a way of evaluating impact, textual, oral and arts-based methods offer particular value in: creating new knowledge spaces (Byrne et al., 2016); eliciting new perspectives on a theme or topic (Boydell et al., 2012; Daykin et al., 2017; van der Vaart et al., 2018); overcoming or challenging power imbalances (van der Vaart et al., 2018); facilitating genuine knowledge exchange (Byrne et al., 2018); and eliciting evidence on "sensitive" or "hard-to-verbalise topics" (van der Vaart 2018). In doing so, this type of evaluation can generate unexpected data layers (Greenwood, 2012) and enhance the communication of both research and impact (Douglas and Carless, 2018).

4.4. Indicator-based approaches

Indicator-based approaches identify variables that indicate the achievement of impacts. Indicators may be used prospectively during planning as milestones and targets, and then retrospectively to see if planned impacts were achieved. Indicators may be identified, organised and evaluated in categories (e.g. see SIAMPI and DPSIR frameworks below) or logical structures (e.g. logic models and Theory of Change). Any method may then be used to evaluate each indicator (e.g. economics and interviews are commonly used to evaluate benefits arising from seven stages of the research cycle in the Payback Framework). Similar to systematic reviews (Section 4.5), which analyse evaluations carried out using any method, theory of change and logic models are a type of approach rather than a type of method.

A theory of change explains how, in theory, research might lead to successive impacts, which can each be measured in turn, providing evidence of clear causal chains from research to impact. Logic models provide a common structure in which expected impacts are systematically measured to generate easily comparable case studies. For example, the Payback Framework (Donovan and Hanney, 2011) organises measurement of impact across seven stages and two interfaces that are typically seen in the research cycle. Methods used to evaluate impact across these stages and interfaces differ from project to project, ranging from quantitative economics methods to qualitative interviews. Similarly, the Fast Track Impact Planning Template (Reed et al., 2018) asks for indicators and means of verification to evaluate the success of engagement and progress towards impact, followed by an assessment of risks to engagement and impact. Depending on the indicators identified, impacts may be measured using very different methods in any given application of the logic model.

As a type of impact evaluation, indicator-based approaches should be seen as a way of identifying and ordering relevant methods in an evaluation, rather than as methods in their own right. They trace causal chains from research to impact, based on an anticipated logic or a theory of likely or desirable change. The closer that reality corresponds to what was expected in theory at the outset, the stronger the case for assuming the research contributed to the outcomes (Bamberger, 2012).

Indicator-based approaches may be used to provide evidence that research was either sufficient or necessary to generate impact, but the explicit consideration of risks and assumptions in both approaches make them well suited to evaluating whether the research was a necessary cause of impact in the context of other contributory/confounding factors. Although they tend to be used *ex-ante* to plan for impacts, they can also be used in evaluation to compare actual impacts to those that were planned.

A logic model (also, called logical framework, Julian et al., 1995) or Theory of Change (Stachowiak, 2013) is typically developed at the start of a research project, working back from the ultimate benefits (in the case of a Theory of Change) or working forwards from impact goals (in the case of a logic model). It consists of mapping out the steps that would be necessary to move from the planned research activities, to the generation of research outputs, intermediate outcomes, short-term impacts and the ultimate benefits that are sought (Alvarez et al., 2010). If the links in the causal chain (also referred to as "programme theory") accurately enable the design of the pathway to impacts and reflect the impact delivery process, then it is possible to design an evaluation to look for each of the causal links and measure indicators to infer whether or not the research is making progress towards impact. For example, an evaluation may assess whether or not capacity has been built and awareness raised by the end of the first year of a project, as envisaged in its Theory of Change, by stress testing procedures or services or surveying staff. Alternatively, national statistics may be used to monitor indicators of malnutrition or morbidity in a project designed to enhance the health of a population. A Theory of Change may be used to work out with greater detail and flexibility how the measurable targets and objectives in a logic model might be delivered in a given context (but it is rare for a logic model to be based on a Theory of Change).

Developing a logic model includes an identification of the different beneficiaries or users of the research output(s), assessments of risk (e.g. internal and external factors that may influence the delivery of each outcome along the causal chain) and identification of assumptions behind the causal links that have been inferred [ibid; Funnell and Rogers, 2011; Douthwaite et al., 2011]. The causal chain in a Theory of Change is usually expressed visually using diagrams, whereas logic models tend to be presented as tables (e.g. Logical Framework Analysis or the Fast Track Impact Planning Template; Reed et al., 2018), and both may also be turned into narrative. Theories of Change tend to focus more on the multiple, potentially alternative links that can be made in the causal chain from research to impact, whereas logic models tend to focus more on activity and impact indicators (and their means of verification). Both Theories of Change and logic models may be developed by a project or research, or may be co-developed in collaboration with stakeholders. For example, Participatory Impact Pathways Analysis enables researchers and stakeholders to jointly describe a project's theories of action, develop logic models, create network maps and use them for planning and evaluation (Alvarez et al., 2010).

One advantage to logic model approaches to impact evaluation is their ability to standardise the collection of data in the creation of case studies that are easily comparable. Similar to the Payback Approach (described above), the ASIRPA method (Socio-economic Analysis of Impacts of Public Agronomic Research) is based on standardized case studies that combine three analytical tools: a chronology that underlines the role of specific actors and the context; an impact pathway (there is no chronology in the impact pathway) that describes the productive configuration, the outputs, the intermediary stage and the impacts; and a vector of impacts that scores the intensity of five impact dimensions (economic, health, political, social and environmental) (Joly et al., 2015; Matt et al., 2017). Public Value Mapping (Bozeman and Sarewitz, 2011) identifies the public value of policies and then tracks the evolution and impacts of policies as they lead to social outcomes.

Contribution analysis also takes a logic model approach, focusing on tracing pathways to impact as a way of assessing the relative contribution of the research to the impact (Morton, 2015). It involves mapping a

pathway to impact, and identifying assumptions and risks for each stage of the pathway. Impact indicators are identified to collect evidence for each element of the pathway, and thus write a ‘contribution story’ that considers various alternative explanations.

The Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions project (SIAMPI) developed an approach to contribution analysis that acknowledged the complexity of attribution between research activities and observed impacts. It focused specifically on reflecting the ‘productive interactions’ between actors, such as the researcher-stakeholder interaction where knowledge is produced and valued that is both scientifically robust and socially relevant (Sanjari et al., 2014; Spaapen and van Drooge, 2011). The Driver-Pressure-State-Impact-Response (DPSIR) framework identifies and monitors indicators within these five categories that are causally linked (OECD, 2001). In this framework, impacts are generally negative outcomes, and so in impact evaluation, the focus is on the effectiveness of the response to the negative impact.

Both Theories of Change and logic models typically involve the identification of activity and impact indicators and criteria. Reed et al. (2006) provided a list of attributes for designing indicators for use by researchers and/or stakeholders that combine accuracy and ease of use. Others have adapted SMART indicators from the management world to suggest that impact indicators should be specific (capture the essence of the desired result and able to pick up changes over the time), measurable in either quantitative or qualitative terms; achievable (feasible in terms of equipment, funding, competences and time), relevant (capture what is to be measured accurately and consistently), and timely (able to provide information in a timely manner) (Douthwaite et al., 2003). The design of impact indicators follows two broad methodological paradigms: i) an expert-led and top-down approach whereby indicators are collected rigorously, scrutinized, and assessed often using statistical tools (this top-down approach enables evaluators to present trends and make comparisons, but such evaluations usually fail to engage local communities); and ii) a community-based and bottom-up paradigm that is rooted in an understanding of local context and local perceptions of the environment and society, but that may be difficult to compare to other contexts (Reed et al., 2006, 2008; Richards and Panfil, 2011).

Alternatively, criteria-based approaches evaluate impacts against pre-established, theory-driven criteria, designed to predict or explain why impacts arise (Rau et al., 2018). For example, Mitchell (2019) developed a survey approach in which data from publics and stakeholders is collected to measure outcomes in different categories, rating their usefulness (based on Likert scale answers to questions about instrumental, conceptual and symbolic use) to create a numeric impact index against which different case studies can be compared. A number of others have proposed the “usability” of research as a key evaluation criterion (Kirchhoff et al., 2013; Lemos, 2014), categorising research according to the ways in which it can be used, for example conceptual use, instrumental use and capacity-building (Meagher and Lyall, 2013). Alternatively, based on criteria arising from participatory research with researchers, Mårtensson et al. (2016) proposed that impact should be evaluated in relation to the credibility of the underpinning research, its contribution to society, the extent to which the research can be effectively communicated and the extent to which it conforms to established ethical and research quality standards.

4.5. Evidence synthesis approaches

While each of the preceding methods or approaches can be used as part of a project cycle, evidence synthesis typically takes place at the programme level and draws on bodies of work emerging from multiple projects. Evidence synthesis is especially useful where there is apparently contradictory evidence across a range of studies about the relationship between an intervention arising from research (e.g. a new process or product) and impact (e.g. studies reporting positive, negative or no association with outcomes that are valued as impacts). Evidence

synthesis is a process of carrying out a review of existing data, literature and other forms of evidence with pre-defined methodological approaches, to provide a transparent, rigorous and objective assessment of whether something arising from research is a sufficient cause of impactful outcomes. Its use is now widespread across many sectors of society in which research can be used to influence and inform decision-making (Game et al., 2018).

Efforts to improve the connections between policy decisions and research evidence have resulted in a number of approaches to evidence synthesis (Game et al., 2018), from meta-analysis to different forms of narrative-based synthesis. Many of these can be broadly grouped under the umbrella term of ‘systematic reviews’. The utility of systematic reviews is well established across a broad range of research disciplines (Victora et al., 2011; Game et al., 2018), including the medical and public health sectors (Egger et al., 2003), development and humanitarian interventions (Mallett et al., 2012), and conservation and environmental management (Pullin and Knight, 2001; Sutherland et al., 2004). Systematic reviews locate information from the peer-reviewed and grey literature, critically appraise methodologies and synthesise findings to deliver answers to research/practice/policy questions. Indeed, by engaging stakeholders in the co-development of a search protocol, as is recommended practice, the probability that review outcomes are relevant enough to generate impact is increased. Stakeholder confidence in systematic reviews is enhanced by the fact that they follow a transparent and repeatable protocol, and give an extensive account of the available evidence. This approach minimises the incorporation of bias into the review. For example, a conventional review may reflect the author(s)’ own opinions and can be based on a selection of literature that is in itself potentially biased.

The methods for reviewing the literature, and for the subsequent synthesis of evidence, under the broad family of systematic reviews, can be very varied. One of the critiques of a full systematic review is that it is time and labour intensive as it requires considerable consultation with likely end-users and searching of unpublished and grey literature, often by hand and often at geographically disparate locations. Further criticisms include that the traditional format of a systematic review (and the meta-analysis that is subsequently carried out on the data) is that it is “mechanistic, driven more by concerns about reliability and replicability than about adding to understanding of phenomena of interest” (Slavin, 1995). As response to those criticisms, alternative ways of synthesizing evidence have emerged in which some of the most rigid principles of systematic reviews and meta-analysis are relaxed (Mallett et al., 2012; Slavin, 1995). These alternative ways include: rapid evidence assessments/synthesis, scoping reviews, systematic maps, semi or flexible systematic reviews and best-evidence synthesis and simply following systematic and repeatable search strategies (Koricheva et al., 2013). More ‘informal’ rapid reviews and “realist-based” synthesis have also emerged. These often use broad inclusion criteria for evidence (qualitative and quantitative) to facilitate comparison of impact evaluation methods, develop a transferable theory, and attempt to provide policy-makers with knowledge in response to time sensitive and emerging issues (Victora et al., 2011; Saul et al., 2013; Pawson, 2002). However, the lack of transparency and repeatability might render these informal processes less useful for impact evaluation.

Systematic review approaches have also been developed which utilise qualitative evidence (Noyes, 2010) and are centred predominantly on exploring and progressing theoretical frameworks (Dixon-Woods et al., 2006), investigating system complexity (Sheppard et al., 2017) and placing research within its social context via meta-narratives (Greenhalgh et al., 2005). A configurative systematic review is one example (Gough et al., 2017). Such reviews set out to interpret and understand a concept by configuring information and generating new knowledge/perspectives and are largely concerned with identifying patterns (Barnett-Page et al., 2009).

The methods used for data analysis as part of the review process include configurative and aggregative approaches, or a combination of

the two. Configurative methods aim to formulate ways of understanding phenomena and their meaning/value, usually through the review of qualitative data. Aggregative methods combine the (generally quantitative) findings of similar studies to judge the strength of a conclusion and normally follow a more traditional statistical/meta-analytical approach (Gough et al., 2017). Whereas classic quantitative aggregative reviews are likely to be meta-analysing similar forms of data, configurative reviews are concerned with identifying patterns provided by heterogeneity (Barnett-Page et al., 2009). As such, they are ideal for synthesising evidence from different disciplines or methodologies. The choice between them, or how they are combined, usually depends on data quality and availability, which is often driven by the heterogeneity in methods used by researchers to address the questions underpinning the impact that needs to be evaluated.

The different variables measured, methods used and ways of reporting outcomes is a significant constraint preventing evidence synthesis in systematic reviews. In response to this challenge, a number of attempts have been made to develop standards of evidence in specific domains. For example, the Alliance for Useful Evidence reviewed 18 standards of evidence currently used in UK social policy and called for the creation of a single set of standards that could enable more effective comparison between policy appraisals (Puttick, 2018). This is similar to approaches to evidence in the medical research community (e.g. the use of common outcome measures for chronic pain clinical trials enabling findings to be synthesised across studies in meta-analyses to inform evidence-based medicine policy and practice; Turk et al., 2003) and could in theory be applied to the generation of evidence for research impact.

Regardless of the specific approach taken to the review, or to the analysis of resultant data, one of the great strengths of following systematic approaches, is that reviews are updatable as new evidence becomes available. Thus, systematic approaches allow tracking, through time, of the nature and pathways through which evidence travels through the literature resulting in impact on wider society.

5. A methodological framework for research impact evaluation

In this penultimate section, we explain how the different types of impact evaluation identified in the previous section fit into a broader methodological framework. Fig. 2 shows how research leads to possible impacts via an impact plan and pathways to impact (in the case of serendipitous impacts, the impact plan is missing but pathways can still typically be traced). However, these possible impact claims may be contested in terms of their significance or reach, or on the basis of the evidence that significant or far-reaching impacts can be attributed to the research. Therefore, for impacts to be considered demonstrable, an impact evaluation needs to be designed (denoted by the grey box in Fig. 2). Ideally evaluations can draw on monitoring that has been designed to track progress towards planned impacts (however an evaluation can proceed in the absence of monitoring, drawing on alternative sources of evidence). Monitoring can provide formative feedback that can help adapt and refine pathways (the feedback loop in Fig. 2), increasing the likelihood of delivering impacts. Various types of monitoring can be used as part of the evaluation process depending on the nature and purpose of the impact evaluation.¹ In addition to monitoring data (such as intervention outcome data), the evaluation may produce other evidence (such as health economics evidence of cost savings resulting from the intervention), which taken together demonstrate that significant and far-reaching impacts were derived from the research.

¹ Monitoring can be categorized as follows: i) surveillance monitoring is about assessing long-term changes in conditions resulting from an activity; ii) operational monitoring consists of implementing additional measure for cases where there is risk of failure of not meeting initial directives; and iii) investigative monitoring determines reasons to failure.

Table 1 identifies five types of evaluation design, and Fig. 2 suggests that there are two key factors likely to inform the choice between these evaluation designs. First the choice of evaluation design must be suited to the context in which it is to be used, including the resources available (some types of evaluation design, such as experimental methods, can be time consuming and resource intensive), the scope of the evaluation (e.g. in spatial or temporal scale or the range of linked systems to be considered), the types of impact being evaluated (as noted in Table 1, some types of evaluation design are suited to evaluating certain types of impact), and the ontology and epistemology of the team selecting the evaluation design (see introduction to Section 4). Based on the theoretical constructs that emerged from the analysis of literature (described in the introduction to Section 4), the choice of evaluation design will also reflect the aims of the evaluation, for example the extent to which the evaluate aims to provide summative versus formative feedback, or provide evidence of necessary versus sufficient causal links between research and impact. Evaluations are typically designed to establish relationships between research and impacts along causal chains (which often include the evaluation of knowledge exchange activities or pathways to impact). It can be possible to attribute impact to research through long causal chains, however the strength of evidence for research impact is only as strong as the weakest link in the chain. As a result, attribution in long causal chains is often partial, indicating that research may have been necessary amongst other factors or may have only made a minor, contestable contribution to impact, give the range of confounding factors at play at the end of a long causal chain.

6. Conclusion

With sufficient time and resources, there are now evaluation methods that can be used to monitor and assess almost any impact arising from research. Knowing what delivers impact (and what does not) can help researchers and research evaluators anticipate challenges and avoid using methods that are unlikely to work or that might lead to unintended negative consequences. When things do not go according to plan, evaluation findings can give researchers ideas about how to get things back on track or do things better next time. Whether for funders, the media or the wider public, the process of evaluating impact often enables researchers to communicate the value of research to wider audiences.

In this paper, we have provided new definitions of research impact and impact evaluation informed by our analysis of the literature, including a new way to conceive of reach as scaling up and/or out, that can be applied in any disciplinary context. Based on these definitions, we have sought to simplify the bewildering range of methods and approaches available into five types of evaluation design that can be used to guide the selection of relevant evaluation methods and approaches. Like any typology, there are many alternative ways we could have divided and named the types of evaluation we came across in the review. As a typology of evaluation designs, it includes types of method (e.g. experimental or arts-based) and types of approach (e.g. indicator-based approaches or systematic review). Indicator-based and systematic review approaches may be operationalised using any number of methods, including methods from other parts of the typology. While this introduces potential overlap between types, indicator-based and systematic review approaches are widely used in impact evaluation, and to remove these from the typology to avoid potential overlap would significantly constrain the utility of the typology for identifying the most relevant type of evaluation design for any given purpose or context.

This typology then formed the basis for a wider methodological framework to guide anyone who needs to select a relevant evaluation design and methods to causally link impacts to research and assess their significance and reach. There are almost as many evaluation methods and approaches as there are impacts, and as researchers seek to demonstrate new impacts, methods will continue to evolve. The audience for this paper is also diverse, and the needs of researchers may

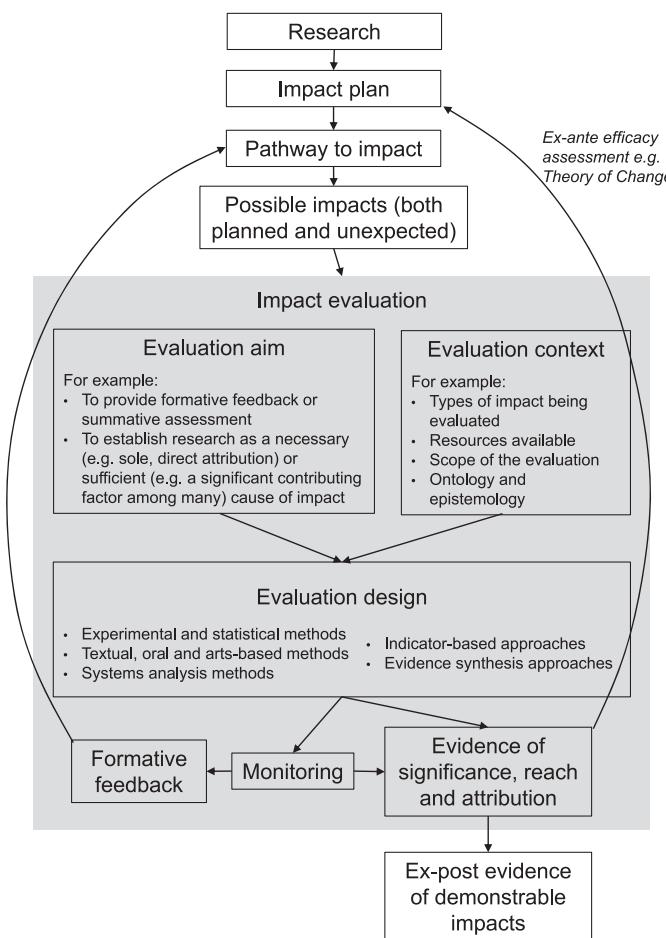


Fig. 2. Methodological framework for evaluating research impact.

differ substantially from those of funders and other stakeholders seeking to evaluate impact. While we have sought to generalise as far as possible through the construction of our typology and methodological framework, to provide methods that can be used across contexts and for different purposes, it is important to recognise the differences between these groups, and how their contexts, perceptions and beneficiaries are likely to change over time. Although it is impossible to capture all possible methods for evaluating impact, we hope that the examples provided under each type of evaluation design will stimulate additional reading and experimentation. Using the methodological framework described in this paper, it should be possible for researchers, funders and other stakeholders working across multiple disciplines to design more effective evaluations to evidence the impact of research.

CRediT authorship contribution statement

M.S. Reed: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Project administration. **M. Ferré:** Writing - original draft, Writing - review & editing. **J. Martin-Ortega:** Writing - original draft, Writing - review & editing. **R. Blanche:** Writing - original draft, Writing - review & editing. **R. Lawford-Rolfe:** Writing - review & editing. **M. Dallimer:** Writing - original draft, Writing - review & editing. **J. Holden:** Writing - review & editing, Funding acquisition.

Declaration of Competing Interest

Professor Mark Reed is CEO of Fast Track Impact Ltd. All other

authors declare that they have no known competing interest.

Acknowledgements

The authors were supported by the Integrated Catchment Solutions Programme (iCASP) funded by the UK Natural Environment Research Council's Regional Impact from Science of the Environment scheme (grant NE/P011160/1).

References

- Alla, K., Hall, W.D., Whiteford, H.A., Head, B.W., Meurk, C.S., 2017. How do we define the policy impact of public health research? a systematic review. *Health Res. Policy Systems* 15, 84.
- Alvarez, S., Douthwaite, B., Thiele, G., Mackay, R., Córdoba, D., Tehelen, K., 2010. Participatory Impact Pathways Analysis: a practical method for project planning and evaluation. *Dev. Pract.* 20, 946–958.
- Australian Research Council (ARC), 2017. Engagement and impact assessment. Access via: <https://www.arc.gov.au/engagement-and-impact-assessment>.
- Bamberger M., 2012. Introduction to mixed methods in impact evaluation. *Impact Evaluation Notes No. 3*.
- Barnett-Page, E., Thomas, J., 2009. Methods for the synthesis of qualitative research: a critical review. *BMC Med. Res. Methodol.* 9, 59.
- Baumeister, R.F., Leary, M.R., 1997. Writing narrative literature reviews. *Rev. General Psychol.* 1, 311.
- Bayley, J.E., Phipps, D., 2017. Building the concept of research impact literacy. *Evidence Policy*. <https://doi.org/10.1332/174426417X15034894876108>.
- Befani, B., Stedman-Bryce, G., 2017 Jan. Process tracing and bayesian updating for impact evaluation. *Evaluation* 23 (1), 42–60.
- Bormann, L., 2012. Measuring the societal impact of research: research is less and less assessed on scientific impact alone—we should aim to quantify the increasingly important contributions of science to society. *EMBO Rep.* 13 (8), 673–676.
- Bozeman, B., Sarewitz, D., 2011. Public value mapping and science policy evaluation. *Minerva* 49, 1–23.

- Bozeman, B., Youtie, J., 2017. Socio-economic impacts and public value of government-funded research: lessons from four US National Science Foundation initiatives. *Res. Policy* 46, 1387–1398.
- Boydell, K., Gladstone, B., Volpe, T., Allemand, B., Stasiulis, E., 2012. The production and dissemination of knowledge: a scoping review of arts-based health research. *Forum Qual. Sozialforschung / Forum: Qual. Social Res.* 13 (1). Art. 32.
- Braham, M., Van Hees, M., 2009. Degrees of causation. *Erkenntnis* 71 (3), 323–344.
- Brewer, J.D., 2011. The impact of impact. *Res Eval* 20, 255–256.
- Bulaitis, Z. (2017) 'Measuring impact in the humanities: learning from accountability and economics in a contemporary history of cultural value', Palgrave Communications, 3: 71–9.
- Byrne, E., Daykin, N., Coad, J., 2016. Participatory photography in qualitative research: a methodological review. *Visual Methodol.* 4 (2), 1–12.
- Byrne, E., Elliott, E., Saltus, R., Angharad, J., 2018. The creative turn in evidence for public health: community and arts-based methodologies. *J. Public Health (Bangkok)* 40 (1), i24–i30.
- Chamberlain, K., McGuigan, K., Anstiss, D., Marshall, K., 2018. A change of view: arts-based research and psychology. *Qual. Res. Psychol.* 15 (2–3), 131–139.
- Chambers, R., 2013. Ideas For Development. Routledge.
- Chapman, D.S., Ttermansen, M., Jin, N., Quinn, C.H., Cornell, S.J., Fraser, E.D.G., Hubacek, K., Kunin, W.E., Reed, M.S., 2009. Modelling the coupled dynamics of moorland management and vegetation in the UK uplands. *J. Appl. Ecol.* 46, 278–288.
- Checkland, P., 2000. Soft systems methodology: a thirty year retrospective. *Syst. Res. Behav. Sci.* 17 (S1), S11–S58.
- Chubb, J., Reed, M.S., 2018. The politics of research impact: implications for research funding, motivation and quality. *British Politics* 13, 295–311.
- Chubb, J., Watermeyer, R., Wakeling, P., 2017. Fear and loathing in the academy? the role of emotion in response to an impact agenda in the UK and Australia. *Higher Educ. Res. Dev.* 36 (3).
- Coates P., Brady E., Church A., Cowell B., Daniels S., DeSilvey C., Fish R., Holyoak V., Horrell D., Mackey S., Pitt R., Stibbe A., Waters R. Arts & humanities perspectives on cultural ecosystem services. Arts and Humanities Working Group Final Report, 2014. Available from: http://randd.defra.gov.uk/Document.aspx?Document=12303_WP5_AandHAnnex1_ArtsandHumanitiesPerspectivesonEcosystemServices_25June.pdf.
- Coemans, S., Wang, C., Leysen, J., Hannes, K., 2015. The use of arts-based methods in community-based research with vulnerable populations: protocol for a scoping review. *Int. J. Educ. Res.* 71, 33–39.
- Crossick, G., Kaszynska, P., 2016. Understanding the Value of Arts and culture: the AHRC Cultural Value Project. Arts and Humanities Research Council.
- Danto, A., 1962. Narrative sentences. *Hist. Theory* 2 (2), 146–179.
- Daykin, N., Gray, K., McCree, M., Willis, J., 2017. Creative and credible evaluation for arts, health and well-being: opportunities and challenges of co-production. *Arts Health* 9 (2), 123–138. May 4.
- Dickson I.M., Butchart S.H.M., Dauncey V., Hughes J., Jefferson R., Merriman J.C., Munroe R., Pearce-Higgins J.P., Stephenson P.J., Sutherland W.J., Thomas D.H.L., Trevelyan R. PRISM – toolkit for evaluating the outcomes and impacts of small/medium-sized conservation projects. Version 1. 2017. Available from www.conservativevaluation.org.
- Dimick, J.B., Ryan, A.M., 2014. Methods for evaluating changes in health care policy: the difference-in-differences approach. *JAMA* 312 (22), 2401–2402. Dec 10.
- Directorate-General for Research and Innovation (European Commission). A new horizon for Europe: impact assessment of the 9th EU framework programme for research and innovation. ISBN 978-92-79-81000-8; 2018.
- Dixon-Woods, M., Cavers, D., Agarwal, S., Annandale, E., Arthur, A., Harvey, J., Hsu, R., Katbamna, S., Olsen, R., Smith, L., Riley, R., 2006 Dec. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Med. Res. Methodol.* 6 (1), 35.
- Donovan, C., Hanney, S., 2011. The 'payback framework' explained. *Res. Eval.* 20 (3), 181–183.
- Douglas, K., Carless, D., 2018. Engaging with arts-based research: a story in three parts. *Qual. Res. Psychol.* 15, 2–3.
- Douthwaite, B., Kuby, T., van de Fliert, E., Schulz, S., 2003. Impact pathway evaluation: an approach for achieving and attributing impact in complex systems. *Agric. Syst.* 78, 243–265.
- Douthwaite, B., Schulz, S., 2011. Spanning the attribution gap: the use of program theory to link project outcomes to ultimate goals in INRM and IPM. In: Paper presented at the INRM Workshop, Cali (Colombia).
- Dunn, V., Mellor, T., 2017. Creative participatory projects with young people: reflections over five years. *Res. All* 1, 284–299.
- Edler, J., Georghiou, L., Blind, K., Uyarra, E., 2012. Evaluating the demand side: new challenges for evaluation. *Res. Eval.* 21 (1), 33–47.
- Egger, M., Juni, P., Bartlett, C., Holenstein, F., Sterne, J., 2003. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? *Empir. Study. Health Technol. Assess* 7, 1–76.
- Ewen, J., Parkin, G., O'Connell, P.E., 2000. SHETRAN: distributed river basin flow and transport modeling system. *J. Hydrol. Eng.* 5, 250–258.
- Falagas, M.E., Pitsouni, E.I., Malietzis, G.A., Pappas, G., 2008. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB J.* 22 (2), 338–342.
- Fraser, K.D., Sayah, F., 2011. Arts-based methods in health research: a systematic review of the literature. *Arts Health* 3 (2), 110–145.
- Funnell, S.C., Rogers, P.J., 2011. Purposeful Program theory: Effective use of Theories of Change and Logic Models. John Wiley & Sons.
- Game, E.T., Tallis, H., Olander, L., Alexander, S.M., Busch, J., Cartwright, N., Kalies, E.L., Masuda, Y.J., Mupepele, A.C., Qiu, J., Rooney, A., 2018. Cross-discipline evidence principles for sustainability policy. *Nature Sustain.* 1, 452.
- Gaunand, A., Hoede, A., Lemarié, S., Matt, M., De Turckheim, E., 2015. How does public agricultural research impact society? a characterization of various patterns. *Res. Policy* 44, 849–861.
- Garbarino, A., Holland, J., 2009. Quantitative and Qualitative Methods in Impact; Issues Paper. Governance and Social Development Resource Centre. Available at: <http://www.gsdrc.org/docs/open/eirs4.pdf>.
- Garcia, C., Dray, A., Aubert, S., Reibelt, L.M., Waeber, P.O., 2015. Scenarios of biodiversity exploring possible futures for management. *Akon'ny Ala* 32.
- Gerstenblatt, P., 2013. Collage portraits as a method of analysis in qualitative research. *Int. J. Qual. Methods* 12 (1), 294–309.
- Gertler P.J., Martinez S., Premand P., Rawlings L.B., Vermeersch C.M.J. Impact evaluation in practice, Washington DC: World Bank; 2011.
- Geuna, A., Piolatto, M., 2016. Research assessment in the UK and Italy: costly and difficult, but probably worth it (at least for a while). *Res. Policy* 45, 260–271.
- Gough, D., Oliver, S., Thomas, J., 2017. An Introduction to Systematic Reviews. editors. Sage. Mar 28.
- Grant, J., 2015. The nature, scale and beneficiaries of research impact: an initial analysis of Research Excellence Framework (REF). In: 2014 impact case studies. Research Report 2015/01. King's College, London, UK. Access via. <https://www.kcl.ac.uk/sspp/policy-institute/publications/Analysis-of-REF-impact.pdf>.
- Greene, E.J., Darley, J.M., 1998. Effects of necessary, sufficient, and indirect causation on judgments of criminal liability. *Law Hum. Behav.* 22 (4), 429–451.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., Peacock, R., 2005. Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Soc. Sci. Med.* 61, 417–430.
- Greenhalgh, T., Thorne, S., Malterud, K., 2018. Time to challenge the spurious hierarchy of systematic over narrative reviews? *Eur. J. Clin. Invest.* 48 (6), e12931.
- Greenwood, J., 2012. Arts-based research: weaving magic and meaning. *Int. J. Educ. Arts* 13 (Interlude 1).
- Halse, C., Mowbray, S., 2011. The impact of the doctorate. *Stud. Higher Educ.* 36, 513–525.
- Harper, D., 2002. Talking about pictures: a case for photo elicitation. *Visual Stud.* 17 (1), 13–26.
- Hewlett, K., Bond, K., Hinrichs-Krapels, S., 2017. The Creative Role of Research: Understanding research Impact in the Creative and Cultural Sector. London: Kings College, London.
- Treasury, H.M., 2011. The Magenta Book: Guidance for Evaluation. HM Treasury, London.
- Joly, P.-B., Gaunand, A., Colinet, L., Larédo, P., Lemarié, S., Matt, M., 2015. ASIRPA: a comprehensive theory-based approach to assessing the societal impacts of a research organization. *Res. Eval.* 24 (4), 1–14.
- Julian, D.A., Jones, A., Deyo, D., 1995. Open systems evaluation and the logic model: program planning and evaluation tools. *Eval. Program. Plann.* 18 (4), 333–341.
- Kenter, J.O., Reed, M.S., Irvine, K.N., O'Brien, E., Brady, E., Bryce, R., Christie, M., Church, A., Cooper, N., Davies, A., Hockley, N., Fazey, I., Jobstvogt, N., Molloy, C., Orchard-Webb, J., Ravenscroft, N., Ryan, M., Watson, V., 2014. UK national ecosystem assessment follow-on. In: Work Package Report 6: Shared, Plural and Cultural Values of Ecosystems. UNEP-WCMC, LWEC.
- Khandker, S., B. Koolwal, G., Samad, H., 2009. Handbook On Impact evaluation: Quantitative Methods and Practices. The World Bank.
- Koricheva, J., Gurevitch, J., Mengerson, K., 2013. Handbook of Meta-Analysis in Ecology and Evolution. editors. Princeton University Press.
- Lance, P., Guilkey, D., Hattori, A., Angeles, G., 2014. How Do We Know If a Program Made a difference? A guide to Statistical Methods For Program Impact Evaluation. MEASURE Evaluation, Chapel Hill, North Carolina.
- Mackie, J.L., 1974. The Cement of the universe: a study of Causation. Oxford University Press, Oxford, UK.
- Mallett, R., Hagen-Zanker, J., Slater, R., Duvendack, M., 2012. The benefits and challenges of using systematic reviews in international development research. *J. Dev. Effectiveness* 4 (3), 445–455.
- Martín-Martín, A., Orduna-Malea, E., Thelwall, M., López-Cózar, E.D., 2018a. Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. *J. Informetr.* 12 (4), 1160–1177.
- Martín-Martín, A., Orduna-Malea, E., López-Cózar, E.D., 2018b. Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics* 116 (3), 2175–2188.
- Matt, M., Gaunand, A., Joly, P.B., Colinet, L., 2017. Opening the black box of impact-ideal-type impact pathways in a public agricultural research organization. *Res. Policy* 46 (1), 207–218.
- Mayne, J., 2012. Making causal claims. In: ILAC Brief 26. CGIAR. Available at: https://cgspace.cgiar.org/bitstream/handle/10568/70211/ILAC_Brief26_Making%20causal%20claims.pdf?sequence=1.
- Mitchell, V., 2019. A proposed framework and tool for non-economic research impact measurement. *Higher Educ. Res. Dev.* 1–4. Jun 7.
- Moon, K., Blackman, D., 2014. A guide to understanding social science research for natural scientists. *Conserv. Biol.* 28 (5), 1167–1177.
- Moore, G., Croxford, B., Adams, M., Refaei, M., Cox, T., Sharples, S., 2008. The photo-survey research method: capturing life in the city. *Visual Stud.* 23 (1), 50–62.
- Morris, Z.S., Wooding, S., Grant, J., 2011. The answer is 17 years, what is the question: understanding time lags in translational research. *J. R. Soc. Med.* 104, 510–520.
- Morton, S., 2015. Creating research impact: the roles of research users in interactive research mobilisation. *Evidence Policy* 11 (1), 35–55.
- Moss, M.L., 1981. Genetics, epigenetics, and causation. *Am. J. Orthod.* 80 (4), 366–375.

- National Science Foundation (NSF) Perspectives on Broader Impacts. 2014; Available from: https://www.nsf.gov/od/ia/publications/Broader_Impacts.pdf.
- Niederman, F., Crowston, K., Koch, H., Kremer, H., Powell, P., Swanson, E.B. 2015. Assessing IS research impact. *CAIS* 36, 7.
- Noyes, J., 2010. Never mind the qualitative feel the depth! The evolving role of qualitative research in Cochrane intervention reviews. *J. Res. Nurs.* 15, 525–534.
- Nutley, S., Walter, I., Davies, H.T.O. 2007. Using evidence: How research Can Inform Public Services. Policy Press, Bristol.
- Oancea, A., 2019. Research governance and the future(s) of research assessment. *Palgrave Commun.* 5 (1) art. no. 27.
- OECD, 2001. Environmental Indicators for Agriculture: Methods and Results, vol. 3. OECD, Paris.
- Parascandola, M., Weed, D.L., 2001. Causation in epidemiology. *J. Epidemiol. Community Health* 55 (12), 905–912.
- Patton, M.Q., 1996. A world larger than formative and summative. *Eval. Pract.* 17 (2), 131–144.
- Pawson, R. 2002. Evidence-based policy: in search of a method. *Evaluation* 8 (2), 157–181.
- Pearl, J., 1999. Probabilities of causation: three counterfactual interpretations and their identification. *Synthese* 121 (1–2), 93–149.
- Penfield, T., Baker, M.J., Scoble, R., Wykes, M.C. 2014. Assessment, evaluations, and definitions of research impact: a review. *Res. Eval.* 23, 21–32.
- Pullin, A.S., Knight, T.M., 2001. Effectiveness in conservation practice: pointers from medicine and public health. *Conserv. Biol.* 15, 50–54.
- Puttick, R., 2018. Mapping the Standards of Evidence used in UK Social Policy. Big Lottery Fund, the Economic and Social Research Council and Nesta. Access via: <https://www.alliance4usefulevidence.org/assets/2018/05/Mapping-Standards-of-Evidence-A4UE-final.pdf>.
- Rau, H., Goggins, G., Fahy, F. 2018. From invisibility to impact: recognising the scientific and societal relevance of interdisciplinary sustainability research. *Res. Policy* 47, 266–276.
- Rebora, G., Turni, M., 2013. The UK and Italian research assessment exercises face to face. *Res. Policy* 42 (9), 1657–1666. Nov 1.
- Reed, M.S., Bryce, R., Machen, R. 2018a. Pathways to policy impact: a new approach for planning and evidencing research impact. *Evidence Policy* 14, 431–458.
- Reed, M.S., Dougill, A.J., Baker, T. 2008. Participatory indicator development: what can ecologists and local communities learn from each other? *Ecol. Appl.* 18, 1253–1269.
- Reed, M.S., Evelyn, A.C., Cundill, G., Fazey, I., Glass, J., Laing, A., Newig, J., Parrish, B., Prell, C., Raymond, C., Stringer, L.C. 2010. What is social learning? *Ecol. Soc.* 15 (4), r1 online.
- Reed, M.S., Fraser, E.D., Dougill, A.J. 2006. An adaptive learning process for developing and applying sustainability indicators with local communities. *Ecol. Econ.* 59 (4), 406–418.
- Reed, M.S., 2018. The Research Impact Handbook, 2nd Edition. Fast Track Impact, Huntly, Aberdeenshire.
- Research England. Guidance on Submissions. REF 2019/01, 2019. Access via: <https://www.ref.ac.uk/publications/guidance-on-submissions-201901>.
- Richards, M., Panfil, S.N., 2011. Towards cost-effective social impact assessment of REDD+ projects: meeting the challenge of multiple benefit standards. *Int. Forestry Rev.* 13 (1).
- Richards, M. Issues and challenges for social evaluation or Impact Assessment of 'multiple-benefit' Payment for Environmental Services (PES) projects. Unpublished review for United Nations Forum for Forests. Forest Trends, Washington, DC. 2008. Available at: http://moderncms.ecosystemmarketplace.com/repository/moderncms/documents/SFCM_2009_smaller.pdf.
- Rihoux, B., Ragin, C.C. 2008. Configurational Comparative methods: Qualitative comparative Analysis (QCA) and Related Techniques. Sage Publications.
- Samuel, G.N., Derrick, G.E. 2015. Societal impact evaluation: exploring evaluator perceptions of the characterization of impact under the REF2014. *Res. Eval.* 24 (3), 229–241.
- Sanjari, M., Bahramnezhad, F., Fomani, F.K., Shor, M., Cheraghi, M.A. 2014. Ethical challenges of researchers in qualitative studies: the necessity to develop a specific guideline. *J. Med. Ethics Hist. Med.* 7, 1–6.
- Saul, J.E., Willis, C.D., Bitz, J., Best, A. 2013. A time-responsive tool for informing policy making: rapid realist review. *Implem. Sci.* 8, 103.
- Scrimner, M., 1991. Beyond formative and summative evaluation. In: McLaughlin, M.W., Phillips, D.C. (Eds.), *Evaluation and Education: At Quarter Century*. The University of Chicago Press, Chicago, IL, pp. 18–64.
- Sheppard, C., Davy, S., Pilling, G., Graham, N. 2017. *The Biology of Coral Reefs*. Oxford University Press.
- Slavin, R.E. 1995. Best evidence synthesis: an intelligent alternative to meta-analysis. *J. Clin. Epidemiol.* 48, 9–18.
- Spaapen, J.M., Van Drooge, L., 2011. Introducing productive interactions in social assessment. *Res. Eval.* 20, 211–218, 1995.
- Spanish Government. Resolución de 28 de noviembre de 2018, de la Secretaría de Estado de Universidades, Investigación, Desarrollo e Innovación, por la que se fija el procedimiento y plazo de presentación de solicitudes de evaluación de la actividad investigadora a la Comisión Nacional Evaluadora de la Actividad Investigadora. BOE-A-2018-16379; 2018.
- Stachowiak, S., 2013. *Pathways For change: 10 Theories to Inform Advocacy and Policy Change Efforts*, ORS Impact. Center for evaluation Innovation.
- StarMetrics, 2016. *Science and technology for America's reinvestment measuring the effects of research on innovation. Competitiveness and Science. Process Guide*. Office of Science and Technology Policy, Washington DC.
- Stem, C., Margoluis, R., Salafsky, N., Brown, M. 2005. Monitoring and evaluation in conservation: a review of trends and approaches. *Conserv. Biol.* 19 (2), 295–309.
- Strauss, A., Corbin, J.M., 1997. *Grounded Theory in Practice*. Sage.
- Sundin, A., Andersson, K., Watt, R. 2018. Rethinking communication: integrating storytelling for increased stakeholder engagement in environmental evidence synthesis. *Environ. Evidence* 7 (1), 6.
- Sutherland, W.J., Pullin, A.S., Dolman, P.M., Knight, T.M. 2004. The need for evidence-based conservation. *Trends Ecol. Evol. (Amst.)* 19 (6), 305–308. Jun 1.
- Tian, J., Pearl, J., 2000. Probabilities of causation: bounds and identification. *Ann. Math. Artif. Intell.* 28 (1–4), 287–313.
- Turk, D.C., Dworkin, R.H., Allen, R.R., Bellamy, N., Brandenburg, N., Carr, D.B., Cleeland, C., Dionne, R., Farrar, J.T., Galer, B.S., Hewitt, D.J. 2003. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain* 106 (3), 337–345. Dec 1.
- UKRI 2018 (UK Research Innovation) pathways to impact. Available at: <https://www.ukri.org/innovation/excellence-with-impact/pathways-to-impact/>.
- UNEG, 2013. *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*. United Nations Evaluation Group, New York. University Grants Committee. Framework for Research Assessment Exercise (RAE) 2020. Access via: <https://www.ugc.edu.hk/doc/eng/ugc/rae/2020/framework.pdf>; 2017.
- USAID, 2011. *USAID Evaluation Policy: Evaluation, Learning from Experience*. United States Agency for International Development, Washington DC.
- Van der Vaart, G., van Hoven, B., Huigen, P. 2018. Creative and arts based research methods in academic research: lessons from a participatory research project in the Netherlands. *FQS Forum Qual. Social Res.* 19 (2), 19.
- Victora, C.G., Black, R.E., Boerm, J.T., Bryce, J. 2011. Measuring impact in the Millennium Development Goal era and beyond: a new approach to large-scale effectiveness evaluations. *Lancet* 377, 85–95.
- VSNU/KNAW/NOW. Protocol for Research Assessments in the Netherlands. Access via: <https://www.knaw.nl/nl/actueel/publicaties/standard-evaluation-protocol-2015-2021>; 2014.
- Wang, C.C., Yi, W.K., Tao, Z.W., Carovano, K. 1998. Photovoice as a participatory health promotion strategy. *Health Promot. Int.* 13 (1), 75–86.
- Watermeyer, R., 2019. *Competitive Accountability in Academic life: the Struggle For Social Impact and Public Legitimacy*. Edward Elgar Publishing.
- Woolcock, M., 2013. Using case studies to explore the external validity of "complex" development interventions. *Evaluation* 19, 229–248.
- Woolcott, G., Keast, R., Pickernell, D. 2019. Deep impact: re-conceptualising university research impact using human cultural accumulation theory. *Stud. Higher Educ.* 1–20. Mar 21.