A proposed framework and tool for non-economic research impact measurement

Article *in* Higher Education Research and Development · March 2019

DOI: 10.1080/07294360.2019.1590319

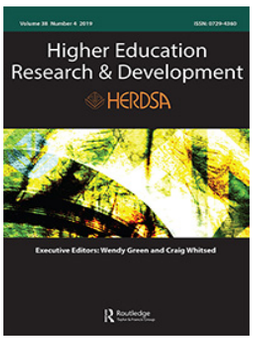
CITATIONS
10

READS
275

1 author:

Vincent-Wayne Mitchell
The University of Sydney
158 PUBLICATIONS **10,863** CITATIONS

SEE PROFILE



Higher Education Research & Development

ISSN: 0729-4360 (Print) 1469-8366 (Online) Journal homepage: https://www.tandfonline.com/loi/cher20

A proposed framework and tool for non-economic research impact measurement

Vincent Mitchell

To cite this article: Vincent Mitchell (2019) A proposed framework and tool for non-economic research impact measurement, Higher Education Research & Development, 38:4, 819-832, DOI: 10.1080/07294360.2019.1590319

To link to this article: https://doi.org/10.1080/07294360.2019.1590319

HIGHER EDUCATION RESEARCH & DEVELOPMENT
2019, VOL. 38, NO. 4, 819–832
https://doi.org/10.1080/07294360.2019.1590319

Routledge
Taylor & Francis Group

A proposed framework and tool for non-economic research impact measurement

Vincent Mitchell

Discipline of Marketing, The University of Sydney Business School, Sydney, Australia

ABSTRACT

Research impact features heavily in debates about 'the measured university' and is now formally assessed by governments in the UK and Australia. Yet clear guidance on how impact can be measured in non-monetary ways is often lacking because of confused thinking and the context-specific nature of outcomes. To help resolve this, we first propose a general impact model of inputs, processes, outputs and outcomes framework. Then, using a survey approach of research users, we measure outcomes at different levels of abstraction using usefulness as the central construct and impact categories from the European (EU) and Association to Advance Collegiate Schools of Business (AACSB). The survey measures are simple, comparable between different impact cases, cost-effective, externally verifiable, and easily administered by those for whom impact measurement is new and puzzling. They can also be combined to form an impact index to address criticism that there is little standardisation in impact measurement. To improve the standardisation of context-specific measures, we suggest a common methodology for deriving these (SROI). The article discusses limitations of using surveys including administration, self-report data, and impact timescale problems, and suggests ways to reduce these. Implications for researchers, research managers, and assessors interested in measuring impact, such as who should do it and pay for it, are discussed.

ARTICLE HISTORY
Received 24 April 2018
Accepted 22 November 2018

KEYWORDS
Impact policy; impact measurement; research outcomes; research evaluation

Introduction

> *The philosophers have only interpreted the world in various ways. The point however is to change it.* (Karl Marx's gravestone)

Although the impact debate is not new (Grey, 2001), there has been an increasing focus on how to measure and demonstrate research impact effectively as more countries adopt measurement frameworks, including Italy's Research Quality Evaluation (VQR), the UK's Research Excellence Framework (REF), Australia's Excellence in Research for Australia (ERA), France's Agency for the Evaluation of Research and Higher Education (AERES), the Netherlands' Evaluating Research in Context (ERiC), and Belgium's Industrial Research Fund (IOF) (Adam et al., 2018; Technopolis, 2014). Consequently, universities around the world are under increased pressure to demonstrate their value or

CONTACT Vincent Mitchell vince.mitchell@sydney.edu.au

usefulness to a wider group of stakeholders (Learmonth, Lockett, & Dowd, 2012), as part of a wider phenomenon in academic life, known as 'the measured university' (Peseta, Barrie, & McLean, 2017). Apart from resisting the impact agenda (Chubb, Watermeyer, & Wakeling, 2017), an alternative approach is to make the case that academic research has always had an impact, but previously academics have seldom bothered to measure it. This is because the measurement is difficult and there has been an absence of readily available measures (Rymer, 2011, p. 4). This has led to universities intensifying their efforts in activities that can be measured, rather than activities that are harder to measure, but are more useful to society (Ernø-Kjølhede & Hansson, 2011). Even measures in the most comprehensive impact assessment to date, the UK REF, were found to be diverse and inconsistent in their use and expression and could not be synthesised (Grant, 2015). Coupled with this is a concern that this inconsistency of evidence is evaluated differently by the various expert assessors making comparisons between disciplines and institutions difficult (Mårtensson, Fors, Wallin, Zander, & Nilsson, 2016). Unsurprisingly then, doubts have been raised about the adequacy of the operationalisation of impact for evaluating the ways in which research and knowledge translation are actually carried out (Smith, Ward, & House, 2011). The problem can be characterised by Cameron's famous quote, 'Not everything that can be counted counts, and not everything that counts can be counted' (Cameron, 1963, p. 43).

Responding to calls that assessments of societal research impact are much needed (Bornmann, 2013) and building on a suggestion by the National Innovation and Science Agenda (NISA) that surveys of end-users or beneficiaries of research may be valuable indicators (NISA, 2016, p. 15), this article proposes a survey measure of research usefulness as a way of capturing non-economic impact. It contributes by first clarifying what is impact using an outcomes perspective that expands on previous work on hierarchies of research quality (Mårtensson et al., 2016). Second, it proposes a new simple Net Recommendation Score, as well as a 2-level outcome measure that can be combined to form an impact index out of 98. Third, it proposes a common method (SROI) for assessing context-specific outcomes. These not only solve the comparability problem, but also the problem of not knowing what to assess or how to assess it for those new to impact. Finally, it critically self-evaluates survey outcome measures to present a realistic assessment of their worth and gives ideas for further research. The article could change both researchers' and institutional practice in demonstrating the value of our research work by responding constructively to the measured university agenda.

Defining research impact

Although how exactly impact should be defined is widely debated, there is some agreement from the definitions which follow that in order for research to be impactful, it needs to be useful. For example, taking 'academic work and turning it into knowledge that is useful and used by business, government, and society more broadly' (Mason, 2015, p. 3), creating 'an effect on, change or benefit to the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia' (Higher Education Funding Council for England [HEFCE], 2015a) and providing a 'contribution … to the economy, society, environment and culture beyond the contribution to academic research' (ARC,

2017, p. 4). Although usefulness has also been framed as 'relevance' (Butler, Delaney, & Spoelstra, 2015, p. 5), the two concepts are distinct, since researchers can work on relevant practitioner problems, but come up with useless research findings.

Measuring research impact

The most dominant impact 'measure' used is case studies (Abrahamson & Fairchild, 1999; Millo & MacKenzie, 2009; Nicolai & Dautwiz, 2010), both in the UK REF and Australian ERA. Case studies comprise both quantitative and qualitative data, allowing evidence to be contextualised and a story told that preserves a distinctive account or disciplinary perspective (Penfield, Baker, Scoble, & Wykes, 2014), and enable authors to articulate a range of impacts which would not been have captured through a 'top-down taxonomy' (HEFCE, 2015b, p. 71).

But the question remains about what information should be presented as evidence of impact within each case? Merkx, van der Weijden, Oostveen, van den Besselaar, and Spaapen (2007) discuss the benefits and drawbacks of a range of evaluation tools such as: bibliometrics (Schulz & Nicolai, 2015), economic rate of return, peer review, case study, logic modelling and benchmarking (Grant, 2006), historical approaches (Augier & March, 2011; Wensley, 2007), and textual approaches (Beech, MacIntosh, & MacLean, 2010; Nicolai & Seidl, 2010; Pearce & Huang, 2012). Measuring Impacts Under CERIF (MICE), which has over 100 indicators, has been criticised for being insufficiently context specific (Penfield et al., 2014) while other tools for evaluating specific types of research, like the effect of training (Kirkpatrick, 1996), can be criticised as too complex to administer and is mainly applicable to organisational behaviour research. Although recent UK government guidance on how to report measures (Parks, Ioppolo, Stepanek, & Gunashekar, 2018) is useful, it stops far short of saying what the impact measures should be, apart from obvious outcomes, such as jobs created, money made, and pollution emission reduction. Other categorisations (Lähteenmäki-Smith, Hyytinen, Kutinlahti, & Konttinen, 2006; Martin & Tang, 2007; Meagher, Lyall, & Nutley, 2008), are limited in their coverage of the possible beneficial outcomes of a research project, do not distinguish high versus low usefulness, and fail to identify how such outcomes can be rigorously and convincingly measured, especially when it comes to non-economic, social benefits.

One general problem often seen in impact measures is that they capture four different points in the process: research creation, dissemination of findings, research use, and potential benefits (de Jong, Barker, Cox, Sveinsdottir, & van den Besselaar, 2014). For example, REF2021 advice still includes measuring impact using engagement by number of people, and media mentions (Parks et al., 2018), while Appendix F of ARC's advice on engagement metrics talks about book sales or sitting on advisory boards (ARC, 2018). But media mentions, downloads, books sales and event participation, and advisory boards are output measures, which do not actually measure what the end-user did with the knowledge which is the outcome. Here, we clarify this confusion and conflation (see Figure 1). Outcomes are the final effect of outputs and include changes to behaviour or thinking such as: changing government policy or a professional body's practice and changes to firm performance or jobs created because of the research. Outputs are measures of how the research knowledge was conveyed as a result of processes such as: media coverage, non-

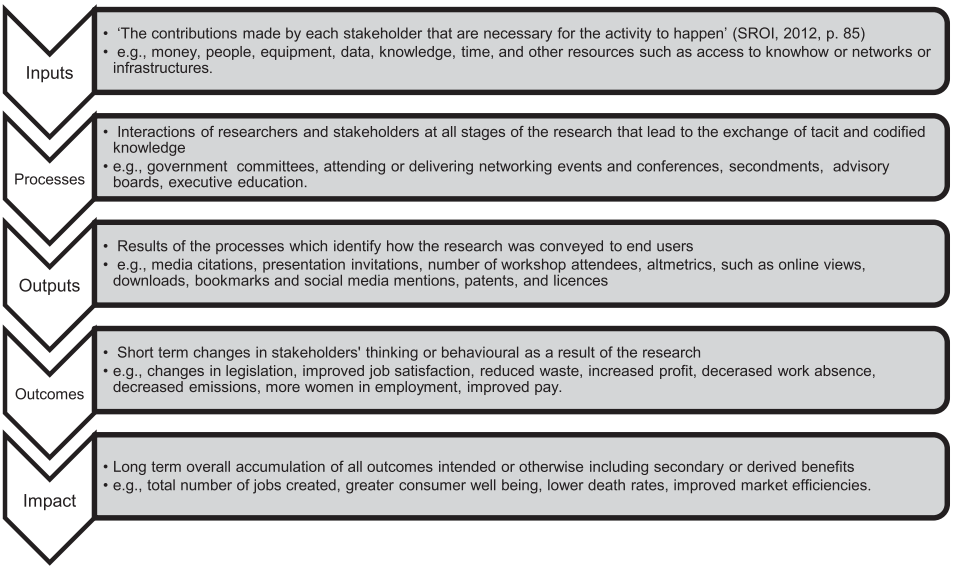


Figure 1. A general framework of research inputs, processes, outputs, outcomes and impact.

academic citations, direct policy advice and expert commentary to government, and website activity. Processes are necessary to translate inputs into outputs and might include: university–industry knowledge transfer partnerships, research-in-action workshops, executive education, webinars, and other dissemination activities. Inputs are resources such as money, time, data, people, and research knowledge.

This model is used by other organisations (INTRAC, 2015; OECD, 2010) for assessing social impact. They view impact as being the total collection of outcomes – negative/positive, direct/indirect, and intended/unintended – and look at them more holistically over a longer period of time. The differences then between outcomes and impact are the level at which the benefit or outcome is measured, with impacts being a higher order or resultant benefit of a given outcome and the longer timescale involved. For example, an outcome of a new method for testing people's sight might be to improve 100 people's sight this year. However, the impact may be a decreased number of accidents, improved workforce participation, and better quality of life for those 100 individuals over the next 50 years. Such secondary outcomes can sometimes be identified by asking why the first outcome is important (Reynolds & Gutman, 1988), for example, 'Why is improving sight important?' This results in secondary potential outcomes, for example, decreased number of accidents, and 'Why is that important', because this results in a better quality of life and reduction of emergency treatment costs.

In summary, existing impact measures are often neither generic enough to make comparisons, nor cost effective, understandable, and simple to administer to most cases, or confuse processes, outputs, and outcomes. To address some of these problems, this article proposes two generic measures based on the concept of usefulness and a generic method for assessing context-specific outcomes.

A proposed new impact outcome measures: what, how and who?

What is useful?

In understanding our concept of research usefulness, we draw on the concept of *relative advantage* (Greenhalgh, Robert, Macfarlane, Bate, & Kyriakidou, 2004; Rogers, 1995), which we define as 'the degree to which the research knowledge is perceived as better than the idea it supersedes'. This could be comprised of three components. First is *newness*, which is how new the knowledge is to the recipient. Second is *disruption*, namely how much of our previous thinking needs to be undone or changed or how difficult and different is the new way of thinking (Christensen, Raynor, & McDonald, 2015). Disruption can be problematic if not considered carefully as 'useful' research may only help to consolidate existing vested interests or ways of thinking rather than to challenge and change them, for example, initial research on climate change was not useful to those with vested interests in the energy sector. Third is *resonance*, which is how much the research connects with the assumptions of organisational members and conforms to their interests (Beyer & Trice, 1987; Nicolai & Dautwiz, 2010). These interests and assumptions can cover a manager's identity, the social rules in specific situations (Wilhelm & Bort, 2013) and audience receptiveness (Blumler & Katz, 1974; Groß, Heusinkveld, & Clark, 2015). Whilst disruption and resonance appear to conflict, it is possible that disruptive ideas can also resonate with an audience. Next, we consider the level at which the research might be useful within a given organisation, namely, their job (micro), their department (meso), and their organisation (macro).

How to measure usefulness: a simple, generic comparable measure

Table 1, section one, shows the items used to measure the above concepts using simple survey questions (see Table 1, level 1). Seven-point scales were used as there is little difference between 5, 7, and 10-point scales (Dawes, 2008). Similar to the Net Promoter Score for firms (Keiningham, Cooil, Andreassen, & Aksoy, 2007), we propose an overall recommendation question, where 1–3 are considered negative, 4 is considered neutral, while 5–7 are considered positive. This could be converted into a simple Net Recommendation Score by looking at the percentage of people who scored 5–7 minus those who scored 1–3. In addition, questions 1–7 in level one could be added to form a usefulness index.

What are the impact outcomes categories?

Impact has essentially two elements – as Davis puts it, 'Good for what? Good for whom?' (Davis, 2015, p. 180). In answering the 'Good for what?' question, the UK REF categorises impacts as things to be impacted on such as: 'products, processes, behaviours, policies, practices; and avoidance of harm or the waste of resources' (REF, 2014, p. 27). Taking a different approach, the Association to Advance Collegiate Schools of Business (AACSB, 2012) refers to three categories of usefulness: *instrumental use* is when 'the findings of a research study directly influence managerial action' (Astley & Zammuto, 1992, p. 452); *conceptual use* is when 'ideas, concepts, or scientific research results influence how a practitioner conceptualises a problem without specific, direct usage' (Astley & Zammuto, 1992,

Table 1. Possible measurement items for a survey of 'What' and 'Who' the research is useful for.

Level 1 Simple Usefulness Measure	1. How likely are you to recommend this research to your peers? (1 not at all likely, 4 neutral, 7 extremely likely) 2. How new would you rate these research findings? (1 not very new, 4 neutral, 7 extremely new) 3. How much of your previous thinking needs to be undone or changed because of these findings? (1 not very much, 4 neutral, 7 a lot; used for rest of questions) 4. How much does the research resonate with your assumptions regarding this topic? 5. How useful was this research for you in your job?', 6. How useful was this research for your department?' 7. How useful was this research for your firm?' (1 not very useful, 4 neutral, 7 extremely useful) 8. Can you give a brief example of which type of people the research is useful for and how it is useful to them? Open text box.
Level 2 Outcome Type Measure	Please answer all the following questions with reference to what you know about the research. 1. To what extent does this research lead to economic benefits? If you rated above 4, please give a brief example. Open text box. 2. To what extent does this research lead to social benefits (e.g., quality of life). If you rated above 4, please give a brief example. 3. To what extent does this research lead to environmental benefits? If you rated above 4, please give a brief example. 4. To what extent does this research lead to cultural benefits (e.g., stimulating creativity)? If you rated above 4, please give a brief example. 5. How much does the research directly influence managerial action? 6. How much does the research influence how practitioners think about a problem? 7. How much is the research used to undermine a current debate within your sector?
Who is the research useful for?	1. How many people do you think this research will be useful for? (please give a number) 2. How many years do you think the implications of this research will take to change your activities? 3. How many years do you think the implications of this research will last in your industry? 4. Can you give a brief description of the type of people it may useful for?

p. 452); *symbolic use* is when research results are used 'to legitimate and sustain predetermined positions' (Beyer & Trice, 1987, p. 598). Others have used this framework and added *widespread use* (Ozanne et al., 2016), which is not so much a different category of use, but rather a quantification of conceptual or instrumental use. It can also be argued that symbolic use is not a separate category, but a case of conceptual use targeted at changing an agenda. More conventionally, the European Commission and UK REF proposes outcome areas such as: (a) economic benefits; (b) social benefits (e.g., quality of life); (c) environmental benefits; and (d) cultural benefits (e.g., stimulating creativity) (European Commission, 2011; Parks et al., 2018). Our proposed second level of measurement incorporates a combination of the EU and AACSB approaches.

How to measure outcome categories

Specifically, our level-two impact assessment uses the four macro outcome areas (European Commission, 2011) and the three uses to which research is put (AACSB, 2012) (see Table 1, level 2 for specific questions). To extend this, respondents could be given a free-text box for each question and asked to give an example or data that would be helpful to justify their rating. Measurement levels one and two could be combined to form some type of impact index, for example, questions 1–7 in level one could be added to form a usefulness index and combined with questions 1–7 in level 2 could

form a simple impact index out of 98 with 14–28 very poor, 29–56 poor, 57–84 good, 85–98 exceptional. The assumption made here is that each item is equally weighted. This is unlikely to be true, but for the sake of simplicity, we will work with that assumption for now. As well as being reported as part of the evidence for impact in case studies, if there is a need from the funders or ERA or REF reporting requirements for additional impact evidence, those who scored between 85–98 could be followed up to participate in a more detailed context-specific analysis outlined next.

How to measure context-specific outcomes

In more complicated context-specific impact measures, such as The Payback Framework, the authors link research with the specific associated benefits (Hanney & González-Block, 2011; Scoble, Dickson, Hanney, & Rodgers, 2010). To do this systematically we advocate using the Social Return on Investment (SROI) framework, which measures outcomes in ways that are relevant to the people or organisations that contribute to it (see SROI Network, 2012, and www.socialvalueuk.org). To explore this method, we use an example of an intervention which helps people recover from mental illness through recycling computers (NEF, 2008). It involves six stages. First, establishing the scope and identifying key stakeholders, for example, family of mentally-ill employees, other employees, the mentally-ill employees themselves, the company, the local health services. Second, by engaging with stakeholders we establish the relationship between inputs (e.g., skills and time, processes, such as training in computer recycling), outputs (e.g., IT skillset and number of recycled computers), and outcomes which can be both intermediary (e.g., increased self-confidence, improved mental health), and final outcomes or impact (e.g., sustainable long-term employment over 10 years). For those outcomes that can be foreseen, it is worthwhile establishing current benchmarks prior to the research being carried out, for example, what are the typical self-confidence measures of mentally-ill employees or what were their employment data over the past 10 years. For those outcomes that cannot be predicted, for example, work-place accidents, this is where having good interaction processes with stakeholders will allow for their capture later. Third, is to evidence outcomes and give them a value while being careful not to confuse outputs with outcomes (see Figure 1), for example, weekly mental health measures, participation in non-work social events, numbers who move into permanent jobs after participating in the scheme. SROI uses financial proxies to estimate the social value of non-traded goods to different stakeholders and includes cost savings as well as increases in income, for example, improvements in mental health could be costed in terms of fewer drugs, less counselling, fewer hospital appointments. Fourth, is to estimate, where possible, those aspects of change that would have happened anyway because of other factors, sometimes called 'deadweight', for example, how many computers would have been recycled anyway without the scheme or what is the typical mental health improvement of patients over time who do not work? This partly tackles the problem of attributing the research to its impact. 'Drop-off' is used to account for how long the impact lasts and outcomes are only calculated if they that last more than one year. This helps to reduce the problem of research fads. Fifth, calculating the final SROI involves adding up all the benefits, for example, number of participants, their mental health improvement, number of computers recycled, and subtracting any negatives, for example, the costs of the scheme, the effect on other co-

workers who became demotivated because of participants' erratic behaviour, and comparing the result to the investment to arrive at a return ratio of £1 value for £1 investment. Finally, there is a stage of reporting, using, and embedding, including sharing findings with stakeholders and responding to them (for more details see NEF, 2008).

Who are the beneficiaries of outcomes?

By *who* we mean the types of beneficiary (individuals, organisations, communities, regions, and other entities) (REF, 2014, p. 27). Knowing who the research benefits is enormously helpful in collecting data on how it has benefitted them. The next question is 'how many' beneficiaries have used the research, which is called 'reach' in the UK REF. We see reach as having two conceptual components: *scope*, which is the degree to which the knowledge is useful for one person/firm or many; and *time*, which is how long the research will take to change thinking and behaviour and how long this will last (see Table 1, bottom section). Measuring scope and time may also help to distinguish true impact (i.e., longer-lasting effects) from management fads (i.e., shorter-term effects) where the initial idea is over-hyped compared to its long-term actual benefit. In the next section we critique our proposed level one and two measurement system to make users fully aware of its limitations.

Limitations of the proposed measures

The traditional limitations of self-report survey data and impact timing consideration apply. For example, there may be non-response bias and under-reporting, which may drastically underestimate usage. While the survey questions are simple and straightforward to answer, they are general perceptual measures that suffer from self-reporting biases such as over-optimism. These limitations may be reduced by the triangulation of data sources such as text mining of social media and sociometrics (Ozanne et al., 2016). For example, after a piece of research is released to a relevant LinkedIn group, the discussion about the research for the next year could be analysed to extract practitioners' views about the work (e.g., 'I used X technique and it saved me weeks of time'). Once example comments have been extracted, an automated search function can be created to identify the number of times 'weeks of time' is mentioned as an outcome. Such textual analysis may also provide previously unexplored outcomes. That said, it should be acknowledged that online sources too have limitations in that not all practitioners are willing, have the skills, and/or have access or authority to record their thoughts and feelings on social media. This is especially apparent when engaging in commercially or politically sensitive research, as well as with clients in the defence industry or invention and IP creation.

A further troubling issue with one-off surveys relates to the timescale of impact measurements, which may underestimate long-term impact (Buxton, 2011). In business and management this may be up to seven years.[1] This is because impact 'varies over time and can change, positively or negatively, at the one-point snapshot whenever it is measured' (Brewer, 2011, p. 256) and because utilised knowledge becomes edited and translated (Czarniawska & Sevon, 1996), or re-interpreted (Seidl, 2007). What is useful in one point in time (i.e., a management fad) may diminish over time and even be

revealed to damage organisations over time (Starkey & Madan, 2001). Alternatively, even useless research may over time prove to have uses (Learmonth et al., 2012). To reduce this problem, several surveys at different times will be needed depending on the rate of change within an industry and the rate of adoption of the research findings. In addition, the survey could be adapted to specify a particular time period for respondents to focus on, for example, in the last year or three years which may introduce errors due to memory.

While the survey measures benefit from being easy to understand, administer, and interpret for non-experts and, as such, they are well within the scope of the individual researcher, research centre, university, grant awarding body, or the government body which assesses impact, they still require some time and money. These resource considerations become much more problematic when doing an evaluative SROI analysis that can take several months (SROI Network, 2012, p. 13) and may take many 'scientists beyond the bounds of their disciplinary expertise' (Holbrook & Frodeman, 2011, p. 244). This raises the question of where does the responsibility lie for measuring impact and who will be responsible for training, motivating and rewarding researchers to do this well? Historically there have been few incentives that value the practical impact of research (Starkey, Hatchuel, & Tempest, 2009) other than altruism. The present system of it being the responsibility of individual researchers seems untenable and a more institutionally supported model needs urgent consideration.

Discussion

We acknowledge the argument that some have put forward that such measures may not only distort the nature of academic enquiry and freedom (Martin-Sardesai, Irvine, Tooley, & Guthrie, 2017), but also fail to capture the soul of academic labour and may result in demoralisation (Sutton, 2017), especially for qualitative researchers (Mertkan & Bayrakli, 2018) and early career academics (Smith, 2017). However, impact and the measured university are long-term issues which are unlikely to go away and thus we advocate an engaged approach to try and work with these ideas to shape the agenda and make them work for us, rather than resist them. Our focus on usefulness measures benefits from being simple, easily understood by non-impact experts, comparable, and easy to implement across different projects and disciplines. For example, the online surveys might be administered either via attendance lists at workshops, conferences, or seminars about the research, or via email lists gathered from people who have provided their email in order to download the research report or presentation having heard about it in the press or online. To improve reach and make more people aware of the research, researchers need a media strategy for their research and need to ask their marketing department to help with SEO keywords, to maximise potential interest via internet searches.

As comparability is key for fair assessment, both when judging impact cases internally for promotion, or externally in national impact assessments, the suggested measures deliver on this criterion. The impact index could be useful for comparing between impact cases being considered for which to fund more or which to do a more in-depth impact analysis on or even deciding which to submit to national impact assessments. In particular, it would allow for easy discovery of those people who found the research most beneficial and who would be most likely to cooperate in additional data collection

of these benefits. Our proposal of SROI as a default and common methodology for identifying and measuring context-specific non-comparable outcomes between cases helps in at least being able to understand and compare the methodology by which these outcomes were achieved. This will not only help academics new to impact by providing a ready-made, structured way of deriving these impact measures for their specific projects, but also should give assessors greater confidence in, and understanding of, how these outcomes are derived. The suggested measures might also be used by research centre managers and internal university research assessment managers looking to advise researchers on developing an impact case. In addition, research funding bodies and national assessment organisations might use it as part of their reporting advice and in assessing the outcomes of grants. This would then improve their requirement to provide guidance on how to measure impact in the increasing number of national assessments around the world (VQR, REF, ERA, AERES, ERiC, IOF).

Conclusion and further research

The article responds to calls to develop tools for assisting with impact evaluation (Penfield et al., 2014). If we are unable to demonstrate effectively the non-monetary impact of our research, this will place even more reliance on financial measures and money generated as the prime indicator of engaged research as has happened, for example, in Australia (ARC, 2017). In addition, the lack of readily-available non-monetary measures may seriously underestimate social science's current impact on the world. In tackling this complicated issue, the article contributes by clarifying confusion around impact by suggesting an input, process, output, outcome, and impact framework. Second, it expands the impact concept of 'applicable result' (Mårtensson et al., 2016) to not only propose a simple Net Recommendation Score, but also other outcome measures based on the idea of usefulness across different outcome domains that can be combined to form an impact index to address criticism that there is little standardisation in impact measures. This allows for easier benchmarking and comparative evaluation to occur. Third, for context-specific outcomes, it advocates for the use of a common SROI methodology for assessing these. Fourth, we contribute constructively to 'the measured university' debate by suggesting a productive way forward is to acknowledge the reality of measurement pressures, but also to help academics respond to the challenge of measuring non-economic outcomes in a less confusing, fairer, more achievable and efficient way than is currently the case.

Although our suggestions are a useful starting point, further research might examine how other sectors such as law, accounting, management consulting, and the charity sector efficiently measure non-economic impacts. Further work is also needed into the use of text analysis of social media sites as a measure of usefulness or beneficial outcomes. Like our survey measures, such an analysis tool may even be able to give an automated, quantifiable, and rapid estimate of impact which would not supplant important qualitative evidence, but rather complement it. Finally, as the impact index assumes all items are equally weighted, which is unlikely to be true, data collected at levels one and two could be possibly be re-analysed using factor analysis and structural equation modelling to explore the relationships between these items with a view to establishing which outcome measures best predict overall recommendation. This may give some indication of the relative value of each outcome category.

Note

1. The data should be treated with caution given that this is a function of how the data were calculated by HEFCE, which has not been validated by other sources (HEFCE, 2015b, p. 45).

Acknowledgements

We gratefully acknowledge two anonymous reviewers for their constructive suggestions to improve the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Vincent Mitchell http://orcid.org/0000-0002-6018-9488

References

Abrahamson, E., & Fairchild, G. (1999). Management fashion: Lifecycles, triggers, and collective learning processes. *Administrative Science Quarterly*, *44*(4), 708–740.

Adam, P., Ovseiko, P. V., Grant, J., Graham, K. E., Boukhris, O. F., Dowd, A. M., … Sued, O. (2018). ISRIA statement: Ten-point guidelines for an effective process of research impact assessment. *Health Research Policy and Systems*, *16*(1), 8.

ARC. (2017). *Engagement and impact pilot assessment report 2017*. Retrieved from http://www.arc.gov.au/sites/default/files/filedepot/Public/EI/Engagement_and_Impact_Assessment_Pilot_2017_Report.pdf

ARC. (2018). *Australia Research Council EI submission guidelines*. Retrieved from https://www.arc.gov.au/engagement-and-impact-assessment/ei-key-documents

Association to Advance Collegiate Schools of Business. (2012). *Impact of research. A guide for business schools*. Insights from the AACSB international impact of research exploratory study. Retrieved from http://www.aacsb.edu/~/media/AACSB/Publications/research-reports/impact-of-research-exploratory-study.ashx

Astley, W. G., & Zammuto, R. F. (1992). Organization science, managers, and language games. *Organization Science*, *3*(4), 443–460.

Augier, M., & March, J. (2011). *The roots, rituals, and rhetorics of change: North American business schools after the second world war*. Palo Alto, CA: Stanford University Press.

Beech, N., MacIntosh, R., & MacLean, D. (2010). Dialogues between academics and practitioners: The role of generative dialogic encounters. *Organization Studies*, *31*(9–10), 1341–1367.

Beyer, J. M., & Trice, H. M. (1987). How an organization's rites reveal its culture. *Organizational Dynamics*, *15*(4), 5–24.

Blumler, J. G., & Katz, E. (1974). *The uses of mass communications: Current perspectives on gratifications research*. Beverly Hills, CA: Sage.

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology*, *64*(2), 217–233.

Brewer, J. D. (2011). The impact of impact. *Research Evaluation*, *20*(3), 255–256.

Butler, N., Delaney, H., & Spoelstra, S. (2015). Problematizing 'relevance' in the business school: The case of leadership studies. *British Journal of Management*, *26*(4), 731–744.

Buxton, M. (2011). The payback of 'payback': Challenges in assessing research impact. *Research Evaluation*, *20*(3), 259–260.

Cameron, W. (1963). *Informal sociology: A casual introduction to sociological thinking*. Michigan: Random House.
Christensen, C. M., Raynor, M., & McDonald, R. (2015). What is disruptive innovation? *Harvard Business Review*, *93*(12), 44–53.
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 Education Research & Development*, *36*(3), 555–568.
Czarniawska, B., & Sevon, G. (1996). *Translating organizational change*. Berlin: de Gruyter.
Davis, G. F. (2015). Editorial essay: What is organizational research for? *Administrative Science Quarterly*, *60*(2), 179–188.
Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, *50*(1), 61–104.
de Jong, S., Barker, K., Cox, D., Sveinsdottir, T., & van den Besselaar, P. (2014). Understanding societal impact through productive interactions: ICT research as a case. *Research Evaluation*, *23*(2), 89–102.
Ernø-Kjølhede, E., & Hansson, F. (2011). Measuring research performance during a changing relationship between science and society. *Research Evaluation*, *20*(2), 131–143.
European Commission. (2011). *Estimating the economic value of the benefits provided by the tourism/recreation and employment supported by Natura 2000*. Final Report. Retrieved from http://ec.europa.eu/environment/nature/natura2000/financing/docs/Estimating_economic_value.pdf
Grant, J. (2006). *Measuring the benefits from research*. RAND Europe. Retrieved from http://www.rand.org/pubs/research_briefs/2007/RAND_RB9202.pdf
Grant, J. (2015). *The nature, scale and beneficiaries of research impact: An initial analysis of Research Excellence Framework (REF) 2014 impact case studies* (*HEFCE Report* 2015/01). London.
Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*, *82*(4), 581–629.
Grey, C. (2001). Re-imagining relevance: A response to Starkey and Madan. *British Journal of Management*, *12*(S1), S27–S32.
Groß, C., Heusinkveld, S., & Clark, T. (2015). The active audience? Gurus, management ideas and consumer variability. *British Journal of Management*, *26*(2), 273–291.
Hanney, S. R., & González-Block, M. A. (2011). Editorial. Yes, research can inform health policy; but can we bridge the 'Do-knowing it's been done' gap? *Health Research Policy and Systems*, *9*(23), 1–5.
Higher Education Funding Council for England. (2015a). *REF impact*. Retrieved from http://www.hefce.ac.uk/rsrch/REFimpact
Higher Education Funding Council for England. (2015b). *The nature, scale and beneficiaries of research impact. An initial analysis of Research Excellence Framework (REF) 2014 impact case studies*. King's College London and Digital Science. Retrieved from http://www.hefce.ac.uk/media/HEFCE,2014/Content/Pubs/Independentresearch/2015/Analysis,of,REF,impact/Analysis_of_REF_impact.pdf
Holbrook, J. B., & Frodeman, R. (2011). Peer review and the ex ante assessment of societal impacts. *Research Evaluation*, *20*(3), 239–246.
INTRAC. (2015). *Outputs, outcomes and impact*. Retrieved from https://www.intrac.org/wpcms/wp-content/uploads/2016/06/Monitoring-and-Evaluation-Series-Outcomes-Outputs-and-Impact-7.pdf
Keiningham, T. L., Cooil, B., Andreassen, T. W., & Aksoy, L. (2007). A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing*, *71*(3), 39–51.
Kirkpatrick, D. (1996). Great ideas revisited. *Training & Development*, *50*(1), 54–60.
Lähteenmäki-Smith, K., Hyytinen, K., Kutinlahti, P., & Konttinen, J. (2006). *Research with an impact: Evaluation practises in public research organizations* (VTT Research Notes, 2336). Espoo.

Learmonth, M., Lockett, A., & Dowd, K. (2012). Promoting scholarship that matters: The uselessness of useful research and the usefulness of useless research. *British Journal of Management*, *23* (1), 35–44.

Martin, B., & Tang, P. (2007). *The benefits of publicly funded research* (SPRU Electronic Working Paper Series no.161). Retrieved from https://www.sussex.ac.uk/webteam/gateway/file.php?name=sewp161.pdf&site=25

Martin-Sardesai, A., Irvine, H., Tooley, S., & Guthrie, J. (2017). Government research evaluations and academic freedom: A UK and Australian comparison. *Higher Education Research & Development*, *36*(2), 372–385.

Mason, R. (2015). *The impact of business school research: Economic and social benefits*. Chartered Association of Business Schools. Retrieved from http://charteredabs.org/wp-content/uploads/2015/03/Chartered-ABS-Impact-Case-Studies-Publication-2015.pdf

Mårtensson, P., Fors, U., Wallin, S. B., Zander, U., & Nilsson, G. H. (2016). Evaluating research: A multidisciplinary approach to assessing research practice and quality. *Research Policy*, *45*(3), 593–603.

Meagher, L., Lyall, C., & Nutley, S. (2008). Flows of knowledge, expertise and influence: A method for assessing policy and practice impacts from social science research. *Research Evaluation*, *17*(3), 163–173.

Merkx, F., van der Weijden, I., Oostveen, A. M., van den Besselaar, P., & Spaapen, J. (2007). *Evaluation of research in context: A quick scan of an emerging field*. Den Haag: Rathenau Instituut/ERiC.

Mertkan, S., & Bayrakli, H. (2018). Re-inventing researcher identity: When the individual interacts with the contextual power dynamics. *Higher Education Research & Development*, *37*(2), 316–327.

Millo, Y., & MacKenzie, D. (2009). The usefulness of inaccurate models: Towards an understanding of the emergence of financial risk management. *Accounting, Organizations and Society*, *34*(5), 638–653.

National Innovation and Science Agenda. (2016). *Engagement and impact assessment consultation paper*. Retrieved from http://www.arc.gov.au/sites/default/files/filedepot/Public/ARC/consultation_papers/ARC_Engagement_and_Impact_Consultation_Paper.pdf

NEF. (2008). *Measuring value: A guide to Social Return on Investment (SROI)*. Retrieved from https://commdev.org/wp-content/uploads/2015/06/Measuring-Value-A-Guide-to-Social-Return-on-Investment.pdf

Nicolai, A., & Seidl, D. (2010). That's relevant! Different forms of practical relevance in management science. *Organization Studies*, *31*(9–10), 1257–1285.

Nicolai, A. T., & Dautwiz, J. M. (2010). Fuzziness in action: What consequences has the linguistic ambiguity of the core competence concept for organizational usage? *British Journal of Management*, *21*(4), 874–888.

OECD. (2010). *Glossary of key terms in evaluations and results based management*. Retrieved from http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

Ozanne, J. L., Davis, B., Murray, J. B., Grier, S., Benmecheddal, A., Downey, H., … Seregina, A. (2016). Assessing the societal impact of research: The relational engagement approach. *Journal of Public Policy & Marketing*, *36*(1), 1–14.

Parks, S., Ioppolo, B., Stepanek, M., & Gunashekar, S. (2018). *Guidance for standardising quantitative indicators of impact within REF case studies*. Retrieved from https://www.ref.ac.uk/media/ref,2021/downloads/Guidance%20for%20standardising%20quantitative%20indicators%20of%20impact.pdf

Pearce, J. L., & Huang, L. (2012). The decreasing value of our research to management education. *Academy of Management Learning & Education*, *11*(2), 247–262.

Penfield, T., Baker, M. J., Scoble, R., & Wykes, M. C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*, *23*(1), 21–32.

Peseta, T., Barrie, S., & McLean, J. (2017). Academic life in the measured university: Pleasures, paradoxes and politics. *Higher Education Research & Development*, *36*(3), 453–457.

REF. (2014). *Panel criteria and working methods*. Retrieved from http://www.ref.ac.uk/media/ref/content/pub/panelcriteriaandworkingmethods/01_12.pdf

Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, *28*(1), 11–31.
Rogers, E. M. (1995). *Diffusion of innovations*. New York, NY: Free Press.
Rymer, L. (2011). *Measuring the impact of research – The context for metric development*. The Group of Eight. Retrieved from https://go8.edu.au/sites/default/files/docs/go8backgrounder23_measimpactresearch.pdf
Schulz, A. C., & Nicolai, A. T. (2015). The intellectual link between management research and popularization media: A bibliometric analysis of the *Harvard Business Review*. *Academy of Management Learning & Education*, *14*(1), 31–49.
Scoble, R., Dickson, K., Hanney, S., & Rodgers, G. J. (2010). Institutional strategies for capturing socio-economic impact of academic research. *Journal of Higher Education Policy and Management*, *32*(5), 499–510.
Seidl, D. (2007). General strategy concepts and the ecology of strategy discourses: A systemic-discursive perspective. *Organization Studies*, *28*(2), 197–218.
Smith, J. (2017). Target-setting, early-career academic identities and the measurement culture of UK higher education. *Higher Education Research & Development*, *36*(3), 597–611.
Smith, S., Ward, V., & House, A. (2011). 'Impact' in the proposals for the UK's research excellence framework: Shifting the boundaries of academic autonomy. *Research Policy*, *40*(10), 1369–1379.
SROI Network. (2012). *A guide to social return on investment*. Retrieved from http://socialvalueuk.org/
Starkey, K. A., & Madan, P. (2001). Bridging the relevance gap: Aligning stakeholders in the future of management research. *British Journal of Management*, *12*, S3–S26.
Starkey, K., Hatchuel, A., & Tempest, S. (2009). Management research and the new logics of discovery and engagement. *Journal of Management Studies*, *46*, 547–558.
Sutton, P. (2017). Lost souls? The demoralization of academic labour in the measured university. *Higher Education Research & Development*, *36*(3), 625–636.
Technopolis. (2014). *Measuring scientific performance for improved policy making*. European Parliamentary Research Service. Retrieved from www.europarl.europa.eu/RegData/etudes/etudes/join/2014/527383/IPOL-JOIN_ET(2014)527383(SUM01)_EN.pdf
Wensley, R. (2007). *Beyond rigour and relevance: The underlying nature of both business schools and management research* (Advanced Institute of Management Research Paper, 051). London: Advanced Institute of Management.
Wilhelm, H., & Bort, S. (2013). How managers talk about their consumption of popular management concepts: Identity, rules and situations. *British Journal of Management*, *24*(3), 428–444.