Cover page (not anonymised)

Title: Obstacles to co-producing evaluation knowledge; power, control and voluntary sector dynamics

Author name(s) and affiliation(s):

Louise Warwick-Booth
Leeds Beckett University, Centre for Health Promotion Research

Ruth Cross
Leeds Beckett University, Centre for Health Promotion Research

James Woodall
Leeds Beckett University, Centre for Health Promotion Research

Abstract:

Background

Despite literature recognising the huge potential of co-production as a positive approach to evidence creation, there is a dearth of evidence about how co-production principles can problematise knowledge exchange, specifically in evaluation work.

Aims

To critically examine 3 evaluation projects commissioned by voluntary sector stakeholders to illustrate challenges in knowledge exchange linked to the co-production of evidence exchange.

Methods

We critically compare the challenges experienced in co-producing evidence across 3 evaluations, reflecting on power dynamics, co-productive ways of working and emotions, which all impact upon successful knowledge exchange.

Findings

In project 1, internal monitoring data required for reporting was not shared. In project 2, the commissioners' need to evidence success resulted in limited knowledge sharing, with valuable learning about partnership issues and service delivery held internally. In project 3, evidence demonstrating the failure of a local authority model of area management for community members was partially discredited by statutory stakeholders (state actors).

Discussion and conclusions

Bias in evaluation reporting and academic publication can arise from current knowledge exchange processes, including co-production. Voluntary sector funding is problematic as stakeholders delivering programmes also commission evaluations. Knowledge exchange is influenced by vested interests arising from the political context in which data is gathered. Evaluators can face aggression, challenge and unfair treatment resulting in damaged relationships, and failures in knowledge exchange. The emotional elements of knowledge exchange remain under-reported. Varying and shifting power dynamics also limit knowledge exchange. Changing research practice, to support power sharing needs further exploration to facilitate improved knowledge exchange.

Key messages (if applicable):

1. Knowledge exchange is influenced by vested interests, power dynamics and the political context in which data gathering occurs.

2. Evaluation and associated knowledge exchange is emotional, which is under-reported in the academic literature.

3. Co-production principles can problematise knowledge exchange.

Key words/short phrases:

Co-production, evaluation, voluntary sector, power, politics, emotions

JEL codes (if applicable):

N/A

Word count:

7,040

Introduction

In recent years, universities have been tasked with considering the impact of their research, especially in terms of how they exchange research knowledge with relevant users, within wider society. Halsall & Powell (2016) define knowledge exchange as a process in which people and institutions share ideas, learning and experiences, with this definition used by the authors in relation to the evaluations discussed in this paper. This is described as a two-way process by funding bodies such as the ESRC (2022). There are however multiple conceptualisations of the process of knowledge exchange, and an evidence base tending to focus on implementation as rational and logical, despite this being a social process taking place in complex environments (Ward et al, 2012). Context is key, but it is a challenging concept narrowly defined in the literature which focuses on individuals rather than organisations. Knowledge exchange is also seen as an implementation problem in a wide range of studies, rather than a social and political problem (Ward et al 2012). For Ward et al (2012), it is grounded in dynamic processes as much as being reflected in outputs. They therefore argue that knowledge exchange processes need to be revised as the research work evolves, to support changes in practice. Davies et al (2008) also note that no single source of knowledge can provide definitive answers, as findings can be contradictory instead suggesting that the term knowledge interaction is more appropriate as it reflects uncertainties, complexities and contextual ways of knowing.

Diez-Vial & Montoro-Sanchez (2014) discuss the success of contextual knowledge exchange within specific geographical spaces and places. Their research findings highlight the importance of social capital in knowledge exchange particularly in terms of relationships (trust and friendliness) as well as shared culture. The importance of positive interpersonal relations within place-based knowledge exchange processes is also noted in the literature (Wain et al, 2021: ESRC, 2022). Place however remains a problematic concept as a driver of knowledge exchange (Wain et al, 2021). Despite definitional challenges, Wain et al (2021) detail a range of local place -based knowledge exchange impacts including policy changes, co-created research, the creation of learning environments and skills development. However, there is a limited evidence base about knowledge exchange, specifically in the voluntary and community sector, hence the focus of this paper. Hardwick et al (2015) note that VCSE's (Voluntary Community and Social Enterprise) can be evidence-informed in their decision making, but they experience barriers to exchanging research due to lack of time, staff skill, resource challenges and the acontextual nature of some academic data collection. Mann (2019) also discusses tensions in VCSE research between the types of outcomes that funders want to measure, and the slower, less quantifiable progress that happens in community-based interventions, often referred to as soft(er) outcomes. Furthermore, knowledge exchange involving co-production approaches, remains underexplored in the academic literature.

Despite a lack of an agreed definition or concept of co-production in the wider literature (Brandsen and Honingh, 2015), and even what some describe as 'confusion' (Nabatchi et al., 2017), there is general consensus about what co-production comprises of. Co-production is underpinned by a set of values that are realised in working together, working in partnership, and recognising that everyone has a part to play and has expertise, therefore the authors work to apply these values in research projects. Co-production is underpinned by reciprocal relationships within the research process which support all involved to work together towards a mutually agreed outcome (Marshall et al., 2019; Nabatchi et al., 2017; Needham and Carr, 2009).

Co-production is a way of working that has been championed in the health and social care services to involve people more actively in service delivery and is intended to promote service users as partners in care (NHS England and C4PC, 2020). A key principle of co-production is that people work together as equals. Marshall et al. (2019) outline five key features of a co-production model

developed for use in the NHS: ownership, understanding and support of co-production by all; a culture of openness and honesty; a commitment to sharing power and decisions; clear communication in plain English; and a culture in which people are valued and respected.

Nabatchi et al's (2017) 3x4 typology of co-production has 3 *levels* relating to *who* is involved in the co-production activity – individual, group, and collective; and four *phases* relating to *when* it is taking place – commissioning, design, delivery and assessment. Using this typology, we can situate the evaluation work discussed here at the 'group' level and in the 'design' and 'assessment' phases. The typology defines both parties in the group (evaluators and evaluated) as 'state' actors, given that they are 'serving in a professional capacity' and recognises that lay actors (members of the public) are also involved in co-production (Nabatchi et al, 2017: 769). This typology was however, developed from involvement in public services, rather than evaluation per se, and there remains a relative dearth of literature about co-production in evaluation.

Co-production has advantages to those who engage in it and is assumed to be a positive, beneficial process for all involved (Warwick-Booth et al., 2022) however, co-production in the context of evaluation brings specific challenges in particular, power-sharing. Evaluators might be said to hold the balance of power given that they are, essentially, assessing a service. Yet, the evaluated could also be said to hold the balance of power as they have oversight and specific knowledge about the project being evaluated and can withhold or block access to certain information or stakeholders (Bechar and Mero-Jaffe, 2014). The evaluated may hold money as the research client and need to ensure further income generation for their organisational survival. It is therefore useful to conceive of power in the evaluation dynamic as a 'tug of war.' Power moves back and forth between the evaluator and the evaluated at different points in the process.

One of the key components of co-production is trust (Needham and Carr, 2009) however, there will inevitably always be an element of reticence when one party (the evaluated) is being 'measured' by another (the evaluator). In addition, some of the key facets of co-production may not be possible given the relationship between the evaluated and the evaluator – for example, peer support and developing truly collaborative relationships.

The idealised principle of 'shared ownership' in co-production is potentially problematic when evaluations are led by people external to the project. Vested interests may differ depending on what is at stake for both parties. The most successful co-production occurs when everyone feels secure and has a sense of belonging or shared purpose (Needham and Carr, 2009). However, this not always the case in evaluation contexts as will be discussed. Despite power-sharing being a central ideal in co-production this is difficult to realise in practice in the context of evaluation. Questions are therefore introduced regarding issues of power – who holds it? who can exercise it? Often the evaluator is at the behest of the organisation or powerful, senior stakeholders (state actors) within it, which sits within Laverack's (2007) notion of 'power-over' which is about one person or group having control over another. In addition, Laverack's (2007) discusses power-with, where public health practitioners share their power, and power-within, where self-confidence and empowerment are held internally at an individual level.

A commissioning model of funding sets the context for the relatively small-scale VCSE projects that we have evaluated. Some VSCEs can and do self-evaluate, and toolkits exist to support this. Indeed, the literature suggests an increasing number of 'in-house' evaluations conducted by individuals on their own organisation (Pattyn and Brans, 2013). That said, there is, as described by Mann (2019), a drive towards external accountability for such projects, in an increasingly challenging and competitive funding environment. Thus, with a political backdrop that promotes short-termism

against the aftershocks of austerity, those being evaluated are often genuinely worried about keeping their jobs and about project sustainability. This is particularly the case for third sector and voluntary organisations that might lurch from one funding stream to the next with little continuity and the constant threat of extinction. There is a lack of long-term security for staff or for the projects themselves (and therefore the clients who access that service provision and support), contributing to pressure to conform to research language of outcomes and evidence even when working with marginalised community members (Mann, 2019). Consequently, all evaluations are political, some organisations do not like their projects being evaluated and findings themselves may also be politicised (Hall, 2019). The challenges of accounting for both stakeholder interests and political sensitivities are under-appreciated in small scale evaluation research (Bryson et al, 2011). All of these issues can contribute to tensions, mistrust and challenges in the relationships that are needed for successful knowledge exchange.

Interventions

This paper reports on data generated from 3 evaluation studies, across different VCSE interventions in the north of England. These interventions were selected because they constitute the most challenging evaluations experienced in over a decade of such work, and in line with Hall's (2019) position, conflict is important information that needs to be evidenced within evaluation work, yet it remains underreported. Table 1 provides contextual detail about each of the 3 interventions, detailing their intended aims, and delivery models.

<Insert table 1 here>

Table 1 - Overview of each VCSE intervention

Methods

Although the specific nature of the interventions differs as illustrated in table 1, the evaluation approaches utilised across all 3 intended to qualitatively explore these interventions and provide recommendations for improvements. Each evaluation drew upon co-production principles, attempting to share power and responsibility with stakeholders during the research process. This meant tailoring the methodological design and data collection tools towards stakeholder requirements (evaluation 1, and 2), as well as working transparently to discuss research processes, analysis, findings, and reporting (evaluation 3). Co-production differs in each instance here due to the adaptation of principles, levels of stakeholder involvement, and varying power dynamics associated with commissioners, delivery partners, service users and community members. In each of these 3 evaluations, co-producing knowledge was intended to facilitate knowledge exchange. Table 2 provides a comparative summary of methodology used across each evaluation, illustrating the data underlying each evaluation. However, this paper is drawing upon the final evaluation reports, that were generated from these data.

<Insert table 2 here>

The final evaluation reports upon which this work is based are available as follows https://eprints.leedsbeckett.ac.uk/id/eprint/1930/, https://eprints.leedsbeckett.ac.uk/id/eprint/7910/ and https://eprints.leedsbeckett.ac.uk/id/eprint/8286/.

Sampling

Each evaluation used purposive sampling, attempting to select knowledgeable stakeholders (Creswell & Plano Clark, 2011). Those sampled had experience of delivering the interventions, referring into them, using them or were community members affected by the intervention. Evaluators are challenged with recruitment as not all potential stakeholders are willing or able to participate and reaching those who decide not to engage with interventions is an ever present issue (Warwick-Booth & Coan, 2022). However, involving stakeholders in evaluation research as participants remains crucial for learning (Bryson et al, 2011). In co-producing knowledge, we relied upon gatekeepers to supply us with details of those who consented to participate and were never granted access to internal records which would have provided us with a fuller sampling frame to draw from. Commissioners across all of these evaluations also limited access to other internal data, as discussed later. Table 3 provides an overview of the complete samples for each evaluation.

<Insert table 3 here>

Table 3 – summary of sampling across the interventions

In each of the intervention data sets, the evaluation team were not fully cognisant with what was missing data-wise from each project.

Analysis

Evaluation data were analysed using one of two approaches depending on the requirements of each evaluation. We used Framework Analysis (project 1 and 2) or Thematic Analysis (project 3). In all instances involving primary data collection, focus group discussions and individual interviews were recorded (with consent from all participants) then transcribed verbatim. In project 2, one participant asked for their interview not to be recorded, so notes were made instead.

We used framework analysis to develop a hierarchical scheme to classify and organise the data that we had gathered according to each theory of change, and evaluation aims. We also documented emergent categories. Frameworks are an analytical matrix to organise themes, patterns, and connections across data sets (Ritchie, Spencer & O'Connor, 2003). Thematic analysis (Braun & Clarke, 2006) was used in project 3 to identify, analyse and report patterns (themes) within data sets from community events, and open text supplied via questionnaires. Descriptive statistics were also used to analyse questionnaire data in project 3. Across all 3 evaluations, themes were always agreed by members of the research team, who worked together to reach consensus on reporting of the findings.

In this paper, further analysis of co-productive knowledge exchange is provided by comparing the processes experienced across the evaluations (table 4), examining the political context in which the evaluations took place (table 5) and applying researcher reflexivity to our experiences through using Marshall et al's (2019) framework (table 6).

Ethical Approval

Each evaluation received ethics approval through university procedures and adhered to standard practices such as gaining informed consent from all participants involved in both interviews and focus groups. Where participants were below the age of consent (project 1), parental assent was obtained. Attention was paid to risk management, with careful monitoring of emotions during data collection, and debrief strategies in place for participants and researchers alike. No personal information was used in data reporting, to ensure confidentiality as well as anonymity by use of labels and pseudonyms. However, given the small-scale nature of these evaluations, some participants (service users and workers alike) would have been known to professionals involved in

delivery. In project 2, one senior staff member withdrew consent for an interview to be included in the evaluation analysis, due to in-depth discussion within it related to the partnership, and associated concern about comments being identifiable. Finally, in terms of secure information management, security was maintained by password protected data management systems, and pseudonymisation of all internal data provided by stakeholders, prior to sharing it with evaluators. Detailed information about organisations has been replaced with general description in this paper, due to the small scale nature of the evaluations and the need to ensure anonymity in reporting. The authors have considered the ethics of reporting these negative experiences and contend that given the many ethical challenges that qualitative research raises it is important to offer analysis of these and to develop guidelines to enhance practice (Sanjari et al, 2014) especially in studies using co-production.

Findings

Each evaluation resulted in a detailed final report being both produced, and made available to funders, workers, service users and the general public. As evaluation outputs, these reports illustrate successful intervention delivery in two cases, alongside some issues for consideration (1 and 2), and in evaluation 3, a more complex and challenging picture of the intervention impact was co-created. Reporting is one component of knowledge exchange. Using co-production principles means that knowledge exchange varied across the interventions being compared. Dissemination of evaluation findings and knowledge exchange was included as part of our proposals but was to be more firmly negotiated with commissioners in each instance. Table 4 outlines the different forms of knowledge exchange that were successful in each project, but also summarises the challenges which restricted full evidence disclosure and therefore limited knowledge exchange.

<Insert table 4 here>

Table 4 – Comparison of knowledge exchange across the interventions

Table 4 illustrates that final evaluation reports and accompanying knowledge exchange activities do not include all evidence, with some level of nondisclosure present in each instance. In projects 1, and 2 commissioners withheld internal monitoring data instead requesting the evaluation team to collect primary data. Stakeholders chose to share only a limited sample of their internal data. Co-creation also impacted upon full disclosure in each evaluation. Earlier described as tugs of war, power dynamics shifted and changed over the course of the work, sometimes working in favour of the evaluation team, and at other times leading to challenges and disagreements. Despite commissioners voicing their commitment to the value of the independence brought by the evaluation team in each instance, there were attempts to control and shape the final reporting by stakeholders who requested that some content was not included, or that findings be reported in specific ways, reflecting vested interests at play. Whilst vested interests in each evaluation context were different, each political context served to influence and limit knowledge exchange (see table 5), though this is not evident in the final evaluation products.

<Insert table 5 here>

Evaluators' reflections

The interventions' discussed here offered challenge in terms of knowledge exchange, and co-production, as detailed in table 6, which applies Marshall et al.'s (2019) framework of co-productive ways of working to the 3 evaluation projects, considered in this paper.

<Insert table 6 here>

Table 6 – reflections on co-production

As illustrated in table 6, emotions were evident in each evaluation context, for all of those involved. Emotions can themselves be seen as evaluation evidence and are determined by contextual and political influences. At various points, the evaluators experienced frustration, especially where communication challenges arose, and requests were unheard. Evaluators also experienced upset, when challenged aggressively in meetings (evaluation 1), during data collection (evaluation 2) and by stakeholders (all evaluations). These feelings created additional emotional labour for the evaluation team (Hall, 2019), above and beyond that arising from qualitative data collection exploring sensitive issues (Warwick-Booth, et al 2023), which can be overwhelming and burdensome in itself for all involved, including stakeholders and service users.

Discussion

In summary the findings reported here show that in project 1, internal monitoring data required for reporting was not shared. In project 2, the commissioners' need to evidence success resulted in limited knowledge sharing, with valuable learning about partnership issues and service delivery held internally. In project 3, evidence demonstrating the failure of a local authority model of area management for community members was partially discredited by statutory stakeholders (state actors). This paper is biased towards illustrating challenges in co-production and associated evidence exchange, we note that we have had many positive relationships with commissioners, and as evaluators we have been able to produce qualitative products that are fair, balanced and empirically sound, which were subsequently peer-reviewed and published. Unlike Hall's (2019) self-professed naivety we, as an evaluation team, are acutely aware of the political context in which we work. Research suggests that evidence-production can be less than neutral and, on a practical level, external evaluation of programmes – especially in the voluntary and community sector where funding is very time-limited – can cause anxiety and fear for jobs and sustainability (Moretti, 2021). We appreciate that there is a lot at stake for small projects in terms of funding- and staffing-sustainability and that decisions can be made based on evaluation outputs which puts significant pressure not just on the state actors, but on the evaluation team itself. In the type of work we evaluate, the people working in the projects are highly invested, passionate about what they do and the services, and people, that they support. They are emotionally charged and therefore are not neutral participants in their role, or in the evaluation process. They want the project to succeed and to be shown in the best light. This might be down to a number or combination of factors – concern for the clients, personal reputation, or even security of employment. Indeed, researchers bring their own positionality and values to their practice, and therefore do not occupy neutral positions (Warwick-Booth et al, 2023). Consequently, as we outlined in the findings, emotions abound and should therefore be considered as evaluation evidence (Hall, 2019 & Mann, 2019) and more specifically in relation to knowledge exchange processes. Emotional labour is underreported in the literature on knowledge exchange, yet it is important due to vested interests, and what Hall (2019: 161) refers to as 'emotional magnitude'. Despite our own emotional labour, we managed emotional reactions through peer support between ourselves and communicated with commissioners in a professional manner most often via email. On reflection, communication could have been better managed, if this were conducted in person, to agree expectations, and next steps.

Furthermore, if we reflect on Marshall et al.'s (2019) five key features outlined earlier in this paper the co-productive nature of the ways of working in the 3 evaluation projects we are considering in this paper can be called into question. Firstly, ownership, understanding and support of co-production by all was questionable depending on the specific context. Secondly, a culture of openness and honesty was not always experienced by the evaluation teams. Thirdly, a commitment

to sharing power and decisions was not consistently seen. Fourthly, at times, clear communication in plain English did not happen and finally, and a culture in which people are valued and respected was not necessarily promoted. We recognise, however, that there will be varying truths, perspectives and experiences for the other parties involved in the evaluation projects and further work should seek these viewpoints. Given the issues experienced the authors limited communication with commissioners following final reporting. The authors were contractually bound to seek permission for publications, and sent emails in relation to this, receiving agreement from project 1 and project 2. No requests were made to project 3. Power was a tug of war and is a useful metaphor for analysing the evaluation projects discussed here. In addition to Laverack's (2007) interpretation of power as power-within, power-over and power, the authors suggest another interpretation of power – 'power-holding', whereby one party holds (back) pertinent information that would be of value to the other party's endeavours. This power-holding serves to limit full co-production, as well as knowledge exchange. The literature recognises many different types of co-production (Brandsen and Honingh 2015), however there does not seem to be a type described that explicitly analyses co-production between evaluator and evaluated, exploring knowledge exchange limitations arising in such contexts. Whilst the literature also suggests that boundary negotiation is needed in the co-creation of research evidence, the practical and emotional challenges of this in reality remain under explored (Nicholas et al, 2019). Knowledge exchange processes are underpinned by risk, and risk management is therefore evident in relation to reputations, relationships, and compromises in terms of which findings are presented publicly and all that remain hidden.

Conclusion and recommendations

Recommendations for research

- Future studies are needed to gather detail on how changing research practices can enhance power-sharing and facilitate more transparent and honest knowledge exchange.

Recommendations for practice

- Some of issues raised in this paper related to knowledge production could be minimised by universities implementing stronger contracting processes which make clear the independence of the evaluators and their requirement to report openly and fairly as academics. Terms of reference specifically relating to expectations could also be included in university documents for commissioners.

- Knowledge exchange can be supported by the creation of opportunities for the evaluated and evaluators to share concerns openly and honestly in safe spaces, with mediation available if required. In person communication, focused on negotiating solutions is needed to support difficult conversations. Email messages can exacerbate frustrations when sent quickly, or if their content is misunderstood.

- Evaluators need to pay attention to emotional risk management when co-producing knowledge. Considerate reporting and paying attention to semantics may reduce upset, frustration and stress for those being evaluated.

- Evaluation researchers should also have access to independent support enabling them to discuss and manage any emotional labour associated with their research practices. Space for

respective and constructive debrief can be created through the use of models based upon clinical supervision and peer support.

This paper documents learning from 3 co-produced evaluation projects funded via voluntary sector commissioners, exploring the ways in which evaluation knowledge production was challenged and indeed limited in a variety of ways. Challenges underpinning transparent and effective knowledge exchange result from funding pressures, vested interests, and the dynamics associated with stakeholder power in each context. Various compromises result in terms of the methods used, sample access, evidence presentation, reporting and knowledge exchange. The emotional elements of problematic knowledge exchange processes are evident here too, for all involved. Co-production principles can problematise knowledge exchange. Voluntary sector funding currently serves to create a dynamic and socially driven knowledge exchange process, highly at risk of challenge, bias and failure especially when co-production principles are used in evidence generation.

Research ethics statement

Research ethics approval was sought and granted via Leeds Beckett University procedures, for all 3 evaluations. Evaluation 1 was approved on 09.04.13. Evaluation 2 received approval on 16.10.17, and evaluation 3 was granted approval on 12.12.19.

Funding details

All 3 evaluations were funded by VCSE organisations, rather than research council grants, and were subject to standard university contracting procedures.

Contributor statement

LWB, RC and JW all designed and developed the manuscript. LWB wrote the literature section, and findings, RC wrote the discussion, and inputted into recommendations for practice. JW provided critical review, guidance on relevant literature, and inputted into the recommendations for practice.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Acknowledgements

The authors would like to acknowledge Susan Coan, for her support with data collection (evaluations 1 and 2), analysis and reporting in all 3 evaluations. Susan also acted as a critical friend, and commented on draft versions of this paper.

References

Bechar, S. & Mero-Jaffe, I. (2014) 'Who is afraid of evaluation? Ethics in evaluation research as a way to cope with excessive evaluation anxiety: insights from a case study.' *American Journal of Evaluation,* 35**:** 364-376.

Brandsen, T. and Honingh, M. (2015) 'Distinguishing Different Types of Co-production: A Conceptual Analysis Based on the Classical Definitions.' *Public Administration Review,* 76 (3): 427-435.

Braun, V. & Clark, V. (2006) 'Using thematic analysis in psychology' *Qualitative Research in Psychology,* 3 (2):77-101.

Bryson, J.M., Quinn Patton, M. & Bowman, R.A. (2011) 'Working with evaluation stakeholders: A rationale, step-wise approach and toolkit' *Evaluation and Program Planning,* 34: 1-14.

Creswell, J.W. & Plano Clark, V.L. (2011) *Designing and conducting mixed methods research,* 2nd Edition. California, Thousand Oaks, Sage.

Cross, R. & Warwick-Booth, L. (2015) 'Using storyboards in participatory research.' *Nurse Researcher,* 23 (3): 8-12.

Davies, H., Nutley, S. & Walter, I. (2008) 'Why 'knowledge transfer' is misconceived for applied social research' *Journal of Health Service Research,* 13 (3)188-190.

Diez-Vial, I. & Montoro-Sanchez, A. (2014) 'Social capital as a driver of local knowledge exchange: A social network analysis' Knowledge *Management Research and Practice,* 12: 276-288.

ESRC (2022) *How to do effective knowledge exchange* available at https://www.ukri.org/councils/esrc/impact-toolkit-for-economic-and-social-sciences/how-to-do-effective-knowledge-exchange/developing-relationships/#contents-list (accessed 6.12.22)

Hall, A.C. (2019) 'Evaluations that fail: nasty emails, small samples and tenuous futures' *Evidence & Policy,* 15 (1): 161-171.

Halsall, J.P. & Powell, J. (2016) 'Crafting knowledge exchange in the social science agenda' *Cogent Social Sciences* 2 (1): 1244145, DOI: 10.1080/23311886.2016.1244145.

Hardwick, R., Anderson, R. & Cooper, C. (2015) 'How do third sector organisations use research and other knowledge? A systematic scoping review' *Implementation Science.* 10: 84, DOI 10.1186/s13012-015-0265-6.

Hayden, M.C., Weib, M., Pechriggi, A. & Wutti, D. (2018) 'Insights into university knowledge transfer in the social sciences and humanities (SSH) and other scientific disciplines – more similarities than differences' *Front. Res, Metr. Anal*, 3 (32) https://doi.org/10.3389/frma.2018.00032.

Laverack, G. (2007) *Health Promotion Practice: Building Empowered Communities.* Maidenhead, Open University Press.

Mann, S. (2019) 'Making it count: learning from a knowledge exchange with voluntary sector workers and software providers, and data from auto/biographical reflection as an activist academic' *Methodological Innovations,* 11: 1-9.

Marshall, C., Zambeaux, A., Ainley, E., McNally, D., King, J., Wolfenden, L. and Lee, H. (2019) 'NHS England Always Events® program: Developing a national model for co-production' Patient *Experience Journal,* 6 (1): 154-165.

Moretti, E. (2021) 'Navigating the awkward, challenging social context of evaluation.' *Evaluation Journal of Australasia,* 21**:** 163-176.

Nabatchi, T., Sancino, A. and Sicilia, M. (2017) 'Varieties of Participation in Public Services: The Who, When, and What of Co-production' *Public Administration Review,* 77 (5): 766-776.

Needham, C. and Carr, S. (2009) *SCIE research briefing 31: co-production: an emerging evidence base for adult social care transformation, 2009.* Available online at: https://www.scie.org.uk/publications/briefings/briefing31/ [Accessed 9th December 2022]

NHS England and C4PC (2020) *A Co-production Model.* 11 October 2020. Available online at: coalitionforpersonalisedcare.org.uk/resources/a-co-production-model/ [Accessed 9th December 2022].

Nicholas, G., Foote, J., Kainz, K., Midgley, G., Prager, K. & Zurbriggen, C. (2019) 'Towards a heart and soul for co-creative research practice: a systematic approach' *Evidence and Policy*, 15 (3) 353-370.

Pattyn, V. & Brans, M. (2013) 'Outsource versus in-house? An identification of organizational conditions influencing the choice for internal or external evaluators.' *Canadian Journal of Program Evaluation,* 28: 43-63.

Ritchie, J., Spencer, L., & O'Connor, W. (2003) 'Carrying out qualitative analysis' in J. Ritchie and J. Lewis (Eds.), *Qualitative research practice: a guide for social scientist students and researchers* (pp. 219-262). London, UK: Sage.

Sanjari, M., Bahramnezhad, F., Fomani, F.K., Shoghi, M. & Cheraghi, M.A (2014) 'Ethical challenges of researchers in qualitative studies: the necessity to develop a specific guideline' *Journal of Medical Ethics and History of Medicine* 7: 14.

Wain, M., Rosemberg, C., Farla, K., Seth, V., Maki, N., Dijkstal, F., Sutinen, L. & D'hont, J. (2021) *Knowledge exchange and place: A review of the literature*. *Final Report* Technopolis Group available at https://www.ukri.org/wp-content/uploads/2022/05/RE-270522-KnowledgeExchangePlaceReviewLiteratureFinalReport.pdf (accessed 6.12.22).

Ward, V., Smith, S.O., House, A., & Hamer, S. (2012) 'Exploring knowledge exchange: a useful framework for practice and policy' *Social Science and Medicine*, 74 (3): 297-304.

Warwick-Booth, L., Cross, R. and Coan, S. (2022) 'You're not going to give a monkey's chuff': exploring co-production in the design of services for women who have experienced sexual violence' *Perspectives in Public Health* 142 (4): 231-236.

Warwick-Booth, L. & Coan, S. (2022) 'Lessons learned from a gender-specific educational programme supporting young women with experience of domestic abuse' *Health Education Journal* https://doi.org/10.1177/00178969221129110.

Warwick-Booth, L., Cross, R., Coan, S., (2023 in press) 'The practices of feminist stakeholders supporting social change and health improvement for women experiencing domestic abuse in England' in Jourdon, D. & Potvin, L. (Eds) *Global Handbook on health promotion research. Doing Health Promotion Research Vol 3* New York, Springer.

Table 1 - Overview of each VCSE intervention

Project	**Intervention**
Evaluation 1 2013-2015	This women-only, VCSE project was aimed at young women aged 13-18 years, with unmet needs, who were deemed to be slipping between existing offers of service provision. The work focused on early intervention to prevent them from entering adulthood with severe and escalating levels of disadvantage. Project delivery involved a key worker giving comprehensive support to women on their case load. The work was supported by a steering group of workers from a range of other agencies, which met quarterly. One VCSE organisation delivered the work.

Evaluation 2 2017-2021	This intervention to support marginalised and vulnerable women was based in one city, using specialist support workers, to deal with women's complex needs. The work was supported by a partnership of VCSE delivery organisations working as a new consortium for the period of funding. The partnership offered frontline services to support women and girls to lead safer and healthier lives, information sharing and various opportunities for women to be involved in the work. The focus of the work was on providing support to the most disadvantaged communities in one city, with the aim of reaching greater numbers of the most vulnerable women, ensuring that they received holistic, joined-up support.
Evaluation 3 2019-2020	In 2014, a Local Authority implemented a Managed Zone/Approach to street sex work, attempting to create a safe area, without attempts to prosecute those involved. A partnership of statutory and VCSE agencies delivered this work.

Table 2 – summary of methods used in each evaluation

Intervention	**Research Design**	**Methods**
Evaluation 1	The evaluation team responded to a tender for the work, proposing an action research approach. This was adapted into a standard qualitative study, to assess the outcomes of the intervention.	The evaluation ran 2013-2015, using a co-designed theory of change. We were commissioned to evaluate the project from its inception, and to attend steering groups throughout the lifetime of the intervention. Individual semi-structured interviews captured staff views. Focus groups with young women receiving support, used creative methods to support them in storytelling (Cross & Warwick-Booth, 2015). All data collection followed traditional qualitative conventions, guided by a semi-structured schedule. Some internal qualitative monitoring data was also provided to the evaluators.
Evaluation 2	The evaluation team responded to a tender for the work, which requested a qualitative evaluation design. The proposal included an offer of a range of qualitative tools	The evaluation ran 2017-2020 (data collection was completed before Covid19 restrictions). Funders developed their own theory of change.

	including focus groups, interviews, creative methods and peer research. We reported on both process and outcomes.	We were commissioned to evaluate the project from its inception, and to annually report findings. Internal monitoring was implemented by staff, using a specifically designed outcomes tool. We used focus groups and/or individual interviews with service users, and observed partnership meetings.
Evaluation 3	The evaluation team were approached by the VCSE organisation and asked to provide support with analysis of qualitative data that had already been collected. The evaluation team had no role in the research design.	This evaluation ran 2019-2020 and used structured community events in local venues to generate data by gathering perspectives about the intervention. Volunteers collated notes and key points, and one event was recorded (each table discussion was a voice file). Much of the data gathering was completed prior to the evaluation team involvement. Evaluators attended, observed, and participated in one community event.

Table 3 – summary of sampling across the interventions

Intervention	**Sampling**
Evaluation 1	• Anonymised internal monitoring data included women's demographic characteristics, the engagement details of the young women, referral data, 11 brief case studies and 12 detailed case notes from worker records. • 2 focus group discussions with young women accessing the project (n=13). • 2 focus groups with staff. One captured learning from referrers (n=3), and 2 referrers unable to attend were interviewed separately. The other focus group documented steering group member perspectives (n=4). • Key workers were also interviewed (n=2).

Evaluation 2	• An internal review of delivery after one year of funding involved 4 focus group discussions, segmented according to staff roles (n= 28). • Service user interviews and/or focus groups (n=34). • 25 anonymised service users exit forms (project workers completed these). • 54 professional interviews/focus groups (19 repeated to capture learning over the intervention delivery period). • Anonymised information and learning collated through quarterly monitoring and reporting (n= 252 services users).
Evaluation 3	• Hand-written notebooks completed at some events. • Hand-written structured templates completed at some events. • Submissions from community members who were unable to attend events (both in paper form, and online). • Voice files from mobile phone recordings taken at one event. • Hand-written paper-based questionnaires from two school-based surveys (N=45 in one sample, and n=227 in the other).

Table 4 – Comparison of knowledge exchange across the interventions

Project	**Successful knowledge exchange in the public domain**	**Challenges to knowledge exchange, hidden from public view**
Evaluation 1	Evaluation report (branded with university logo) Evaluation summary Event held at the university to present findings, to showcase service user voice, and creative methodological outputs Academic conference presentations Publication of journal articles	Evaluation contact repeatedly requested that we provide analysis to show the cost effectiveness of the project, whilst not providing the internal data required for the team to deliver this. In steering group meetings, verbal promises were made about data sharing, but these were never realised despite repeated evaluator requests. No contractual data sharing agreement was in place, so ethically full knowledge

		exchange was not required. Evaluators were not aware if this data was actively withheld, or if it did not exist. Evaluation contact criticised the evaluation team for not delivering the full work required, and in negotiation with finance, reduced the final payment to the university. Evaluation contact engaged a different consultant during the evaluation, who used an alternative methodological approach and worked separately. Evaluation contact requested a final written evaluation report, plus hard copies, but then verbally criticised the evaluators for providing these, noting that they gathered dust in a cupboard.
Evaluation 2	Evaluation report (with university branding, and organisational logo) Presentation of findings at public events organised by the commissioner Academic conference presentations	Evaluators were not invited to present the final report at the end of project event. The internal review of delivery was especially challenging with evaluators verbally attacked in one focus group by senior staff members, and ignored by some workers who refused to comment in another focus group. Hostile emails were sent to the lead evaluator, commenting on the first draft of the final report. Despite being commissioned to deliver a qualitative evaluation, senior staff reported disappointment in the data not being measurable and wrote that it did not represent value for

		money. Senior staff finalised evaluation report, following negotiations.
Evaluation 3	Evaluation report (branded with VCSE details only)	Commissioners were keen to co-produce the report, so wrote some sections. The evaluation team completed data analysis, and co-ordinated the content of the final report. Evaluation reporting was delayed as the VSCE had many disagreements about the work, the data and associated reporting. Staff and volunteer sickness, as well as turnover further contributed to delays. A senior colleague in an external professional role, made a direct complaint to the university Vice Chancellor, suggesting that the lead evaluator had been bribed by the commissioning organisation to produce a report which criticised the intervention, and that the report was flawed. This was verbally communicated to the lead evaluator in a telephone call. No written communication took place in relation to this accusation. No evidence was supplied to support the claim. The evaluation lead was investigated without her knowledge – she discovered this when senior staff emails from within her own institution were supplied to external parties under a freedom of information request. The lead evaluators emails related to the project were also shared following this freedom of information request. The university legal

		department failed to fully redact some email content, and shared family names of one of the lead evaluator's children. Despite highlighting this internally, no further action was taken, and the legal department instead requested that the lead evaluator agree to all of her email content being unredacted (she did not agree). The lead evaluator was pressurised internally by senior management to 'learn lessons', and externally by a senior professional figure who used the co-produced nature of the report to criticise the data as invalid during an online meeting organised to discuss the report. Furthermore, the VCSE organisation commissioning the report also pressured the lead evaluator to submit freedom of information requests and to engage with local media reporting (she did neither). Some VCSE group members added content to the report after receiving it from the evaluators, which was not agreed in advance. The university logo was removed from the final report, with only a small sentence included to acknowledge the evaluation team's contribution to analysis.

Table 5 – summary of political challenges to knowledge exchange

Intervention	Political context
Evaluation 1	• Delivery by one VCSE, in two different community locations. • VCSE was challenged with finding continuous funding to support interventions. • Workers employed on part-time and temporary contracts.

	• Interventions were often short term funded, with external evaluation requirements tied to them.
Evaluation 2	• Partner organisations had historically competed with each other for funding, and whilst they united to deliver this intervention, it was only funded for a four year period). • Each partner organisation was invested in the delivery work, committed to supporting service users, yet all needed to secure future funding for their organisational survival. • Various attempts were made to tackle partnership tensions, including the creation of a no-compete agreement between partners however, this was never realised.
Evaluation 3	• The intervention had a history of contention within the local community. • Despite the intervention aims of supporting sex workers in an area where they had historically operated for a number of years prior to the intervention's implementation, residents frequently challenged delivery partners and made repeated complaints. • Local and national media reported negatively on many occasions. • Charities delivering support, received funding so were at risk of losing income if the intervention was discontinued. • Some organisations drew on religious principles, to challenge the interventions' aims. • The partnership delivering the work, were keen to illustrate it as successful.

Table 6 – reflections on co-production

Intervention	**Ownership, understanding and support of co-production**	**Culture of openness and honesty**	**Commitment to sharing power and decisions**	**Clear communication**	**Culture of value and respect**
Evaluation 1	Action research approach proposed and accepted, but not realised in the research delivery.	Participation in steering group meetings was positive, enabling evaluator insight. Internal data not shared, with time cited as a key barrier by workers.	Evaluator proposed solutions not accepted e.g., offer to support staff with internal data retrieval by attending in house. Decisions taken internally were not	Email updates sent regularly by the evaluation team. Evaluation team reported to agreed deadlines. Evaluation team attended meetings when invited.	Verbal challenges in steering group meetings about data produced, with a focus on what was missing, rather than valuing what had been produced.

		Access to participants to co-produce data limited by staff (in part due to delivery demands).	communicated to the evaluation team (e.g., engagement of a new consultant).	Evaluation team perceived communication to be one sided as their requests were not always addressed, whilst they continued to meet commissioner requirements.	Evaluation lead problem solved by inviting a university research manager to a meeting with the evaluation contact, who offered to host a co-produced event at the university for 'free,' to build trust.
Evaluation 2	A detailed evaluation proposal noting co-production and qualitative methods was accepted. Co-production was evident in the evolution of the research design, as project delivery became established. Evaluators were supported to use peer researchers, and trained some service users and one staff member, but this was not fully implemented	Senior staff were open about the challenges of the partnership from the outset. Open discussions were frequently held with the evaluation team.	Evaluation design was negotiated with staff throughout the project delivery. Frequent reporting by the evaluation team was well received, and welcomed e.g., internal reports, event presentations (by invitation).	Communication in this instance worked well until the final report (in draft format) was sent to the lead contact. Emails were then sent detailing the report as inadequate, not what had been agreed, and not representing value for money.	Evaluation team experienced a culture of respect from contacts until the point of final reporting, when challenges emerged. There were challenging interviews and focus groups involving staff.

	due to the emotional vulnerabilities of the peer researchers.				
Evaluation 3	Co-production had already taken place in this instance, in terms of the community creation of data.	All data gathered was shared with the evaluation team – notes, papers, voice files, survey responses.	Evaluators invited to community meetings. Evaluators invited to participate in a data gathering event. Final report co-produced: evaluation team analysed the data, and reported it, and community members contributed other sections. Evaluation team were careful to present a balanced analysis of views from listening events, these were seen to be watered down by community members, and to be too emotive, and not fact based by wider stakeholders.	Evaluation team requests were slow to be responded to. Repeat reminders and requests were sent via email. Internal conflicts were hidden within the VCSE, with the evaluation team remaining unaware until the final reporting period.	External stakeholders in senior positions wielded power to challenge the research findings, suggesting that data was not sound, and that the lead evaluator was biased due to being bribed by the VCSE. The university was focused on risk and reputational damage, not researcher wellbeing.