Editorial: Ethics in Evaluation

Chris Barnett and Laura Camfield

Abstract

In the editorial to the special section on 'Ethics and Evaluation', we propose broadening the approach and application of ethics in development evaluation by shifting the focus from 'care of the subject' (Camfield and Palmer-Jones, 2015) to the role of evaluation within a wider political and societal context. This entails not only considering the ethical principles that underpin the evaluator-subject and evaluator-commissioner relationships, but also looking more broadly at the ethical obligations of evaluators to society. The challenge is to find ways in which ethical theory and practice can inform who is included and excluded (and how) throughout the evaluation, as well as how and by whom evidence and values are debated and deliberated.

Introduction

Evaluation takes place within a political and organisational context. While it shares the same methodologies as social science research, stakeholders usually have a more immediate interest in the findings of an evaluation; such as an imminent funding decision, a policy change, or an adjustment to a programme's design or management. As such, stakeholders (in theory at least) should take a close interest in what the evaluation says[i], with findings that can sometimes reflect poorly on their organisation, programme or even personal career. It is this more direct interest in the results of an evaluation that shapes many of the relationships during an evaluation process, and yet all too often the power, values and norms that underpin these relationships are hidden or ignored. In this editorial, we explore the role that ethics has to play in guiding evaluation relationships, and argue for a need to move beyond a narrow ethical focus on protecting survey respondents (Camfield and Palmer-Jones, 2015) to a more encompassing view whereby ethics can inform judgements on who is included (and excluded) from the evaluation process, as well as how evidence, knowledge and different values are deliberated to the wider benefit of society (Barnett, 2015).

This special section arose from a series of three events on ethics from 2014-16[ii], organised by the Centre for Development Impact, and an ongoing collaboration with 3ie. These suggested that the space and demand for dialogue around ethics is expanding, with support from major donors such as the UK Department for International Development (DFID) who are devoting significant resources to updating and strengthening their approach to ethics in research and evaluation (Groves, this volume). The CDI events, and related literature review and small-scale primary research (Munslow and Hale, 2015), explored how the concept of ethics can become more relevant within the field of impact evaluation. As discussed in the special section, the findings point towards a broader model of ethics, where diverse values and principles (of researchers, practitioners, donors, and the public) are considered in relation to the process of evaluation, as well as in relation to society.

Drawing on the three papers within the special section, the editorial first considers the differences between evaluation and research, arguing that while there is no clear distinction, evaluations tend to have a more immediate and practical use for decision-makers – and as such, greater stakes for all those involved. It is these stakes that can lie behind various relationships (with the commissioner, policy-maker, implementer, etc.). The editorial argues that ethics has an important role in guiding behaviours and the values that unpin such relations, and reviews the current application of ethics in development evaluation (drawing extensively from Groves Williams, this volume). This highlights a number of limitations in the current approach to evaluation ethics, in contrast with the more established tradition of research ethics. As the editorial notes, however, research ethics can also be limited by its predominant focus on the researcher-respondent relationship rather than other stakeholders in the evaluation process.

So, where does this leave us: with a current ethical practice in evaluation that falls short, and a situation where the wholesale adoption of research ethics may not provide the most appropriate solution? Instead, this special section argues for a different approach to evaluation ethics: one that helps rebalance the primary focus on the respondent (human subject) and ensures that ethics addresses key stakeholder relationships – including duties of evaluation to society more broadly. In short, this requires a shift from a focus on minimising risks (ethical strategies of do no harm) towards increasingly considering the benefits of evaluation to society (a do good ethical strategy). The challenge is to find ways in which ethical theory and practice can inform who is included and excluded (and how) throughout the evaluation process, as well as how evidence, knowledge and values can be better debated and deliberated (and by whom) to achieve greater social progress.

The special section comprises three papers; the first, by Leslie Groves Williams, provides an overview of the topic by sharing the findings of a review of ethical principles, guidance and practice used by evaluation practitioners, researchers and commissioners working in international development. The second, by Caitlin Scott, offers a commissioner's perspective and begins to explore the political and economic background to the challenges identified by Groves Williams. The third, by Peter O'Flynn, completes the picture by offering an evaluator's perspective, which argues that the type and level of ethical scrutiny may not always be appropriate as it is dependent on the methodology of the evaluation rather than the sensitivity of the topic or the power of the interests involved. In the remainder of the editorial, we outline the arguments we draw from these papers and point towards some possible ways forward.

Evaluation and research

As we described earlier, for many, evaluation is simply a subset of research, and indeed evaluation draws extensively on the same methods of data collection and analysis as the social sciences. Others argue that there are important differences between the two, which have implications for ethical considerations (Bloom 2010): evaluation is characterised as focusing on producing practical and approximate knowledge for immediate use by clients for a specific goal or decision, whereas research focuses more on long-term understanding which may or may not have immediate implications. Patton (2014) puts this pointedly as research being something that informs *science*, while evaluation is something that supports *action*. In reality, such a distinction between evaluation and research is rarely so clear cut, particularly with blended notions of policy research, evaluation research, implementation research, and operational research. Advocates for the use of experimental methods in evaluation, and particularly Randomised Controlled Trials (RCTs), also argue that the knowledge they produce is precise rather than approximate, albeit often focused on a narrow set of questions, and produced over a longer timescale than applied research. Their precision enables them to avoid the twin ethical challenges of using either too few respondents (under-powering), which means that they are not able to measure changes with sufficient statistical precision, or too many, which increases the number of often very poor people giving time for research and possibly also not receiving treatment as a result (Djimeu and Houndolo, this volume)[iii].

There is however one characteristic that reoccurs across much of the literature (e.g. ICAI, 2014), and that is that evaluations should have implications for key decisions; i.e., evaluation findings - if timely and relevant - should have implications for stakeholder's plans, priorities and the use of resources (this is discussed more fully in Scott, this volume). It is this characteristic that often means that stakeholders have a stake in the findings.[iv] Often this stake is concentrated in the commissioner-evaluator relationship, where different interests and power asymmetries can distort the evaluation process (Barnett, 2015), particularly through the growth of 'Payment by Results' where evaluation and verification processes that are not wholly positive have real financial consequences. Of course, the political economy of the evaluation process is just one contributory factor to the final outcome of an evaluation (as explored in Anderson, 2014), but it can have a distortionary effect on all aspects:

from the drawing up of evaluation questions through to decisions around resources and methodologies, procurement and how findings are challenged and utilised.

Ethics in research

Ethics is about right and wrong behaviour, and the values that underpin such behaviour. As Bloom (2010) summarises, "Ethics …involves a set of principles of right conduct that is supposed to govern practitioners' behaviours in clinical and social change situations". Approaches to research ethics have evolved out of the need to prevent the worse cases of misuse, with a particular focus on protecting human subjects, especially in clinical and other forms of medical research. The origins of the present day approach can be traced back to the Nuremberg Code, the first international document that advocated voluntary participation and informed consent. This followed criminal proceedings against leading German physicians who conducted medical experiments on thousands of concentration camp prisoners without their consent; most of whom died or were permanently disabled as a result. Other notable large-scale medical research studies (such as on Thalidomide in the 1950s, and the Tuskegee Syphilis Study conducted from 1932-72) highlighted the misuse of research in either prescribing treatment without informed consent, or withholding medical cures once they had become available. The World Medical Association's Helsinki declaration in 1964 established key principles such as independent review of research protocols to establish the balance of risks and benefits. This was followed by the Belmont report on Bioethics in the late 1970s, which set out three core principles of respect for persons, beneficence, and justice. Following the National Research Act 1974 in the United States, regulations established the first Institutional Review Boards (IRBs) as a mechanism through which human subjects would be protected. Today's research ethics codes and committees build on this foundation of protecting those involved in a research project from abuse of their rights to anonymity, to withdraw, etc. They are often modelled on clinical structures (Stanley and Wise, 2010), despite concerns that these may not present the best model for social science research. Stanley and Wise (2010:1.2) describe how 'qualitative knowledge is often in the hands of nurse researchers who have to manage the power relations within RECs [Research Ethics Committees] between themselves and doctors/consultants who can dismiss or challenge its design and methodology without understanding them'; an argument that finds contemporary relevance in O'Flynn's (this volume) description of the attitude of IRBs to qualitative research.

There are some important considerations to draw out from research ethics, bearing in mind that there are distinct differences between different traditions and practices in different countries and their suitability for the evaluation field (e.g. the formal system of IRBs in the USA vs. the university research ethics committees in the UK, and the recent rise in the growth of private IRBs in the global South to respond to an increasing number of clinical trials):

Firstly, research ethics is often mostly concerned with the researcher-human subject relationship, especially around issues of consent and protecting confidentiality. As Scheyvens et al. describe `fieldwork [...] can give rise to a plethora of ethical dilemmas, many of which relate to power gradients between the researcher and the researched. Combined with this are complex issues of knowledge generation, ownership and exploitation' (Scheyvens et al. in Scheyvens and Storey 2003: 139). This contrasts with evaluators' often complex relationships, particularly in international development.[v] These relationships include respondents / human subjects, but also the commissioners, various stakeholders (funders, policy-makers, programme implementers), and citizens more broadly.

Secondly, there is a strong focus in research on ethical principles and mechanisms to achieve this protection (such as IRBs or research ethics committees). In evaluation, practice is much more varied with many (perhaps even most) evaluations falling outside of such mechanisms, as discussed in O'Flynn, this volume. Instead, where evaluation is viewed as an extension of policy making or project

management, then ethics often tends to fall to the professional judgement, guided by evaluation society or organisational principles.

Thirdly, and building on the first point, evaluators' ethical obligations to human subjects can be at odds with their ethical obligations to wider society. An example of this is the way that research ethics is often about minimising risks – not only to the human subjects, but also to the researcher, and the researcher's organisational affiliation – which can affect the types of methods that are preferred (O'Flynn, this volume). This is particularly evident in IRB policies towards data sharing (Wolf et al, 2006)[vi], which can be seen as sacrificing potentially greater goods such as transparency and accountability by showing how and from whom the data was generated to minimise disclosure and reputational risks. That is not to say that evaluation ethics should not be about protecting individuals by minimising risks, however taking this as its focus can involve rejecting other desirable features of evaluation research such as transparency and accountability, which means that the evaluations may not be as much use to society as they could be.

Ethics in evaluation: the case of international development

The ways in which these trends might impact particularly on international development are discussed in all the papers in this section, and are a particular focus of Grove William's structured review. She argues that there are considerable discrepancies in how ethics features in international development evaluation and research and that more attention is paid to logistical or methodological dilemmas, even where failure to systematically consider ethics can have adverse consequences for those intended to benefit from development interventions. One example she provides is where 'during the design of a study of gender based violence in refugee camps it becomes clear that many survivors are afraid to leave their homes to come and talk to the researchers. The solution from a methodological perspective is to go to the survivors' homes to meet with them. From an ethical perspective, however, it becomes clear that this approach would put the survivors at significant risk of retribution from perpetrators as it would quickly become known that researchers are visiting them to discuss their experience of GBV'. A similar critique is made of the 2010 revision of the ESRC Framework for Research Ethics by Stanley and Wise (2010:5.4) who having outlined the scope for misinterpretation of, e.g. guidance on the use of gatekeepers, argue that "the point we come back to is, 'it all depends' – and it takes a knowledgeable expert assessor to evaluate the appropriateness of the stance adopted". In addition to the ambiguity of much of the guidance, Groves Williams and to some extent O'Flynn argue that there is little evidence of it being applied in practice and few mechanisms to police this and that interviews with practitioners and commissioners suggest considerable confusion around ethical principles, standards and norms (Munslow and Hale, 2015). While there is a role for 'knowledgeable expert assessor', we argue that it is also risky to rely on an individual evaluator's professionalism, especially in a context where there may be many other pressures such as shortage of time preventing them from exercising this.

Where next – realising a duty to society?

In the preceding section we argued that ethical practice in evaluation is often narrowly focused, constrained by lack of resources and competing interests, left to the professionalism of the individual evaluator, and poorly scrutinised. Arguably many of these critiques apply to research ethics as well, despite an increasing bureaucratisation that has been strongly resisted by many researchers (e.g. Hammersley, and Traianou, 2012, 2014), due in part to its bias towards more structured methods. So, should evaluation simply take ethical models from research, where codes and institutional set-ups are well established? Indeed, many posit that evaluation ought to be treated as a subset of research - given the methodological similarities between the two fields (this was certainly the feeling of participants in the CDI workshops referred to earlier). There are certainly advantages in the research approach, but we argue, this entails too narrow a focus on the evaluator-subject relationship (consent, anonymity, etc.) and insufficiently engages with the wider role of

evaluation in policy debate and society. In the closing section we discuss other possible approaches, drawing on examples from recent and ongoing evaluations.

Moving beyond protection of the subject: their rights for inclusion

Ethics can inform judgements on who is included (and excluded) in the evaluation process, and how the rights of a broader stakeholder group to engage can be undertaken in a transparent and considered manner. As argued in Scott's paper, the commissioner (typically donor in international development) often has a key stake and influence in the evaluation process – a power and position to direct and determine scope, questions, resources, methodologies and influence eventual findings. Arguments for evaluation independence (Picciotto 2013) recognise this issue and the importance of governance structures that protect the evaluator – and sometimes the evaluation unit / commissioner - from undue influence from operational and project staff. But still, they typically bias the main messages of the evaluation towards the funder of the evaluation (typically a donor), or a group of elite stakeholders (through evaluation committees, peer review structures, etc.). The rights (and inclusion) of those being evaluated are rarely in such an influential or privileged position. Re-orientating an ethical principles to 'protect the respondent' to one where they have a 'right for inclusion' alters the power base in the complex web of evaluation relationships.

This is not however an argument for participatory evaluation – which equally has a set of power relations around inclusion and exclusion – but rather a set of ethical principles that could be applied to any evaluation (c.f. Hammersley and Trainiou's (2014) argument that justice needs to be considered alongside care in any evaluation). The value of this would be that most ethical guidelines are implicitly informed by theory, whether these relate to the primacy of the rights of the subject or to arguments around a greater good used to justify covert observation, deception, and to some extent experimental games (Iversen, 2014). Articulating the different ethical theories and principles that underpin an evaluation leads to greater clarity about what is actually at stake and the likelihood that a middle ground can be found. So for example, this could mean taking a more deliberate (and transparent) approach at every stage of an evaluation to address the trade-off between methodological rigour and ethical principles such as inclusion. Typically debates between more experimental and more participatory perspectives of evaluation have tended to focus on methodological differences – and different views of rigour and evidence. But as Hemelrijck and Guijt (2016) highlight in their pilots in Vietnam and Ghana (IFAD and BMGF 2015, 2014), it is possible to consider this trade-off in a more explicit way; choosing to include (and exclude) in different ways and for different stages of the evaluation process (from questions and designs through to analysis and findings). This is very different to most evaluations that unconsciously (and un-transparently) include and exclude people in the evaluation process – typically the funders and commissioners with the more powerful voices, as discussed earlier.

Deliberating evidence, knowledge and value to society

A second approach to broadening the application of ethics to the evaluation field is to move beyond a duty to the respondent to a duty to society, as advocated by both the Social Research Association in their 2003 guidelines, which address obligations to society, funders and employers, and colleagues, before they address ones to subjects, and the Government Office for Science's (2007) Universal Ethical Code for Scientists, with its keywords 'respect, rigour and responsibility'. This approach also acknowledges that evaluation findings are not a singular objective 'truth' as implied by some experimental studies, but one form of knowledge to be debated and valued by society. One ethical shift towards a duty to society perspective, is the open data access movement and the view that evaluations and their datasets should be open to peer review, as well as re-analysis and re-use (Camfield and Palmer-Jones, 2013). This sits alongside a more deliberative view of policy and practice, where evaluation is one form of knowledge amongst many – and which should be tested and debated alongside other values, perspectives and forms of knowledge (see Holland (2013) on

the power of participatory statistics and McGregor et al (2015) describing the use of 'deliberative forums' in Zambia where statistical data and local perceptions were debated by local people, district officials, national policy makers, and project staff.

In summary, this editorial has discussed how evaluations sit within a political and organisational context, where different stakeholders have interests that can influence the final evaluation outcome. Groves Williams' (this volume) study of current ethical practice in international development highlights the ambiguity of much guidance and limited evidence that it is being applied in practice. O'Flynn (this volume) further emphasises the disjointed nature of how evaluation ethics is viewed and applied in evaluation by contrasting the research-informed and professional-orientated approaches. We argue in this editorial that the research-informed approach presents a too narrow view of ethics for evaluation's purpose, particularly given range and influence of those with a stake in the evaluation. As Scott (this volume) highlights, these relations are often dominated and distorted by the more powerful, typically the funder or commissioner. It is here that ethics has a role in helping evaluators (and all stakeholders) to be more explicit and deliberate in whose values and rights are included – and even excluded. We put forward a couple of proposals in this direction: one that considers the trade-off between methodological rigour and inclusion at each stage of an evaluation and the other that views evaluation as one form of knowledge to be deliberated, not just by policy-makers (the powerful) but citizens more generally. These are only proposals, however, they have in common their attempt to tackle complex and unbalanced relationships within an evaluation in a way that is more transparent and deliberate about their ethical and political underpinnings.

References

1. Andersen, O. W. (2014). Some thoughts on Development Evaluation Processes. In: Befani, B., Barnett, C. and Stern, E. (eds.) Rethinking Impact Evaluation for Development, IDS Bulletin, Volume 45, Number 6, UK: Wiley Blackwell, pp. 77 –84.
2. Barnett, C. (2015) An Evaluator's Perspective: Imbalances of Power in Evaluation Design and Implementation, pages 8-9 of Evaluation Connections, December 2015, European Evaluation Society.
3. Barnett, C. and Munslow, T. (2014) Workshop Report: Framing Ethics in Impact Evaluation: Where Are We? Which Route Should We Take? IDS Evidence Report 98.
4. Bloom, M. (2010). Client-Centered Evaluation: Ethics for 21st Century Journal of Social Work Values and Ethics, Volume 7, Number 1 (2010) (page numbers not supplied).
5. Camfield, L. (2014). Ethics in Evaluation. Unpublished paper presented at the 2nd Manchester Impact Evaluation Workshop (Manchester, autumn 2014).
6. Camfield, L., Palmer-Jones, R. (2013). Editorial: Three "Rs" of Econometrics: Repetition, Reproduction and Replication. Journal of Development Studies, 4(12): 1607-1614.
7. Camfield, L, Palmer-Jones, R. (2015). 'Ethics of analysis and publication in international development' (p273-89) in Nakray et al, Social Science Research Ethics for a Globalizing World: Interdisciplinary and Cross-Cultural Perspectives. London, Routledge.
8. de Wolf, VA, Sieber, JE, Steel, PM, and Zarate, AO. (2006). Part III: Meeting the Challenge When Data Sharing Is Required. IRB: Ethics & Human Research, 28(2):10-15.
9. Djimeu, E. and Houndolo, D.-G. (this issue) "Power Calculation for causal inference in social science: sample size and minimum detectable effect determination" Journal of Development Effectiveness 8(4):xxxx.
10. Government Office for Science. (2007). Rigour, Respect, Responsibility: A Universal Ethical Code for Scientists. UK Department for Innovation, Universities and Skills.

11. Hammersley, M., Traianou, A. (2012). Ethics in Qualitative Research Controversies and Contexts. Sage Research Methods, UK.
12. Hammersley, M., Traianou, A. (2014). An Alternative Ethics? Justice and Care as Guiding Principles for Qualitative Research. Sociological Research Online, 19 (3), 24.
13. Hemelrijck, A. and Guijt, I. (2016) Balancing Inclusiveness, Rigour and Feasibility: Insights from Participatory Impact Evaluations in Ghana and Vietnam, CDI Practice Paper, Number 14, IDS, Brighton.
14. Holland, J. (2013). Who Counts?: The Power of Participatory Statistics. Practical Action Publishing, UK.
15. ICAI, 2014. How does DFID learn. ICAI report 34.
16. IFAD and BMGF (2014) Impact Assessment of the 'Doing Business with the Rural Poor' Project in Bến Tre, Vietnam (2008–2013). Report of the Pilot Application of a Participatory Impact Assessment and Learning Approach, Rome: IFAD and BMGF
17. IFAD and BMGF (2015) Improved Learning Initiative for the Design of a Participatory Impact Assessment and Learning Approach (PIALA): Methodological Reflections following the Second PIALA Pilot in Ghana, Rome: IFAD and BMGF.
18. Humphreys, M. (2015). Reflections on the ethics of social experimentation. Journal of Globalization and Development 6(1): 87–112.
19. McGregor, J.A, Coulthard, S, Camfield, L. (2015). Measuring what matters. The role of well-being methods in development. ODI Project note 4.
20. Munslow, T. and Barnett, C (2015) Event Report: Right or Wrong? What Values Inform Modern Impact Evaluation? IDS Evidence Report 139, IDS, Brighton.
21. Munslow, T. and Hale, K. (2015) Background Report: Right or Wrong? What Values Inform Modern Impact Evaluation? IDS Evidence Report 140, IDS, Brighton.
22. Picciotto, R. (2013). Evaluation Independence in Organisations, Journal of Multi-Disciplinary Evaluation, 9.20, 18-32.
23. Scheyvens, R. and Storey, D. (eds) (2003) Development Fieldwork: A Practical Guide. London: Sage.
24. Social Research Association. (2003). Ethical guidelines, http://the-sra.org.uk/wp-content/uploads/ethics03.pdf (downloaded 29/08/16).
25. Stanley, L. and Wise, S. (2010) The ESRC's 2010 Framework for Research Ethics: fit for research purpose? Sociological Research Online, 15 (4).

---

[i] Not necessarily an interest because they see the evaluation as important or valuable for an immediate decision (that is the ideal, but so often not achieved in reality). But at the very least, a selfish interest in whether an evaluation findings reflect badly on their programme or project.

[ii] See Barnett and Munslow, 2014, Munslow and Barnett, 2015 and Munslow and Hale, 2015.

[iii] Ethical issues specific to experimental methods are discussed in Humphreys (2015).

[iv] Morris and Cohn (2013) (in Camfield, 2014) reported that 42% of respondents in a survey undertaken by the American Evaluation Society experienced pressure to misrepresent findings, and of those, 70% experienced this more than once.

[v] For instance, accountability relationships in evaluations commissioned by the public sector are complicated by donor country and host country dynamics – ultimately accountability of public sector spend to governments and citizens of both countries.

[vi] As discussed at the CDI workshop (Barnett and Munslow 2014), stimulated by one of the author's reflections on the Millennium Villages evaluation and experience with the Columbia University's IRB.