

By Martin Prowse

Aid effectiveness: the role of qualitative research in impact evaluation

Increased aid flows and the harmonisation agenda have focused attention on the effectiveness of the international aid system. Increased emphasis is being placed on donor alignment with national priorities and greater co-ordination in implementation and evaluation. A different and potentially radical approach to increasing effectiveness has been taken by Abhijit Banerjee and colleagues at the Abdul Latif Jameel Poverty Lab at MIT, US (Banerjee et al., 2007). Focusing less on process and more on outcomes, Banerjee and his colleagues argue that aid should be subject to the rigours of Randomised Control Trials (RCTs) to increase efficiency and efficacy.

Aid thinking, Banerjee suggests, is lazy thinking: limited numbers of weak evaluations contribute to a lack of consensus around the simplest of questions – what works? Aid interventions, so Banerjee's argument goes, are too diffuse to be effective, with anecdotal and tentative findings from poorly conducted evaluations often framed as social scientific fact. Such lazy thinking, and the costs imparted by using weak evaluative tools and inappropriate methods and inference only reinforce the arguments of aid pessimists. Banerjee and his colleagues, self-professed aid optimists who sit at MIT across from Massachusetts General Hospital, believe they have an answer to developing a more robust evidence base for aid: to import the medical model of RCTs into the aid business.

Clearly, tools that increase the evidence base for pro-poor decision-making must be welcomed. For example, 'with or without' RCT

comparisons (see Box 1) that use a naturally-occurring experimental design allow for a very straightforward interpretation of project interventions. However, Banerjee and his colleagues go further than this: they imply that all aid disbursements should be based on RCT evidence.

Responses to the plea for RCTs

Above and beyond questions about how policy must align with context, the most salient arguments in the responses to Banerjee's essay can be clustered around four headings: the scale and reach of evaluations; technical concerns; moral and ethical issues; and political dimensions, for both donors and governments.

- **Scale and reach of evaluations:** Whilst RCTs appear ideally suited to small-scale projects, they are not suitable for evaluating broad policy changes: macroeconomic policies such as exchange-rate policy or trade regimes are not suitable subject matter for RCTs (Goldin et al., 2007; Bhagwati, 2007). Nor are RCTs appropriate for labour-market reforms or investments in infrastructure, such as the creation of a power plant or road construction, or the provision of basic services in health and education (White, 2007).
- **Technical concerns:** Whilst RCTs may increase the efficiency in the allocation of aid flows – getting the most 'bang for your buck', if you like – 'before and after' RCTs require baseline data that may not be obtainable. The time required to ensure interventions are firmly embedded may conflict with the short time horizons of donors and govern-

Overseas Development Institute

ODI is the UK's leading independent think-tank on international development and humanitarian issues.

Box 1 - What is a RCT?

A Randomised Control Trial in social science is an evaluation of a public policy intervention. Research is structured to answer a counterfactual question: how would participants' welfare have altered if the intervention had not taken place? This can involve 'before and after' and 'with and without' comparisons. The former are not dissimilar to more conventional evaluation tools that use baseline data, and may suffer from difficulties in isolating the effects of an intervention from wider societal changes. The latter create a robust comparison group who are not directly exposed to the intervention, and whose outcomes would have been similar to participants if the intervention had not taken place. Such 'with and without' comparisons allow researchers to estimate the average effect of the intervention across the participant group. The main difficulty is in minimising selection bias for the two groups – hence the importance of randomisation.

ments (Goldin et al., 2007).

- **Moral and ethical issues:** The case for RCTs in the aid industry must confront similar moral concerns to those faced in medicine some time ago: path dependence, institutional histories, and an aversion to accepting that some interventions just do not work (Goldacre, 2007). However, the ethical case against using portions of the population as a control group may be strong: are we able to make a case for withholding an intervention from potential recipients as part of a trial when the intervention may save lives? (Goldin et al., 2007).
- **Political dimensions:** The case for RCTs on efficiency grounds may ignore some wider political factors contributing to aid flows – that donors have strategic interests, and that a country's status may be enhanced through giving aid (Moore, 2007). More importantly, aid flows now have lower levels of project expenditure: direct budget support and a shift towards working on governance and institutional process severely limit the potential for RCT-style evaluations (Moore, 2007). A further political angle is highlighted by Bhagwati (2007): RCT trials may increase aid effectiveness, but how does this intersect with country 'ownership'? If an intervention is adjudged to be highly cost-effective using RCT, but is rejected by a government on political grounds, the effort spent on rigour and comparison may come to nothing.¹

The challenge of aid: Not just technocratic

Such conundra highlight how the challenge of aid is not just technocratic, but is a social and political endeavour. Acknowledging this line of argument opens up potentially fruitful lines of enquiry. For example, what kinds of questions are RCTs good or bad at answering? And what types of data (and knowledge) can RCTs generate? This Background Note suggests that for RCTs to be able to tell us 'why' or 'how' project interventions work, they may well need to incorporate some of the insights that qualitative research can offer. To be able to make this suggestion, and discuss these wider lines of enquiry, we need to delve briefly into the philosophy

of social science, and discuss the extent to which different research traditions (and the methods they are linked to) can be combined.

Positivism

Randomised control trials, as with much policy-related research, rely on the positivist worldview – the most common standpoint in the social sciences. Positivism seeks to understand the social world by uncovering universal 'laws' through the measurement of the 'constant conjunction of events' between two or more phenomena. These 'laws' are empirical generalisations which are seen to be (mainly) independent of time/space and are neutral and value-free (Steinmetz, 1998). For positivists, the observation that two variables are strongly correlated is often understood to signify a causal relationship. As we've seen with RCTs, positivists discover empirical generalisations through setting up and testing hypotheses in a deductive manner, with non-falsified hypotheses being extrapolated to a wider range of cases (Popper, 1992; Danermark et al., 2002). Non-falsified hypotheses are accepted and extrapolated because positivists understand the social world as a closed system (Steinmetz and Chaen, 2002). In this respect, positivism is a form of naturalism, based on a belief in the unity of the natural and social sciences (Bhaskar, 1989), and hence attempts to replicate the requirements (and methods) of physical science, such as prediction, closed experimentation and the separation of research findings from interpretation in social research (Steinmetz and Chaen, 2002).

However, importing natural science methods into social science research is not straightforward: there are three contextual effects that can make positivist methods, including RCTs, rather problematic, and may lead researchers to incorporate insights from other standpoints (Burawoy, 1998). First, within positivist research, interview effects – the influence of the socio-biographical characteristics of the interviewer on the generation of data – tend to be marginalised and usually ignored due to the assumed objectivity of the interviewer. Second, positivist research ignores respondent effects – when the respondent understands the question in a different way to how the question was intended. Both of these contextual effects relate to the same key point: humans do not act like molecules, we interpret the social world and act accordingly.

Finally, the assumed closed system of positivism ignores field effects – the complex political, social and economic currents which permeate the social world (see Burawoy, 1998). An example from the physical world – that of a falling leaf – helps to elucidate this point (Baert, 1998). If the physical world were a closed system then, according to the law of gravity, one would expect a leaf to fall from a tree in a straight line. Instead, falling leaves are subject to a wide variety of forces, and their trajectories are highly varied and difficult to predict. This is not to

say that the law of gravity doesn't hold. Of course it does. But, even in the physical sciences, closed systems are unusual. The significance of this example is that in the social world, which is an open rather than a closed system, the physical science method of falsification and extrapolation may not be as accurate as we would like. In other words, isolating the influence of one factor from others is extremely hard – a fact that is rarely acknowledged by those in the positivist tradition. We now turn to the second, and less popular, standpoint in the social sciences.

Social constructivism

The basis of much qualitative research is a social-constructivist standpoint. Qualitative researchers are mainly concerned with the meanings and interpretations which individuals ascribe to their surroundings, actions and practices. This is because individuals' beliefs stimulate, and are constitutive of, conduct and action (Baert, 1998). As social constructivists, many qualitative researchers posit that all types of knowledge are constructed within the cognitive framework and theoretical concepts of an individual: therefore all understandings of the external world are not reflections of it, but are solely subjective interpretations (Schwandt, 2000). From this perspective, all knowledge is predicated on the values, ideas and judgements of the individual, and is locally and contextually defined (Danermark et al., 2002). Many social constructivists are 'judgemental relativists' who do not offer any criteria by which to distinguish between different interpretations and, in this respect, are sceptical of universal-truth claims and often question the validity of representations of the social world that are totalising or homogenising. Moreover, social constructivists mainly believe that the aggregation of values, ideas and beliefs, and their enactment through actions and conduct, lead to the creation of 'real' social processes – thereby discourses are not only the myths that we like to live by, but are key structuring principles in society.

Clearly, there are some deficiencies within this standpoint for development research seeking policy traction! There are two main ways in which social constructivism is problematic. First, 'judgemental relativism' is a position that is incompatible with any type of public-policy intervention. Second, some social constructivists believe in the impossibility of social and cultural translation (Steinmetz, 2004). Such researchers posit that as soon as one leaves one's own social community, social researchers are unable to 'translate' social actions and conduct. In combination with unequal power relationships between researchers and respondents, such translation difficulties contribute to social researchers imposing their own distinctions, understandings and beliefs on the social context, thereby reducing cultures and social practices to their own metric (Steinmetz, 2004). The implication of the 'impossibility of translation' argument is that social research outside of one's own community is untenable – a

position at odds with most policy-relevant research.

Can RCTs be combined with qualitative research methods?

Clearly, there may be some difficulty in trying to sequence and integrate qualitative methods within an RCT experimental design. Some argue that the axioms of positivism and social constructivism render the two approaches mutually exclusive (Denzin and Lincoln, 2000). Such an either/or position is not beneficial to proponents of either standpoint, as both research traditions, and the research methods they are most closely linked to (quantitative vs qualitative), are suited to answering very different types of research question.

In essence, positivist methods, such as RCTs, are able to tackle 'what' and 'where' questions, which means they can capture states or conditions (Ellis, 2000). Social-constructivist methods, on the other hand, are generally able to shed light on 'why' and 'how' questions (Woodhouse, 1998), and are good at capturing processes (Murray, 2002). By their nature, RCTs are therefore unable to tell us very much about how or why societal change occurs – they can certainly show us correlates of some aspect of this change, but these are not necessarily causal mechanisms (Green and Hulme, 2005). In the same way, qualitative and participatory methods are not suitable for answering 'what' and 'where' questions. Overall, a strict quantitative/qualitative divide hides much more than it illuminates, and there is a very strong case for combining both research traditions.

Whilst the traditions differ substantially, as illustrated by the five continua in Box 2, the methods are not completely polarised, and often include elements associated with the other tradition. For example, qualitative research requires some numerical data (and may even utilise a random sample of a sufficient size to try to generate statistical significance), whilst quantitative research can include some open-ended questions. Clearly, both methods are important for improving aid effectiveness – RCTs can tell us which interventions are most successful and which are failures, but can't, on their own, tell us precisely why or how success or failure has occurred. In this respect, RCTs can offer limited advice on expanding or rolling-out an intervention if it works – they often can't inform us about key transmission mechanisms.

Box 2: Five continua between positivist and social constructivist research traditions

1. Type of information on population: Numerical to non-numerical
2. Type of population coverage: General to specific
3. Type of population involvement: Passive to active
4. Type of inference methodology: Deductive to inductive
5. Type of value framework: Money-metric to multi-dimensional value.

Source: Kanbur (2001)

Are RCTs incorporating qualitative methods?

One important channel through which RCTs are gaining policy credence and popularity is through the rapidly growing field of Impact Evaluation (IE). However, there appears to be a high degree of scepticism around the extent to which non-quantitative methods should be included within this approach. Recent literature on IE hardly include qualitative methods at all. For example, the report of the World Bank's Independent Evaluation Group (2007) includes a nominal paragraph on qualitative methods, as does a recent Asian Development Bank report (2006). The CGD's Evaluation Gap Working Group report offers even less (Savedoff et al., 2006). It may well be the case that the appetite for RCTs in evaluating aid interventions is a reaction against the haphazard use of participatory and qualitative methods in recent years. And rightly so, for much recent research in these traditions has not been conducted with enough rigour. However, it is vital that we do not

throw the baby out with the bath water – as Baker (2000) demonstrates, qualitative research should be a vital component of RCT trials, and examples can be easily gleaned from the Network of Networks Impact Evaluation Initiative (NONIE) database.

It is clear that RCTs can offer a mode of increasing evidence-based decision-making for particular aid interventions in particular contexts. However, locating the spaces and places where RCTs are most suitable, and how they can be combined and iterated with qualitative methods and evaluation tools, are key questions that governments and aid agencies should start to address. Instead of reverting to an either/or approach to evaluating aid effectiveness, an and/with perspective offers more than an answer to our original question about 'what works' – it can tell us how and why such success occurs. Surely, methods that can shed light on these vital questions should be a core component of any RCT trial.

By ODI Research Officer, **Martin Prowse**
(m.prowse@odi.org.uk).

Endnotes and references

Endnotes

1. The first part of this Background Note first appeared as a book review of Banerjee et al. (2007) in *Development Policy Review*, 2007, 25 (6): 741-767.

References

- Asian Development Bank (2006) *Impact Evaluation: Methodological and Operational Issues*. Manila: Economic Analysis and Operations Support Division, Economics and Research Department, Asian Development Bank, September.
- Baert, P. (1998) *Social Theory in the Twentieth Century*. Cambridge: Polity Press.
- Baker, J. L. (2000) *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Directions in Development series. Washington, DC: World Bank.
- Banerjee, A. V. et al. (2007) *Making Aid Work*. Cambridge, MA: Boston Review Book, MIT Press.
- Bhagwati, J. (2007) 'If it is Hard to Think of Aid Being Spent Productively in Africa, Why Not Spend Elsewhere for Africa?', in Banerjee et al.
- Bhaskar, R. (1989) *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences*. London: Harvester Wheatsheaf.
- Burawoy, M. (1998) 'The Extended Case Method', *Sociological Theory* 16 (1): 4-33.
- Danermark, B. (2001) *Explaining Society: An Introduction to Critical Realism in the Social Sciences*. London: Routledge.
- Denzin, N. K. and Lincoln, Y. S. (eds) (2000) *Handbook of Qualitative Research*. London: Sage Publications.
- Ellis, F. (2000) *Rural Livelihoods and Diversity in Developing Countries*. Oxford: Oxford University Press.
- Goldacre, B. (2007) 'Testing Social Policy', *The Guardian*, Saturday 21 July. Available at <http://www.guardian.co.uk/life/badscience/>
- Goldin, I., Rogers, F. I. and Stern, N. (2007) 'We Must Tackle Development Problems at the Level of the Economy as a Whole', in Banerjee et al.
- Green, M. and Hulme, D. (2005) 'From Correlates and Characteristics to Causes: Thinking About Poverty from a Chronic Poverty Perspective', *World Development* 33 (6): 867-80.
- Kanbur, R. (ed.) (2001) *Q-Squared: A Commentary On Qualitative And Quantitative Poverty Appraisal*. Working Paper No. 105. Ithaca, NY: Cornell University Department of Applied Economics and Management.
- Moore, M. (2007) 'The New Private Philanthropies Could Challenge the Existing Aid Business', in Banerjee et al.
- Murray, C. (2002) 'Livelihoods Research: Transcending Boundaries of Time and Space', *Journal of Southern African Studies* 28 (3): 489-509.
- Network of Networks Impact Evaluation Initiative (NONIE) <http://www.worldbank.org/ieg/nonie/index.html>
- Popper, K. R. S. (1992) *The Logic of Scientific Discovery*. London: Routledge.
- Q-squared Project, <http://www.q-squared.ca/>
- Savedoff, W. D., Levine, R. and Birdsall, N. (2006) *When Will We Ever Learn? Improving Lives Through Impact Evaluation*. Report of the Evaluation Gap Working Group. Washington, DC: Center for Global Development.
- Schwandt, T. A. (2000) 'Three Epistemological Stances for Qualitative Inquiry: Interpretivism, Hermeneutics, and Social Constructionism', in Denzin and Lincoln.
- Steinmetz, G. (1998) 'Critical Realism and Historical Sociology: A Review Article', *Comparative Studies in Society and History* 40 (1): 170-86.
- Steinmetz, G. (2004) 'Odious Comparisons: Incommensurability, the Case Study and "Small N's" in Sociology', *Sociological Theory* 22 (3): 371-400.
- Steinmetz, G. and Chae, O. B. (2002) 'Sociology in an Era of Fragmentation: From the Sociology of Knowledge to the Philosophy of Science and Back Again', *Sociological Quarterly* 43 (Winter): 111-37.
- White, H. (2007) 'Technical Rigor Must Not Take Precedence Over Other Kinds of Valuable Lessons', in Banerjee et al.
- Woodhouse, P. (1998) 'People as Informants', in A. Thomas, J. Chataway and M. Wuyts (eds), *Finding Out Fast: Investigative Skills for Policy and Development*. Oxford: Oxford University Press.
- World Bank (2007) *Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank*. Washington, DC: Independent Evaluation Group, World Bank.



Overseas Development
Institute

111 Westminster Bridge
Road, London SE1 7JD

Tel: +44 (0)20 7922 0300

Fax: +44 (0)20 7922 0399

Email:
publications@odi.org.uk