



IDS WORKING PAPER

Volume **2013** No **431**

CSP WORKING PAPER Number 010

Evaluating Outside the Box: An Alternative Framework for Analysing Social Protection Programmes

Stephen Devereux, Keetie Roelen, Christophe Béné, Deepta Chopra, Jennifer Leavy,
and J. Allister McGregor
August 2013



**Centre for
Social Protection**



**Institute of
Development Studies**

The Centre for Social Protection (CSP) supports a global network of partners working to mainstream social protection in development policy and encourage social protection systems and instruments that are comprehensive, long-term, sustainable and pro-poor. We produce research on conceptual approaches; design issues, including delivery, targeting and affordability; and impacts of different social protection initiatives.

Email: socialprotection@ids.ac.uk

Web: <http://www.ids.ac.uk/idsresearch/centre-for-social-protection>

CSP WP010

CSP Working Paper series editor: Stephen Devereux



The Vulnerability and Poverty Reduction (VPR) Team aims to construct dynamic and multi-dimensional perspectives on vulnerability and poverty in order to transform thinking, policy and practice.

The VPR team produces working papers on social protection; conflict, violence and development; and poverty and inequality. Follow this link to view a full list of publications:

www.ids.ac.uk/research-teams/vulnerability-and-poverty-reduction-team/publications/vpr-working-paper-series

Evaluating Outside the Box: An Alternative Framework for Analysing Social Protection Programmes
Stephen Devereux, Keetie Roelen, Christophe Béné, Deepta Chopra, Jennifer Leavy, and J. Allister McGregor

IDS Working Paper 431

© Institute of Development Studies 2013

ISSN: 2040-0209 ISBN: 978-1-78118-132-4

A catalogue record for this publication is available from the British Library.

All rights reserved. Reproduction, copy, transmission, or translation of any part of this publication may be made only under the following conditions:

- with the prior permission of the publisher; or
- with a licence from the Copyright Licensing Agency Ltd., 90 Tottenham Court Road, London W1P 9HE, UK, or from another national licensing agency; or
- under the terms set out below.

This publication is copyright, but may be reproduced by any method without fee for teaching or nonprofit purposes, but not for resale. Formal permission is required for all such uses, but normally will be granted immediately. For copying in any other circumstances, or for re-use in other publications, or for translation or adaptation, prior written permission must be obtained from the publisher and a fee may be payable.

Available from:

Central Communications, Institute of Development Studies, Brighton BN1 9RE, UK

Tel: +44 (0) 1273 915637 Fax: +44 (0) 1273 621202

E-mail: bookshop@ids.ac.uk

Web: www.ids.ac.uk/ids/bookshop

IDS is a charitable company limited by guarantee and registered in England (No. 877338)

Evaluating Outside the Box: An Alternative Framework for Analysing Social Protection Programmes

Stephen Devereux, Keetie Roelen, Christophe Béné, Deepta Chopra, Jennifer Leavy, and J. Allister McGregor

Summary

The evidence base on social protection programmes is expanding rapidly, largely pointing towards their positive impacts. Most evaluations rely heavily on quantitative techniques and experimental methods to allow for attribution of impacts. In this paper, we argue that new ways of investigation and analysis are needed to expand and deepen the evidence base in support of improved design and implementation of social protection. Greater emphasis on qualitative research, mixed methods and participatory evaluation is crucial, given current evidence gaps about programme dynamics and impacts. This paper proposes a new evaluation framework that goes beyond conventional approaches, by highlighting relatively neglected aspects related to programme processes, social dynamics and feedback loops in programme evaluations.

Keywords: social protection; impact evaluation; mixed methods.

Stephen Devereux is a Research Fellow in the Vulnerability and Poverty Reduction team at IDS and a founding Director of the Centre for Social Protection. He works on food security, rural livelihoods and social protection in sub-Saharan Africa.

Keetie Roelen is a Research Fellow at IDS in the Vulnerability and Poverty Reduction team and a member of the Centre for Social Protection. Her research interests include child poverty, poverty dynamics and social protection. The majority of her research uses mixed methods.

Christophe Béné is a Research Fellow in the Vulnerability and Poverty Reduction team at IDS. A socio-economist and policy analyst with more than 15 years of experience in developing countries, his inter-disciplinary research focuses on poverty, vulnerability and food security.

Deepta Chopra is a Research Fellow in the Vulnerability and Poverty Reduction team at IDS. A social policy researcher with a focus on policy processes, she has extensive experience in managing, designing and understanding social protection and livelihoods programmes.

Jennifer Leavy is a Senior Research Fellow in the School of International Development at the University of East Anglia. Her research interests include youth, social mobility and poverty dynamics. She works on mixed-method impact evaluations in social protection in sub-Saharan Africa.

J. Allister McGregor is a Research Fellow and Team Leader of the Vulnerability and Poverty Reduction team at IDS. He uses his background in economics and anthropology to study how systems of governance and the formulation and implementation of development policies impact upon poor people.

Contents

Summary, keywords, author notes	3	
Acknowledgements	4	
Introduction	5	
1 Research on social protection impacts	7	
2 Impact evaluation methodologies	10	
3 Alternative evaluation framework	13	
3.1 Programme processes	15	
3.2 Social dynamics and less measurable impacts	16	
3.3 Feedback loops	18	
4 Methodological implications	20	
5 Discussion and conclusion	22	
References	23	
Figures		
Figure 1.1	Characteristics of 222 evaluations of social protection programmes, by outcomes measured	8
Figure 3.1	Standard evaluation framework for social protection programmes	13
Figure 3.2	Proposed alternative evaluation framework for social protection programmes	16

Acknowledgements

Earlier versions of this paper were presented at the Maastricht University Graduate School of Governance seminar series in the Netherlands, and at the UEA/IDS workshop on 'Mixed Methods Research in Poverty and Vulnerability' in London. The authors thank participants at both events, as well as Laura Camfield and an anonymous referee, for their useful comments. Any remaining errors are the responsibility of the authors.

Introduction

The emergence of social protection as a development policy agenda since the late 1990s has been accompanied by increasingly rigorous efforts to assess its impacts, especially social transfer programmes that deliver cash, food, assets or inputs to poor individuals or households in the expectation that these transfers will alleviate their poverty and food insecurity. Credible evidence of positive impacts is crucial to assess the effectiveness of social protection programmes, to determine which interventions are most cost-effective, to build political support and to mobilise funding to expand social protection coverage. Standard impact evaluation methodologies have been applied in Africa, Asia and Latin America to unconditional and conditional cash transfers, public works, school feeding, asset transfers, food aid, agricultural input subsidies and other social protection schemes (Hagen-Zanker, McCord and Holmes 2011).

The potential for social protection to enhance the wellbeing of poor and vulnerable people is increasingly recognised. In particular, a growing evidence base on conditional and unconditional cash transfer programmes confirms their positive impacts on a range of indicators, from levels of income and consumption to access to education and health care (DFID 2011; Hagen-Zanker *et al.*, 2011). It is no exaggeration to describe social transfers¹ as ‘an effective intervention to enhance the participation of the poor in economic development and to combat inequality, social exclusion and chronic poverty’ (Stewart and Handa 2011: 13), or even as ‘a paradigmatic shift in poverty reduction’ (Hanlon, Barrientos and Hulme 2010: 4). The rapidly growing numbers of social protection programmes being designed and implemented across the world testify to the momentum and confidence around these interventions. Global economic crises, the persistence of chronic poverty and food insecurity and the HIV/AIDS pandemic have all been identified as triggers for this growing interest in social protection – especially social transfers – as part of the solution (Garcia and Moore 2012).

Evaluations of social transfer programmes have become dominated by quantitative survey-based methods and the use of ‘difference-in-difference’ econometric techniques. In the difference-in-difference approach, an attributable programme impact is identified when a change is observed in a quantifiable indicator of interest, between the baseline (pre-intervention) survey and the follow-up (post-intervention) survey, and this change is statistically significantly larger for programme participants than for a control group of comparable non-participants.² We identify several reasons for this trend, beyond a generic ‘professionalisation’ of development programming.

First, social protection in much of Africa, parts of Asia and, to a lesser extent, Latin America is largely donor-driven and donor-financed, and donors need to demonstrate impact for their aid spending. Statistically proven impacts are powerful tools to underpin ‘evidence-based policy’ – answering the question: what works? – while the standardised nature of numbers allows for comparability across different types of programmes – answering the question: what works best? Second, in a budget-conscious world where ‘value for money’ drives policy choices, assessing which intervention gives the biggest ‘bang per buck’ answers the question: which intervention is most cost-effective? Third, in a target-driven policy environment led by the Millennium Development Goals, measuring the effectiveness of interventions against calibrated indicators of progress has become mandatory.

¹ “Social transfers are non-contributory, publicly funded, direct, regular and predictable resource transfers (in cash or in kind) to poor or vulnerable individuals or households, aimed at reducing their deficits in food consumption, protecting them from shocks (including economic and climatic), and, in some cases, strengthening their productive capacity.” (Europe Aid 2012: 8).

² We prefer the value-neutral terminology of ‘programme participant’ or ‘social transfer recipient’, rather than the value-laden term ‘beneficiary’.

Fourth, because governments and donors need to demonstrate impact, it follows that their attention is focused on outcomes that can be measured and quantified (e.g. changes in poverty headcount or poverty gap, child nutrition status or dietary diversity, and school enrolment rates or attendance). Although a qualitative component is often added to the research design, this material typically only contextualises or validates the tabulated survey data or provides narrative case studies (often pejorated as ‘anecdotal evidence’) in evaluation reports. Good examples of mixed methods evaluations that genuinely integrate qualitative and quantitative methods – from research design through to data collection, data analysis and reporting – remain disappointingly rare (Adato 2008).

A final explanation is the seductiveness of numbers and the notion – imported into the social sciences from the ‘hard’ sciences – that only quantitative data constitutes ‘rigorous’ evidence (Chambers 2008). Even though responses to household surveys are self-reported (and therefore prone to various biases) rather than objectively measured, readers of evaluation reports are led to believe that tabulated percentages are more ‘scientific’ than narrative case studies or transcribed quotations from focus group participants. Thus household poverty statistics, derived by asking the household head to declare his or her income and assets to a stranger holding a questionnaire, enjoy the same credibility as child nutrition status, which is derived from anthropometric measurement.

In this paper, we acknowledge the need for rigorous evaluation methods, not only for credibility and to build political support for social programmes, but mainly because social protection is ultimately about improving the wellbeing of poor and vulnerable people, so these interventions need to be assessed against this fundamental objective: are beneficiaries or participants less poor and less vulnerable after participating in social protection programmes than before? This paper therefore makes the case for an alternative framework for evaluating social protection programmes, which addresses several limitations of current evaluation methodologies and proposes ways of taking overlooked aspects into account. Specifically, we propose an alternative research framework that would strengthen and expand the evidence base in three distinct ways. First, our framework explicitly recognises that programme design and processes – such as targeting criteria and delivery or payment mechanisms – in themselves affect programme outcomes, not least in the extent to which these design choices and implementation features are empowering (or disempowering) of programme participants. Second, our framework incorporates the less measurable impacts of social transfers, including social benefits and costs, unintended impacts (such as perverse behavioural incentives), effects on social relations within households (across genders and generations) and between households (especially participants and non-participants), and participation in social institutions (particularly informal risk-sharing mechanisms). Third, our framework explicitly considers two types of feedback mechanisms; an automatic feedback loop – ‘recursive causality’ – whereby programme impacts may reinforce or compromise programme effects, and a deliberate learning loop, aiming to incorporate lessons learned into improved design and implementation of the social transfer programme.

Note that we conceptualise ‘social’ in two distinct ways in this paper. One refers to the important aspects of peoples’ lives that are not ‘economic’. In the wellbeing framework³ these are described as ‘dimensions’ and are categorised as people’s relational and subjective needs, to differentiate them from physical needs, which are the focus of most development programmes and evaluations. The second meaning refers to social dynamics – those social, cultural and political processes that are not economic in character but also shape and produce outcomes. In the context of social protection evaluations, the first concept of ‘social’

³ “This approach treats human wellbeing as a dynamic state that is generated through ongoing conscious and sub-conscious participation in social, economic, political, cultural and natural processes. So either people will or will not be able to acquire sufficient food to maintain their health (a physical outcome); they will or will not be able to sustain sufficient relationships with others to function effectively in society (a relational outcome); and they will or will not be able to feel satisfied about what they are able to do or to be in society (a subjective outcome)” (McGregor [forthcoming], 2014: 4).

relates to the social impact of social protection interventions, while the second describes the impacts of social protection interventions on social dynamics or processes. Our argument in this paper is that evaluations of social protection programmes typically fail to adequately consider either social dimensions or social dynamics.

Framing all of these specific points is the recognition that social policies and programmes are designed and implemented within a political context, and that politics inevitably shapes the implementation and outcomes of all social transfer programmes. In his insightful work on 'The politics of what works in tackling chronic poverty', Hickey (2007) identified several political factors as being critical, including the nature of the 'political contract' between states and poor citizens, elections, elite perceptions of the causes of poverty, and institutional arrangements within government structures. This paper concludes by arguing that incorporating our three sets of issues – programme processes, social dynamics and feedback loops – and acknowledging the importance of politics in evaluation methodologies could greatly improve the efficiency, effectiveness and sustainability of social protection programmes, and would (most importantly) generate improved wellbeing outcomes for poor and vulnerable people.

1. Research on social protection impacts

Just as the number and scale of social protection programmes have increased in recent years, so have the number of evaluations of these programmes. The evidence base is expanding rapidly. Recent syntheses of this body of literature include DFID (2011); Ellis, Devereux and White (2009); Garcia and Moore (2012); Hagen-Zanker *et al.* (2011); Hanlon *et al.* (2010); ILO, 2010; and World Bank (2009). In Africa, a number of rigorous evaluations of large-scale programmes have been conducted (notably in Ethiopia, Kenya, Malawi, Mozambique and South Africa) and a further set of evaluations are ongoing or planned (including in Ghana, Lesotho, Niger, Rwanda, Tanzania, Uganda, Zambia and Zimbabwe) (Davis, Gaarder, Handa and Yablonski 2012). In Latin America, rigorous impact evaluations of large-scale conditional cash transfer programmes have been conducted in more than half a dozen countries (including Brazil, Chile, Colombia, Ecuador, Honduras, Mexico and Nicaragua), and this represents some of the most convincing evidence on social protection impacts (Adato and Hoddinott 2010; Fiszbein and Schady 2009).

This growing evidence base on programme impacts points to significant and mostly positive effects. However, our focus here is not on the findings of these evaluations, but on which outcomes are assessed, why they are assessed and how they are assessed. Poverty, education and health indicators appear to be the most commonly measured outcomes across all social protection programmes. In their review of conditional cash transfer programmes, Fiszbein and Schady (2009: xii) draw on 'a large number of carefully constructed impact evaluations of CCT programs' to 'consider the impact that CCTs have had on current poverty, education, health, and nutrition outcomes'. A recent systematic review of three major categories of programmes – unconditional cash transfers, conditional cash transfers, and employment guarantee schemes (a subset of public works) – found that ten impact indicators are most commonly reported in evaluations (Hagen-Zanker *et al.*, 2011: 8). Percentages in parentheses indicate the proportion of 222 evaluations reviewed by Hagen-Zanker *et al.* that assessed each impact indicator.

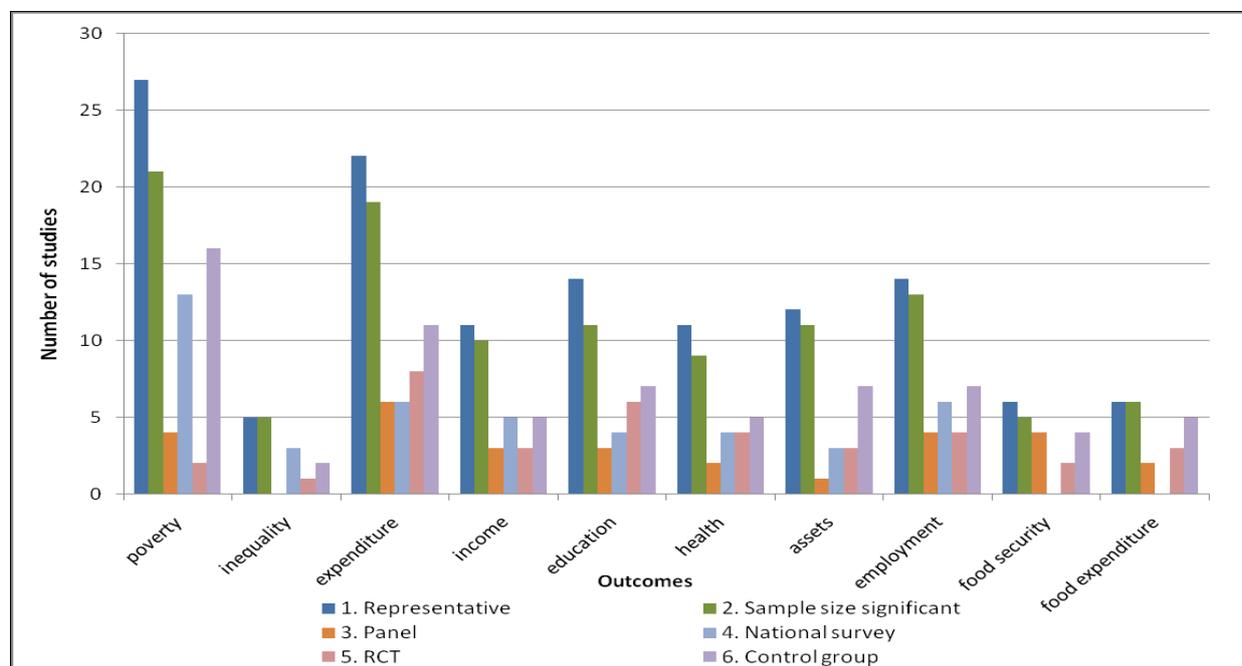
1. **Poverty** (39%): poverty headcount index, poverty gap index, and/or poverty severity index;
2. **Health** (31%): outputs (access to healthcare services) and/or outcomes (e.g. body mass index);
3. **Education** (31%): outputs (e.g. number of children attending school) and/or outcomes (educational performance);

4. **Employment** (25%): amount of paid work created by employment guarantee schemes;
5. **Expenditure** (19%): household spending, or consumption as a proxy for spending;
6. **Assets** (13%): physical, financial and/or social capital that contributes to household wealth;
7. **Income** (12%): household or individual earnings, sometimes including in-kind income;
8. **Inequality** (10%): rural or national inequality, typically measured by the Gini coefficient;
9. **Food security** (6%): indicators include calorie consumption and the number of meals per day;
10. **Food expenditure** (3%) total spending on food, or as a proportion of total household spending.

These indicators share several common characteristics. They are all positive outcomes – there is little interest in assessing negative outcomes such as stigma or conflict. They can be statistically measured and quantified. Their values are likely to change over time. They are expected to be influenced by social protection programmes. For these reasons, these are the outcomes that are included in most impact evaluations.

Hagen-Zanker *et al.*'s systematic review also assessed the methodological quality of the 222 impact evaluations, using six criteria: (1) representative sample (2) adequate sample size (3) panel data (4) national survey (5) randomised control trial (6) control group. Evaluations that meet several of these criteria score highly as having a better quality of evidence. Studies that quantify the impacts of social protection on poverty indices and expenditure tend to meet more of the data quality criteria, while those that assess food security and inequality outcomes tend to meet fewer criteria (Figure 1.1).

Figure 1.1 Characteristics of 222 evaluations of social protection programmes, by outcomes measured



Source: Hagen-Zanker, *et al.* 2011: 21

Figure 1.1 raises two fundamental concerns.

1. **The outcomes measured are limited in focus.** Although these are critically important impacts, a wide range of social impacts (e.g. on intra-community relationships or women's empowerment) and other outcomes are excluded – and some of these overlooked outcomes might also be influencing the intended outcomes.
2. **The methods used to estimate outcomes are limited in scope.** There is an explicit bias in favour of impact evaluations that follow the 'gold standard' rules (i.e. large-scale quantitative surveys with a counterfactual), as evidenced by the six data quality criteria that are used to assess the 222 evaluations in Figure 1.1. The implication is that the closer an evaluation gets to meeting these six criteria, the more 'credible' or 'rigorous' are the findings.

We would argue that the growing tendency to reify experimental methods and quantitative (statistical and econometric) techniques as the 'gold standard' (Khagram and Thomas 2010) for impact evaluations narrows what is measured and how it is measured to such an extent that it compromises the validity of evaluation findings, in several ways. First, the separation of 'monitoring' from 'evaluation' functions in M&E systems focuses analytical attention on outcomes as though these occur independently of programme processes – notwithstanding well-known evidence that how a programme is designed and implemented substantially affects its outcomes and impacts (see Thomas and Grindle 1990). Second, the focus on a relatively limited set of statistically measurable outcomes inevitably leaves substantial knowledge gaps in terms of what these programmes actually achieve – including their under-reported impacts on relationships within households and on social cohesion within communities – beyond what is selectively tested and reported in survey-based evaluations. As a corollary, although social transfers are claimed to have important 'social' effects beyond those that are readily quantifiable (e.g. on women's empowerment or the dignity of pensioners), these are taken into account in evaluations only superficially, if at all, because they cannot be easily or 'rigorously' measured. Third, quantitative methodologies ignore experiences and perceptions and exclude programme participants from contributing to evaluations, except as passive respondents to pre-coded questionnaires or as illustrative but 'anecdotal' case studies. Not only does this overlook or marginalise a potentially rich source of information, it also reinforces the tendency to see social protection impacts in terms of numbers ('before *versus* after') rather than changes in relationships and processes.

Impact evaluations might be more open to exploring social impacts and other secondary outcomes if they were not increasingly constrained by pressures to report statistically significant quantitative impacts that were derived using 'rigorous', 'credible', 'scientific' research methods (Ravallion 2009). An unfortunate consequence of this methods-driven approach to evaluation is that it biases social protection towards objectives that are not aligned with its primary function. By definition, given its origins in social safety nets, social protection should be preoccupied with managing vulnerability, building resilience and installing reliable insurance mechanisms against future livelihood shocks. But since these outcomes are less easy to quantify and track over time than, say, levels of income, these functions have become subordinated to simpler and more measurable outcomes such as poverty reduction, which should not be seen as the primary objective of social protection programmes.

2. Impact evaluation methodologies

Monitoring and evaluation (M&E) systems perform two methodologically related but conceptually independent functions: they monitor programme activities and outputs, and they evaluate programme outcomes and impacts. While monitoring simply tracks programme performance ('how many children received school meals last month?'), impact evaluation is considerably more complex ('how much did learners' performance in exams improve because they received school meals?'). Not only are specific changes 'before and after' observed and quantified, they are also interpreted and attributed. An implicit theory of change is hypothesised and tested – that there is a causal pathway from programme activities or outputs (in this example, school meals) to observed changes in outcomes of interest (exam results).

Impact evaluation is *'evaluating the positive and negative, primary and secondary long-term effects on final beneficiaries that result from a development intervention; assessing the direct and indirect causal contribution claims of these interventions to such effects especially for the poor whether intended or unintended; and explaining how policy interventions contribute to an effect so that lessons can be learnt.'* (Stern, Stame, Mayne, Forss, Davies and Befani 2012: 12)

Experimental or quasi-experimental methods – notably randomised control trials (RCTs), commonly used in medical or pharmaceutical trials– observe and quantify changes in outcomes of interest, before and after a programme is introduced, using baseline and follow-up or 'endline' surveys and applying statistical tests to determine whether any changes observed between a 'control' and 'treatment' group are significant (Banerjee and Duflo 2011). RCTs are able to attribute impact by comparing changes against the 'counterfactual' case – what would have happened in the absence of the intervention. By surveying a statistically similar control group as well as the treatment group, comparisons can be made not only 'before and after' but also 'with and without' the intervention, controlling for any possible biases or confounding factors. This is the essential logic of the 'difference-in-difference' approach to impact evaluation: the difference in the indicator of interest between the treatment and control group, before and after the intervention is introduced, is the attributable impact. (To take a simple example: if all learners average 60 per cent in exams before school meals are introduced, and those who receive school meals score 75 per cent in the next exam while their peers who do not receive school meals average 70 per cent, the attributable impact is not 75 minus 60, but 75 minus 70, or 5 percentage points.)

Many would argue that randomised control trials (RCTs) represent the most appropriate application of the difference-in-difference approach for the estimation of attributable impact (White 2010). So-called 'randomistas' consider the minimisation of selection bias by randomly allocating individuals to a control and a treatment group the only credible way of identifying impact (Ravallion 2009). Nonetheless, RCTs have been subjected to widespread criticism in recent years, largely due to ethical concerns. As Ravallion (2009: 2) describes it: 'The very idea of randomized assignment is antithetical to most development programs, which typically aim to reach certain types of people or places.' In terms of social protection, by definition randomisation means that a sub-group of poor people who genuinely need assistance – the control group – will intentionally be deprived of social transfers. In addition to these ethical concerns, RCTs also require putting methodological considerations ahead of programme design. A random assignment of people to control and treatment groups means that targeting, for example, is no longer possible. Finally, it is argued that although RCTs may be strong in terms of 'internal validity' (i.e. the extent to which they offer a robust and rigorous method for measuring and attributing impact), they are weak in terms of 'external validity' (i.e. the extent to which findings are generalisable and replicable beyond the particular study) (Ravallion 2009; Stern *et al.*, 2012). This is important because qualitative

methods are often criticised for being ‘anecdotal’ and not generalisable, whereas quantitative evidence of positive impacts from one social protection programme is often cited as proof that similar programmes will produce comparable impacts.

In cases where a RCT is not possible due to programmatic considerations (e.g. universal interventions where no control group can be established) or not desirable in light of ethical concerns, other methods are available for reducing selection bias and establishing a counterfactual. Quasi-experimental research design in combination with econometric methods such as Propensity Score Matching (PSM) and Regression Discontinuity Design (RDD) allow a comparison of changes in outcomes for non-randomised control and treatment groups.

There are a number of additional limitations with respect to the ‘difference-in-difference’ approach. Like other statistical methods, one limitation is that even if it measures an observed change in an indicator accurately, it cannot explain this outcome. “The problem is if all these studies do is estimate an average treatment effect: measurement is not evaluation” (White 2010: 156). In our school meals example, it might not be the nutritional impact of the food that generates improved exam results, it could be the boost that children get by being singled out for special attention, or it could be that the introduction of school meals signals the importance of education to parents, so they put more effort into ensuring their children study harder. A related, topical example is the ‘conditionality debate’ – what proportion of rising school attendance rates is due to the requirement on conditional cash transfer programmes that parents send their children to school, and what proportion is due to the ‘income effect’ – the fact that incremental income allows poor people to buy education? An experiment in Malawi with randomised conditional and unconditional cash transfer components found that increases in school attendance were due mainly to the income effect, implying that the conditionality – which has been credited with boosting school attendance rates throughout Latin America – may be secondary, or even redundant (Baird, McIntosh and Özler 2010). The difference-in-difference methodology only quantifies observable changes over time; it does not put this ‘treatment effect’ into its full and proper context.

A second problem with the difference-in-difference methodology is that it is limited to indicators that are amenable to quantification. We know much less about social or political impacts, which are harder to measure through statistical techniques. For example, the difference-in-difference methodology is not well suited to the assessment of ‘hard to measure’ outcomes, such as the effects of social transfers on community relations, women’s empowerment, reciprocity and stigma. As long as the methodology drives the evaluation questions and design rather than the other way around, these and other existing knowledge gaps are likely to remain or even widen further (Ravallion 2009).

A third limitation of quantitative research design and difference-in-difference methodology is that they favour indicators that are likely to respond quickly and measurably to a programme intervention. Though social protection interventions often claim to have long-term impacts, evaluations are usually conducted during the implementation period or immediately after a project ends. For instance, conditional cash transfers undeniably boost household investment in the human capital (education, health and nutrition) of children, which is claimed to improve their livelihood prospects in adult life and ultimately to break the intergenerational transmission of poverty. But this extrapolation of impacts has no empirical foundation – to our knowledge, no data are available as yet on the long-term (inter-generational) effects of any cash transfer programme in Africa, Asia or Latin America. In fact, an equally plausible counter-argument could be constructed: cash transfers alleviate income constraints, but structural barriers – low quality of education and health services for the poor, or discrimination in the labour market on the basis of caste, class or gender – might prevent children from vulnerable families from converting this into higher incomes in adulthood. This might well be the situation facing indigenous people who participate in Mexico’s

Oportunidades, a conditional cash transfer programme that has demonstrated increased uptake of education and health services by children in participating households (Ulrichs and Roelen 2012).

More generally, a ‘sceptic’s complaint’ against social transfers is that they address the symptoms of poverty rather than the causes. Standard evaluations of social transfer programmes could equally be accused of looking superficially at first-order outcomes – short-term increases in consumption of goods and services – rather than testing for transformational changes in personal characteristics, or structural changes in the contextual causes of poverty.

A number of recent contributions to the evaluation literature propose broader approaches to evaluation design and methodologies. Recognising that the kind of statistical analysis described above is not suitable for all evaluations, White and Phillips (2012) draw a distinction between different types of evaluation, depending on the number of ‘units of assignment’ involved and how attribution-focused an evaluation is. So-called ‘large-n’ impact evaluations lend themselves to quantitative studies that allow statistical tests of significant difference between treatment and comparison groups. However, different approaches are needed for ‘small-n’ evaluations, when units of assignment (for example, ‘capacity building in a single institution’) are not sufficient to conduct tests of statistical significance (White and Phillips 2012: 4). White and Phillips cluster conceptual approaches in ‘small-n’ evaluations into two groups. Group I refers to theory-based approaches, concerned with mapping out cause and effect or attribution, essentially by taking an intervention’s theory of change and mapping out the intermediate steps, presenting alternative hypotheses and using whatever methods are appropriate to uncover information and draw conclusions. Group II refers to participatory approaches, which tend not to be so explicitly concerned with cause and effect, and place stakeholder participation at the centre. Our view is eclectic: there is merit in both ‘large-n’ and ‘small-n’ methods, and in both ‘Group I’ and ‘Group II’ approaches. Wherever possible, elements of both quantitative and qualitative methods should be drawn on for holistic ‘mixed methods’ impact evaluations, and alternative hypotheses can also be tested using participatory approaches.

Others have criticised the assumption that development interventions generate impacts that are ‘monotonically increasing and linear’ (Woolcock 2009: 4). In reality, trajectories of change could be ‘J-shaped’ (things get worse before they get better) or could follow a ‘step function’ (little change until a ‘tipping point’ precipitates rapid change), or even a ‘S-curve’ (slow initial change, then rapid increase, followed by a plateau). Woolcock offers the example of a women’s empowerment project that generates violent resistance from men initially before achieving its desired outcomes. ‘[If] a ‘gold standard’ randomised evaluation happens to be conducted when the programme is at the bottom of that curve, it would be unceremoniously (but very inaccurately) deemed a failure’ (Woolcock 2009: 6).

Understanding the ‘black box’ of impact trajectories – what happens between the ‘baseline’ and follow-up or ‘endline’ data points – is crucial, but we argue that it is equally important to look ‘outside the black box’, at the broader context within which the development intervention is situated. As Stern *et al.* (2012: ii) point out: ‘most development interventions are ‘contributory causes’. They ‘work’ as part of a causal package in combination with other ‘helping factors’ such as stakeholder behaviour, related programmes and policies, institutional capacities, cultural factors or socio-economic trends. Designs and methods for IE [impact evaluation] need to be able to unpick these causal packages’.

Summing up this discussion, the need to unpack the causal chain from inputs to impacts – i.e. to explain observed outcomes – is recognised in recent contributions to the evaluation literature (White 2010; Woolcock 2009), especially given the complex context of development programming (Rogers 2008; Stern *et al* 2012), as is the need for finding credible alternatives

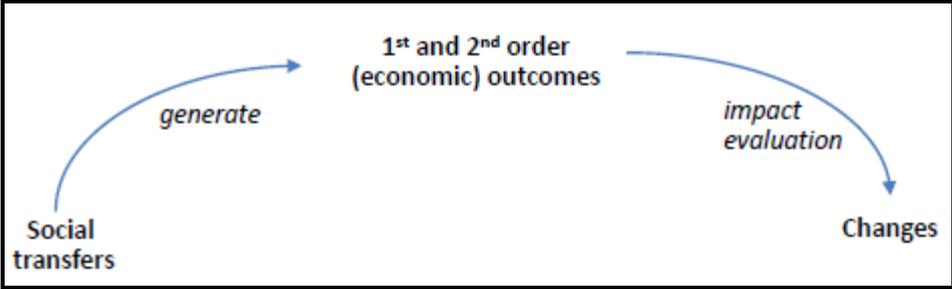
to expensive large-scale longitudinal surveys (White and Philips 2012). However, none of these discussions seeks to look ‘outside the box’ to consider outcomes that are more difficult to quantify and to build in mechanisms that allow lessons learned to feed back into improved programme design and delivery.

The discussion above does not suggest that no research on less measurable and non-quantifiable impacts of social transfers has been conducted. There is in fact a small but steadily expanding evidence base on such impacts (see, for example, Hochfeld and Plagerson 2011; Molyneux and Thomson 2011; MacAuslan and Riemenschneider 2011). Nevertheless, studies that consider the impact of social protection on social dynamics, inter- and intra-household relations, stigma and discrimination are rarely an integral part of evaluation frameworks and rigorously assessed alongside standard quantifiable impacts. The limited literature on social impacts of cash transfers has successfully identified and documented some effects of cash transfer programmes on intra- and inter-household dynamics. These evaluations have however usually focused on only one or two aspects and have failed to offer comprehensive and systematic analyses, and typically have neglected power and political dimensions completely. Also, they have been mainly static and descriptive, i.e. they have not integrated the feedback loops that can occur when changes in social interactions start influencing, in turn, the economic effects of the cash transfers. Because individual and collective decisions are based on perceptions and because people are embedded within social networks and operate on economic, political (power), cultural and social levels in relation to others, the changes induced in social dynamics as a result of cash transfers do in turn affect the economic effects of the programme.

3. Alternative evaluation framework

This section introduces our proposal for an extended evaluation framework, and illustrates how our framework differs from conventional approaches to impact evaluation. Figure 3.1 is a (stylised) graphical representation of the scope of the standard framework for evaluating social protection programmes.

Figure 3.1 Standard evaluation framework for social protection programmes



Source: Authors’ representation

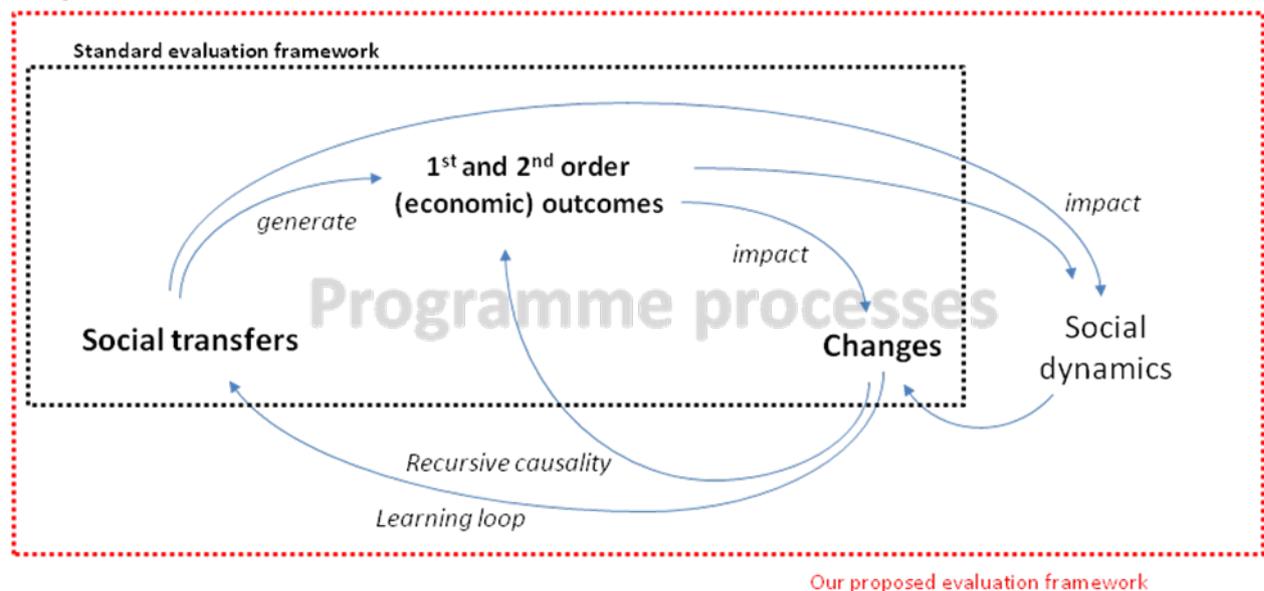
Social protection programmes aim to generate specified 1st and 2nd order (economic) outcomes, mainly related to poverty reduction, such as enhanced consumption (immediate) and/or investment in enterprises and human capital (long-term). Their theory of change is largely based on the assumption that poor households face binding resource constraints that undermine their material wellbeing (usually proxied by consumption) and their capacity to invest and accumulate assets and to access services such as education and health. Providing social transfers alleviates these resource constraints and triggers first- and second-order outcomes – an upward spiral of higher consumption, investment in livelihoods and assets for improved productivity and higher incomes, and stimulus to trade and economic activity. These, in turn, lead to changes such as reductions in the poverty gap or poverty

headcount and better integrated markets, reduced transactions costs, and income and employment multiplier effects. The measured effects are usually limited to quantifiable indicators such as consumption, school enrolment, dietary diversity and immunisation rates.

Based on the preceding discussion, we propose an innovative framework for evaluating social protection programmes that would strengthen and expand the evidence base in three distinct ways:

1. by recognising that programme design and processes – e.g. targeting criteria and delivery or payment mechanisms – can in themselves affect programme outcomes;
2. by incorporating the less measurable social and political impacts of social protection, including unintended positive and negative impacts, effects on social relations within and between households, especially between participants and non-participants, and the potential for empowerment of participants and wider communities;
3. by explicitly incorporating two feedback mechanisms that affect programme impacts: (i) an automatic feedback loop, and (ii) a deliberate learning loop.

Figure 3.2 Proposed alternative evaluation framework for social protection programmes



Source: Authors' representation

In line with the standard evaluation framework, the proposed alternative framework recognises that social protection programmes generate 1st and 2nd order outcomes, including increased consumption, higher incomes and a broader asset base, and that these lead to measurable impacts such as poverty reduction, improved educational outcomes and increased employment. However, its scope extends beyond that.

Firstly, the framework explicitly incorporates programme design and processes. Basic design and implementation choices, such as targeting criteria, delivery mechanisms and grievance procedures, can in themselves affect programme outcomes and impacts, not least in the extent to which these design choices and implementation features are empowering or disempowering of programme participants. They can reinforce or compromise the extent to which transfers generate economic effects – e.g. if targeting is inefficient, the effects will be small (Coady, Grosh and Hoddinott 2004) – and they can also impact on social dynamics –

e.g. if social transfers are delivered to women in male-headed households, this can lead to increased gender-based violence (Holmes and Jones 2013).

Secondly, the framework recognises that transfers may have an impact on social dynamics and other less measurable and more political issues such as stigma, discrimination and power dynamics underpinning these social dynamics, which can reinforce or undermine the desired changes that the transfers are to bring about. Delivering social transfers to poor households in poor communities has profound consequences for local social relations – it establishes a series of relationships (notably between participants and programme staff) and modifies pre-existing relationships (between participants and non-participants within communities, and between recipients and non-recipients within households), and social institutions (particularly informal risk-sharing mechanisms such as reciprocity). While some evaluations do ask about one or two of these aspects, the social impacts of social transfers are rarely – if ever – systematically and comprehensively evaluated.

Thirdly, the framework introduces two feedback loops: (1) ‘recursive causality’; and (2) a ‘learning loop’. The first feedback loop refers to an automatic mechanism whereby programme impacts are reinforced or undermined. For example, if social transfers improve participants’ health, this could increase productivity and reinforce economic effects. The second feedback loop refers to a deliberate mechanism whereby lessons learned about programme outcomes and impacts need to be fed back into programme design and implementation. Both feedback loops are largely left out of consideration in existing evaluation frameworks. The extent to which outcomes and impacts of social transfer programmes reinforce and undermine each other is rarely considered, not in the least due to the limited time frame of many evaluations. An explicit learning mechanism whereby evaluation results are fed back into programme design and implementation in a systematic rather than an *ad hoc* manner is also far from obvious in many frameworks.

Next we move on to discuss each of these three elements in more detail.

3.1 Programme processes

The rapid spread of social protection programmes in recent years, especially in sub-Saharan Africa and Latin America, has been accompanied by investment in a great deal of effort in the design, implementation and evaluation of each stage of the programme cycle, from choice of modality and payment levels to the identification and targeting of eligible participants, to payment delivery systems and assessment of programme impacts (Samson, van Niekerk and MacQuene 2006; Grosh, del Ninno, Tesliuc and Ouerghi 2007; EuropeAid 2012). All too often, however, these decisions are based on technical assessments or choices made by programme personnel operating with limited information, within financial and administrative constraints and political or ideological parameters. Very rarely do such decisions include considerations of how they may interact with programme outcomes and impacts.

Issues surrounding **delivery mechanisms** and **grievance procedures** illustrate the need for greater consideration of these issues in a holistic evaluation framework for social protection programmes.

Experiences with alternative payment delivery mechanisms are expanding rapidly, especially as new technologies allow for a switch from conventional ‘pull’ mechanisms, where participants queue for their transfers at designated pay-points on regular pay-days, to ‘push’ mechanisms, where payments are made directly into post office, bank accounts or through local cooperatives that participants can access at their convenience. Qualitative evidence suggests that innovative delivery mechanisms such as smart-cards or mobile phones are not only more efficient but also reduce the security risks associated with the movement and

management of cash, for both the implementation agency and the recipient. On the Mahatma Gandhi National Rural Employment Guarantee Scheme (MGNREGS) in India, where payments of wages are now made into participants' bank accounts in most places, there is evidence of a reduction in leakages (Chopra 2013). 'Push' mechanisms can empower participants, especially if women are the designated recipients and hence control access to household-level social transfers. Women in Malawi who were given smart-cards to withdraw cash transfers from a mobile bank saw the card as their first sign of official recognition, even as a kind of personal identification document (Devereux and Vincent 2010). On the other hand, a switch from 'pull' to 'push' mechanisms could lead to lost opportunities in terms of social interaction for participants and in the provision of complementary services by government and other agencies (Roelen, Edstrom, Sabates-Wheeler and Davies 2011).

Grievance mechanisms are gradually being introduced to social protection programmes in Africa and elsewhere. These innovations are more fundamentally important than they might appear, because they empower participants and introduce elements of rights and accountability to programme design and delivery that also interact with their potential outcomes and impacts. The credibility and effectiveness of such mechanisms depend crucially on how they are designed. For example, a 'Community Help Desk' was set up during implementation of the Cash and Food Transfers Pilot Project in Lesotho, but this facility was not independent of the project administration and appeared to be designed to explain decisions to complainants rather than to rectify errors (Devereux and Mhlanga 2008). The Hunger Safety Net Programme (HSNP) in Kenya established local 'Rights Committees' that monitor norms for delivering cash transfers that are specified in a 'Citizen's Service Charter' – e.g. payments should be made in full and on time, and all complaints should be addressed within 30 days of being lodged. Local residents in targeted districts (participants or excluded individuals) can also take any query or complaint to their District Social Protection Rights Coordinator (Devereux and White 2010). The Vision 2020 Umurenge Programme (VUP) in Rwanda piloted an 'Appeals and Complaints Process', underpinned by a 'Beneficiary Charter of Rights and Responsibilities', which specifies similar participant rights to those in Kenya's HSNP (Devereux 2011). Mozambique is currently strengthening its systems for implementation of social protection programmes, which will include designing community case management (Roelen 2012) and grievance procedures. In India, the employment guarantee scheme (MGNREGA) has been underpinned by social audit mechanisms that have helped to streamline the implementation processes, made people aware of their rights, improved accountability for programme delivery, and strengthened the social contract between the state and the rural poor (Chopra 2013).

Overall, these developments along the programme cycle are likely to strengthen the programme's intended outcomes and impacts, and for this reason an assessment of these design features and implementation processes should be included in any holistic evaluation.

3.2 Social dynamics and less measurable impacts

Evaluation reports on cash transfer programmes are dominated by material outcomes – income, consumption, asset accumulation, investment behaviour, human capital formation – because these are the areas where the programme implementers hope and expect to make a difference in poor people's lives that is measurable and attributable to the intervention. Despite an expanding evidence base, very little has been documented about the political effects of social protection programmes. One notable exception is a study of the impact of India's employment guarantee scheme on identity politics in Tamil Nadu (de Neve and Carswell 2011). Slightly more is known about the impact of cash transfers on social relations, within and between households and community members (women, men, youth, the elderly, traditional authorities), and especially between recipients and non-recipients of transfers. In particular, the impact of these programmes on informal social networks (reciprocal gift-giving and intra-community transfers, borrowing) can be significant and not always necessarily

positive (increased tension, conflicts, erosion of social cohesion) (Devereux and Getu 2013). However, the conditions under which formal transfers may complement and strengthen informal risk sharing and solidarity mechanisms – or conversely, ‘crowd out’ and weaken them – are still poorly understood.

Of particular importance is the interaction between cash transfers and existing informal risk-sharing mechanisms (reciprocal gift-giving and intra-community transfers, and borrowing). In Kenya and Malawi, some evidence was found of a decline in informal assistance to programme households (Hypher 2011). It is important, however, not to romanticise these informal networks and to recognise that they do not always benefit the poorest in communities – instead, they can entrench power imbalances and create poverty traps, by reinforcing gender, class or ethnic-based inequities. In such instances, the crowding out of informal mechanisms by cash transfer programmes may be desirable for ‘social transformation’ and ‘social justice’ reasons. But more research is needed around the relation between formal and informal transfers. Similarly, evidence that informal transfers do not always adequately substitute for public provision (Ellis 2006), for example, in situations of covariate shocks (such as climate change impacts that affect all local households simultaneously) is another dimension that requires further investigation.

The limited evidence in the literature on the effect of cash transfers on social and economic relationships within and between households and community members is mixed. Some assessments report positive effects on social capital (Attanasio, Fitzsimons, Gomez, Gutiérrez, Meghir and Mesnard 2009), self-confidence and social status (World Bank 2008), or suggest that cash transfers may help to reduce domestic violence (Thakur, Arnold, and Johnson 2009). Hypher (2011), for instance, examining the impact of cash transfers on social networks and social capital in several programmes in sub-Saharan African, found that cash transfers can strengthen social status, contribute to social capital and – through providing access to resources – can also help in building trust and community cohesion. Others note, however, that resentment and jealousy towards recipients of cash transfers are not uncommon (Ellis 2008; MacAuslan and Riemenschneider 2011; Roelen *et al*, 2011). When this happens, negative findings are often referred to through discrete comments on ‘stigma’, or ‘resentment’, and are presented as ‘unfortunate side effects’ that need to be dealt with within the framework of the intervention, without necessarily exploring the actual overall consequences of these secondary effects on the primary outcomes of the interventions.

A review of recent evidence calls for a more comprehensive analysis, and making the assessment of social dynamics and other social and political impacts, a more integral part of the evaluation framework. In the case of the Mchinji Social Cash Transfer Pilot project in Malawi, Miller, Tsoka and Reichert (2011) documented how cash transfers were partly used by participating households as gifts (food, soap, clothing, firewood), loans and wages, suggesting that the provision of cash had generated positive material effects and economic multipliers. However, the study also described how households in the programme declared feeling jealousy from other (non-recipient) households, with 26 per cent of cash recipients reporting that they experienced more conflict in their community since receiving the cash transfers. In Kenya, by contrast, when Ressler (2008) mapped the social impact of cash transfers she found that cash payments strengthened the social networks and social capital of participating households. In particular, these additional resources enabled recipients to participate in community events, share food and borrow when in need because they had more capacity to repay.

An evaluation of the Zimbabwe Emergency Cash Transfer programme (ZECT) showed that although cash transfers had positive material impacts on recipients’ health and education and on their ability to purchase basic goods and food, these transfers also created tension between recipient and non-recipient households (Kardan, MacAuslan and Marimo 2010). At the same time relations within households were reported to have improved – as men and

women were able to discuss how best to share and use the additional cash. However, whether social protection programmes contribute to women's empowerment is contested, not least because this outcome is notoriously difficult to measure. In a nuanced analysis of the gender impacts of conditional cash transfers, Molyneux and Thomson (2011) question claims that they have helped to empower the women and girls who are intended to benefit from receiving cash and enhanced access to education, respectively.

These studies usefully describe some of the complex matrix of inter-related social processes and mechanisms that are positively or negatively affected as a result of the introduction of cash transfers. What these studies do not provide is a dynamic analysis of these changes, in particular an assessment of how these changes might eventually affect the cash transfer programme itself. For example, conflicts or tensions – or even some form of solidarity among the participants – could impede or enhance the likelihood of successfully achieving the programme's objectives. In Ethiopia, for instance, the most recent evaluation of the Productive Safety Net Programme (PSNP) reports the intriguing fact that respondents are reluctant to specify who within their communities should benefit from the programme. This reflects a tension between the preferences of most communities for broader targeting (so that more households are included) and programme planners who would prefer larger transfers for a smaller segment of the population, in order to facilitate (and speed up) programme graduation (Hoddinott, Sabates-Wheeler, Berhane, Handino, Kumar, Lind, Seyoum Taffesse and Tefera 2012).

One objective of our alternative evaluation framework is to identify and assess the various social changes that are triggered by social protection programmes and cannot be adequately captured through the standard (mainly quantitative) methods used for monitoring and evaluation of these programmes. These are investigated at two levels – inter-personal social relations, and social mechanisms – and how these impacts affect programme outcomes.

3.3 Feedback loops

The alternative framework introduces two distinct feedback loops; an automatic feedback loop – 'recursive causality' and a deliberate learning loop.

Criticisms of purely quantitative approaches to impact evaluation are often couched in the context of evaluating 'complicated' or 'complex' programmes (Rogers 2008). Complicated programmes are described as having 'multiple components, multiple agencies, multiple simultaneous causal strands and/or multiple alternative causal strands' (Rogers 2008: 29). Such programmes could be broken down into their constituent components, with quantitative evaluation methods applied to some components where appropriate and alternative approaches applied to others, recognising that a uniform approach may not be appropriate to assessing all components. But in some instances this disaggregated approach may not be desirable – for example, large multi-dimensional programmes may need to be evaluated 'holistically'.

Complexity is anathema to standard impact evaluation models, which favour simple linear causal pathways. Yet in reality the trajectory from inputs and activities to outcomes and impacts is often uncertain and unpredictable in advance, even for social protection programmes with a simple design and a single objective. Complexity in programmes can be of various types: those with feedback or reinforcing loops ('**recursive causality**'); others with non-linear or 'tipping point' relationships; and those with 'emergent outcomes', where the outcomes of an intervention and 'the means to achieve them' emerge during implementation (Rogers 2008). Recursive causality and non-linear relationships may be addressed by experimental/ quasi-experimental and by statistical approaches (White 2010), or via 'good theory-based impact evaluation combining both quantitative and qualitative methods to

unpack the causal chain and uncover how complicated or complex social processes mediate impact' (White and Phillips 2012: 6).

The proposed evaluation framework seeks to explicitly recognise the complexity and non-linearity of social protection programmes. In a study appositely titled 'Broadening the range of designs and methods for impact evaluations', Stern *et al.* (2012) discuss the particular challenges faced when evaluating development projects, noting that evaluators tend to find long-term programmes that are 'embedded in a changing context with extended causal chains or pathways especially challenging' (Stern *et al.* 2012: 3). They draw our attention to the reality that development interventions (which would include social protection programmes) operate in a complex environment which makes isolating their impacts from the array of confounding factors highly challenging. It is because of this complexity that simplistic linear models of causality are inadequate and that an element of 'recursive causality' needs to be introduced in evaluation frameworks.

The deliberate **learning loop** refers to explicitly incorporating lessons learned about all facets of the programme and its impacts back into the programme design and implementation. This is not an automatic mechanism; deliberate action needs to be taken. Emergent systems, whereby outcomes can only be determined during the course of an intervention and not in advance, link to the analysis of complex, adaptive systems. The nature of complexity here points to the value of more empowering, participatory approaches to evaluation: 'local analysis should inform the analysis of the casual chain. And... findings and interpretation should be fed back to the local community in an interactive manner. These mechanisms are being put in place as impact evaluations move more strongly to a mixed methods approach' (White 2010: 162).

A particular element of such a learning loop refers to the inclusion of the opinions, experiences and perceptions of participants, non-participants and community members. It is increasingly acknowledged that the views of community members can bring about changes in behaviour and attitudes that ultimately impact on households' wellbeing and economic development, and that these social effects need to be more carefully assessed, especially in environments and communities where membership of a collective may be more important than individual advancement (Macours and Vakis 2009). In addition, it is recognised that incorporating the richness of personal experiences can increase the accountability of public spending and better communicate results to a wider audience to make public spending decisions and outcomes more transparent.

One example where participants' perceptions influenced programme design is the Productive Safety Net Programme in Ethiopia (PSNP). Initially conceived as an opportunity to shift social assistance away from food aid towards cash transfers, rapidly rising food prices in 2007 resulted in falling purchasing power and a shift in participants' preferences against cash and in favour of food, which retained its real value irrespective of price movements. Between two mixed methods evaluations in 2006 and 2008, more than one million out of seven million PSNP participants switched from cash to food transfers, largely because they requested this shift in payment modality (Sabates-Wheeler and Devereux 2010).

Another example comes from Malawi, where an evaluation of a cash transfer programme in 2006 found that many wives were dissatisfied with how their husbands used cash transfers, and they requested a switch to women being registered as cash transfer recipients, rather than household heads. This recommendation – based entirely on participant feedback – was adopted in a follow-up cash transfer programme in the same locality in 2007, and an evaluation confirmed that this switch did generate improved outcomes (Devereux 2008).

4. Methodological implications

Important methodological implications follow from the proposition that an alternative evaluation framework is needed for social protection programmes. The extension of its scope to include programme processes, social dynamics and other impacts and the inclusion of feedback loops introduces elements of complexity that make linear models of causality and the use of purely quantitative methods simplistic and inadequate. By the same token, although qualitative methods may be better equipped to incorporate non-linearity and complexity, they are also not sufficient to unpack and attribute social protection impacts.

Because quantitative and qualitative methods each have strengths and weaknesses, the case for 'mixed methods' seems overwhelming: their strengths complement each other while their weaknesses should cancel each other out. This is why mixed methods are sometimes described as 'Q-squared' approaches, or even as the 'platinum standard' (Khagram and Thomas 2010), trumping the 'gold standard' of RCTs. As Adato (2008: 222) explains, mixed methods are essential, not only to explain quantitative findings but to explore social impacts. 'Surveys provide reliable measures of program impacts... qualitative research provides explanations of why we do or do not find these impacts, and explores how social processes and social relations are affected by, and in turn shape responses to, the programs'. Following Carvalho and White (1997), these methods can be combined in different ways, including: (1) integrating quantitative and qualitative methodologies in data collection; (2) validating or triangulating information gained through one method by information gained through the other during data analysis; (3) merging findings from both approaches into a synthesised set of policy recommendations.

Perhaps surprisingly, given the tokenistic way in which qualitative methods are typically added (if at all) onto survey-based evaluations, there appears to be a consensus on the need for mixed methods in the mainstream impact evaluation community, as indicated by these selected comments:

The important task of investigating what works and what does not in the fight against poverty cannot be monopolized by one method.
(Ravallion 2009: 5)

Designs need to build on causal inference approaches each of which have their strengths and weaknesses, one of the reasons that combining designs and methods – so called 'mixed methods' – are valuable.
(Stern, *et al* 2012: i)

There are many settings when quantitative methods will be the best available method. But there are also many cases when they are not... a theory-based approach will usually combine methods: quantitative and qualitative and evaluation approaches'.
(White 2010: 162)

A particular element of a more holistic methodological approach to the evaluation of social protection programmes relates to the inclusion or exclusion of the opinions, perceptions and experiences of programme participants and non-participants (e.g. control group members). Their negligible participation in most programme evaluations, except as passive respondents, is regrettable on two levels. Firstly, at an instrumental level this means that valuable information is being overlooked that could greatly improve the implementation of these programmes, leading to enhanced impacts and more effective achievement of programme objectives. Secondly, at a normative level it is unfortunate that progressive practices in terms of participatory approaches to development have not yet been adequately incorporated into social protection programming.

The inclusion of participant and non-participant opinions and experiences is of both intrinsic and instrumental value. Eliciting participant perceptions and opinions can lead to new and important insights into how social protection programmes may (or may not) lead to first- and second-order effects, or help to identify unintended effects. It works towards validating theories of change as they are formulated in programme objectives. In addition, the notion that social protection is 'empowering' is implicit in many programmes, and is an inherent element of their theory of change. The inclusion of participants in the assessment of outcomes allows them to define what they see as empowerment and identify ways towards achieving this, rather than leaving assessors attempting to evaluate a predetermined notion of empowerment with externally defined indicators. This also ensures that participants are able to exert their agency and voice to shape the outcomes of social protection programmes.

Standard qualitative techniques to elicit views, experiences and perceptions of social protection programmes include in-depth interviews (household case studies, key informant interviews), focus group discussions (with programme participants and non-participants), and participatory methods (community wealth ranking, causal linkage diagramming, timelines). The underlying premise of participatory monitoring and evaluation (PM&E) is that M&E should be empowering rather than extractive – so outsiders can play a convening and facilitating role but programme participants are at the centre of the M&E process (Chambers 2008).

Other, less commonly utilised qualitative methods that should be considered in designing M&E systems include qualitative comparative analysis and social network analysis. Qualitative comparative analysis (QCA) is a way of studying cases systematically, allowing data to be interpreted qualitatively while also looking for causality between variables (Ragin 1997; Olsen 2004). QCA is used in comparative research and when using case study research methods. Data collected using qualitative methods are entered into a table where the rows are cases and the columns are attributes or variables, which are derived by coding the case study data using qualitative data analysis software such as NVIVO. Cases can be on multiple levels and are represented as a combination of causal and outcome conditions. Computer algorithms are then used to analyse these data by comparing the combinations of conditions that produce outcomes of interest. This approach retains the depth of qualitative research while drawing on the data handling capacity of quantitative methods.

Social network analysis (SNA) uses participatory visual mapping tools to characterise and analyse the social landscape of households and communities. These tools include social network mapping, community-level institutional mapping, individual life histories and community social histories. After social network architectures have been mapped and described, they can be examined in terms of types of networks, the functions each performs, their relation to inter-household and intra-community risk-sharing mechanisms (such as reciprocity) and how they reinforce social structures and power dynamics (Chopra 2011). In terms of impact evaluation, social network analysis can uncover how social transfer programmes affect resource flows between net donors and net recipients of formal and informal transfers, how these programmes modify relationships within and between households, and whether and how they modify social hierarchies and power inequalities.

All these methods provide complementary and appropriate tools to gain access to different perspectives (perceptions, opinions, experiences) of different individuals and social groups with respect to social protection programmes. Importantly, they can also be used to quantify actual and perceived changes that are attributable to social cash transfer programmes. These generic methods and specific instruments should be selected and adapted for each country context.

5. Discussion and conclusion

The positive impacts of social protection programmes are widely documented, with a rapidly expanding evidence base that is biased towards quantifiable and measurable indicators, following the need to demonstrate impacts, value-for-money and progress towards pre-defined targets such as the Millennium Development Goals. The ongoing 'professionalisation' of social protection programming throughout Africa, Asia and Latin America has been associated with an inexorable trend towards evaluations that attempt to mimic randomised control trials and track quantitative indicators that can be reported as percentage changes over time. Although we acknowledge the need for rigorous evaluations, we identify a number of shortcomings of existing evaluation frameworks for social protection programmes: they ignore programme processes by separating 'monitoring' from 'evaluation', they leave significant knowledge gaps in terms of less quantifiable impacts, and they overlook important sources of information that cannot be tapped into using survey-based methods. This paper has proposed an alternative evaluation framework that aims to address these shortcomings by explicitly taking into account programme processes, social dynamics and other less measurable and more political impacts and two distinct feedback mechanisms – 'recursive causality' and a learning loop. We argue that the integration of such elements in an evaluation framework would allow for a better understanding, and thereby improvement, of social protection impacts.

Given this context, this paper aims to offer a timely contribution to the evaluation literature that argues not only for 'mixed methods' to explain the numbers and assess selected non-quantifiable outcomes, but also for incorporating aspects of social protection programmes that are not normally considered in impact evaluations. We have proposed an alternative evaluation framework that extends beyond conventional approaches in three ways: firstly, by explicitly taking into account programme processes; secondly, by focusing on social dynamics and other less measurable social and political impacts; and thirdly, by introducing two feedback mechanisms – 'recursive causality' (how programme impacts might reinforce or undermine each other) and a 'learning loop' (to feed lessons learned back into programme design and delivery). Several under-utilised research methods and tools can be drawn on for conducting such a holistic evaluation. As discussed above, these include participatory monitoring and evaluation (PM&E) techniques, qualitative comparative analysis and social network analysis.

We argue that integration of these overlooked elements and innovative research methods in an expanded evaluation framework would not only allow for more nuanced understanding of social protection impacts, it would improve the design and delivery of programmes and generate the substantial reductions in poverty and vulnerability that social protection ultimately aims to achieve.

References

- Adato, M. (2008) 'Combining survey and ethnographic methods to improve evaluation of conditional cash transfer programs', *International Journal of Multiple Research Approaches*, 2008(2): 222-236
- Adato, M. and Hoddinott, J. (editors) (2010) *Conditional Cash Transfers in Latin America*. Washington DC: International Food Policy Research Institute
- Attanasio, O. Fitzsimons, E., Gomez, A., Gutiérrez M. I., Meghir, C. and Mesnard, A. (2010) 'Children's Schooling and Work in the Presence of a Conditional Cash Transfer Program in Rural Colombia', *Economic Development and Cultural Change*, 58(2): 181-210
- Baird, S., McIntosh, C. and Özler, B. (2010) 'Cash or Condition? Evidence from a Randomized Cash Transfer Program', *Policy Research Working Paper*, 5259. Washington DC: World Bank
- Banerjee, A. and Duflo, E. (2011) *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs
- Carvalho, S. and White, H. (1997) 'Combining the Quantitative and Qualitative Approaches to Poverty Measurement and Analysis: The Practice and Potential', *World Bank Technical Paper*, 366. Washington DC: World Bank
- Chambers, R. (2008) *Revolutions in Development Inquiry*. London: Earthscan
- Coady, D., Grosh, M. and Hoddinott, J. (2004) 'Targeting outcomes redux', *World Bank Research Observer*, 19(1): 61-85
- Chopra, D. (2011) 'Interactions of 'power' in the making and shaping of social policy', *Contemporary South Asia*, 19(2): 153-171
- Chopra, D. (2013, forthcoming) 'The Mahatma Gandhi National Rural Employment Guarantee Act, India: Examining pathways towards establishing right-based social contracts', *Journal of International Development*
- Davis, B., Gaarder, M., Handa, S. and Yablonski, J. (2012) 'Evaluating the impact of cash transfer programmes in sub-Saharan Africa: an introduction to the special issue', *Journal of Development Effectiveness*, 4(1) 1-8
- De Neve, G. and Carswell, G. (2011) 'NREGA and the Return of Identity Politics in Western Tamil Nadu, India', *Forum for Development Studies*, 38(2): 2005-2010
- Devereux, S. (2008) 'Innovations in the Design and Delivery of Social Transfers: Lessons Learned from Malawi', *Concern Worldwide Policy Paper*. London: Concern Worldwide
- Devereux, S. (2011) 2nd Annual Review of DFID Support to the Vision 2020 Umurenge Programme (VUP), Rwanda. Brighton: Institute of Development Studies
- Devereux, S. and Getu, M. (2013) *Informal and Formal Social Protection Systems in Sub-Saharan Africa*. Addis Ababa: Organisation for Social Science Research in Eastern and Southern Africa (OSSREA)

- Devereux, S. and Mhlanga, M. (2008) *Cash Transfers in Lesotho: An Evaluation of World Vision's Cash and Food Transfers Pilot Project*. Maseru: World Vision
- Devereux, S. and Vincent, K. (2010) 'Using Technology to Deliver Social Protection: Exploring opportunities and risks', *Development in Practice*, 20(3): 367–379
- Devereux, S. and White, P. (2010) 'Social Protection in Africa: Evidence, Politics and Rights', *Poverty and Public Policy*, 2(3): 53-77
- Department for International Development (DFID) (2011) *Cash Transfers Evidence Paper*. London: DFID
- Ellis, F. (2006) *Community Coping, Vulnerability and Social Protection*, Review Paper. Johannesburg: Regional Hunger and Vulnerability Programme (RHVP)
- Ellis, F. (2008) 'We are all poor here: Economic difference, social divisiveness, and targeting cash transfers in sub-Saharan Africa', paper prepared for the international conference, *Social Protection for the Poorest in Africa: Learning from Experience*. Entebbe, Uganda. 8-10 September
- Ellis, F., Devereux, S. and White, P. (2009) *Social Protection in Africa*. Cheltenham: Edward Elgar
- EuropeAid (2012) 'Social transfers in the fight against hunger: A resource for development practitioners', *Tools and Methods Series: EC Reference Document*. Brussels: European Commission
- Fiszbein, A. and Schady, N. (2009) *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington DC: World Bank
- Garcia, M. and Moore, C. (2012) *The Cash Dividend: The Rise of Cash Transfer Programs in Sub-Saharan Africa*. Washington DC: World Bank
- Grosh, M., del Ninno, C., Tesliuc, E. and Ouerghi, A. (2007) *From Protection to Promotion: The Design and Implementation of Effective Safety Nets*. Washington DC: World Bank
- Hagen-Zanker, J., McCord, A. and Holmes, R. (2011) *Systematic Review of the Impact of Employment Guarantee Schemes and Cash Transfers on the Poor*. London: ODI
- Hanlon, J., Barrientos, A. and Hulme, D. (2010) *Just Give Money to the Poor: The Development Revolution from the Global South*. Sterling VA: Kumarian Press
- Hickey, S. (2007) 'The politics of what works in tackling chronic poverty', *CPRC Policy Brief*, No. 5. Manchester: Chronic Poverty Research Centre
- Hochfeld, T and Plagerson, S. (2011) 'Dignity and Stigma among South African Female Cash Transfer Recipients', *IDS Bulletin*, 42(6): 55-59
- Hoddinott J., Sabates-Wheeler R., Berhane G., Handino M., Kumar N., Lind J., Seyoum Taffesse A. and Tefera M. (2012) *Implementing large scale food security programmes in rural Ethiopia: Insights from the Productive Safety Net Programme*. Washington DC: International Food Policy Research Institute
- Holmes, R. and Jones, N. (2013) *Gender and Social Protection in the Developing World. Beyond Mothers and Safety Nets*. London: Zed Books

- Hypher, N. (2011) 'Impact of cash transfers on children – the role of social relations and intra-household dynamics: findings from a Save the Children study', ODI lunchtime seminar. London: ODI. 6 May. [www.odi.org.uk/events/documents/2635-presentation-nicola-hyper.pdf]
- ILO (2010) *Effects of non-contributory social transfers in developing countries: a compendium*, Working Paper. Geneva: International Labour Office
- Kardan, A., MacAuslan, I. and Marimo, N. (2010) *Evaluation of Zimbabwe's Emergency Cash (ZECT) Programme*. Oxford: Oxford Policy Management
- Khagram, S. and Thomas, C. (2010) 'Toward a Platinum Standard for Evidence-Based Assessment by 2020', *Public Administration Review*, S100
- Macauslan, I. and Riemenschneider, N. (2011) 'Richer but Resented: What do Cash Transfers do to Social Relations?' *IDS Bulletin*, 42(6): 60-66
- Macours, K. and Vakis, R. (2009) 'Changing households' investments and aspirations through social interactions: evidence from a randomized transfer program', *Policy Research Working Paper Series*, 5137. Washington DC: World Bank
- McGregor, J.A. (forthcoming 2014) 'Human Wellbeing and Sustainability: Interdependent and Intertwined', chapter in E. Neumayer, G. Atkinson, S. Dietz and M. Agarwala (eds), *The Handbook of Sustainable Development*. Cheltenham: Edward Elgar
- Miller, C., Tsoka, M., and Reichert, K. (2011) 'Impacts on children of cash transfers in Malawi', in S. Handa, S. Devereux and D. Webb (eds), *Social Protection for Africa's Children*. London: Routledge
- Molyneux, M. and Thomson, M. (2011) 'Cash transfers, gender equity and women's empowerment in Peru, Ecuador and Bolivia', *Gender and Development*, 19(2): 195-212
- Olsen, W. (2004) 'Introduction to Qualitative Comparative Analysis', *Focusing on the Case Workshop series*, September 2004
- Ragin, C. (1987) *The comparative method: moving beyond qualitative and quantitative strategies*. Berkeley, University of California Press
- Ravallion, M. (2009) 'Should the randomistas rule?' *The Economists' Voice*, 6(2): 1-5
- Ressler, P. (2008) 'The Social Impact of Cash Transfers: A study of the impact of cash transfers on social networks of Kenyan households participating in cash transfer programs', *RENEWAL Working Paper*. Washington DC: IFPRI
- Roelen, K. (2012) *Referral Mechanisms and Case Management for Vulnerable Children in Eastern and Southern Africa: Mozambique Country Case Study*. Brighton: Institute of Development Studies
- Roelen, K., Edstrom, J., Sabates-Wheeler, R. and Davies, M. (2011) *Child and HIV sensitive social protection in Eastern and Southern Africa: Lessons from the Children and AIDS Regional Initiative (CARI)*. Nairobi: UNICEF ESARO
- Rogers, P. (2008) 'Using programme theory to evaluate complicated and complex aspects of interventions', *Evaluation*, 14(1): 29-48

- Sabates-Wheeler, R. and Devereux, S. (2010) 'Cash Transfers and High Food Prices: Explaining outcomes on Ethiopia's Productive Safety Net Programme', *Food Policy*, 35(4): 274–285
- Samson, M., van Niekerk, I. and MacQuene, K. (2006) *Designing and Implementing Social Transfer Programmes*. Cape Town: Economic Policy Research Institute
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. and Befani, B. (2012) 'Broadening the Range of Designs and Methods for Impact Evaluations – Report of a study commissioned by the Department for International Development', *DFID Working Paper 38*. London: DFID
- Stewart, S. and Handa, S. (2011) 'Reaching orphans and vulnerable children through cash transfers in Sub-Saharan Africa: simulation results from alternative targeting schemes', in S. Handa, S. Devereux and D. Webb (eds), *Social Protection for Africa's Children*. London: Routledge
- Thakur, S., Arnold, C. and Johnson, T. (2009) 'Gender and Social Protection', in *Promoting Pro-Poor Growth: Social Protection*. Paris: Organisation for Economic Cooperation and Development
- Thomas, J. and Grindle, M. (1990) 'After the Decision: Implementing Policy Reforms in Developing Countries', *World Development*, 18(8): 1163-1181
- Ulrichs, M. and Roelen, K. (2012) 'Equal opportunities for all? – A critical analysis of Mexico's Oportunidades', *IDS Working Paper 413/ CSP Working Paper 007*. Brighton: Institute of Development Studies
- White, H. (2010) 'A Contribution to Current Debates in Impact Evaluation', *Evaluation*, 16: 153
- White, H. and Phillips, D. (2012) 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework', *3ie Working Paper*, 15
- Woolcock, M. (2009) 'Towards a Plurality of Methods in Project Evaluation: A Contextualised Approach to Understanding Impact Trajectories and Efficacy', *BWPI Working Paper*, 73. Manchester: Brooks World Poverty Institute
- World Bank (2008) *For protection and promotion: The design and implementation of effective safety nets*. Washington DC: World Bank. [<http://go.worldbank.org/K0Z8SB4VJ0>]
- World Bank (2009) *Conditional cash transfers: reducing present and future poverty*. Washington DC: World Bank