



Review of evaluation approaches and methods used by interventions on women and girls' economic empowerment

Georgia Taylor and Paola Perezniето

Key messages

- Evaluations that explore how interventions (programmes or projects) lead to transformational change in women and girls' lives, enhancing their power, including control of resources, decision-making and agency, and not simply increased income, provide a more meaningful understanding of women and girls' economic empowerment.
- Mixed methods evaluations lead to more useful findings as they consider not only the effects of interventions but also the underlying reasons why they occur. This is especially important for Women and Girls' Economic Empowerment (WGEE) projects as changes in norms, attitudes and behaviours are difficult to fully understand with quantitative data alone.
- This review makes practical recommendations for evaluators and researchers to more adequately capture the impacts of interventions on the multiple dimensions of women and girls' economic empowerment.
- The review found a gap in knowledge around the economic empowerment of adolescent girls. Also, data and analysis in evaluations are not generally disaggregated by age or stages of the lifecycle.
- It is essential to undertake a robust context and gender analysis and to have a Theory of Change to guide the evaluation which builds on a holistic approach to Women and Girls' Economic Empowerment

Acknowledgements

We would like to thank the expert panel members – Nicola Jones, Linda Mayoux, Patrick Nolen and Caroline Pinder – for their comments and insights, and in particular for closely reviewing the shortlisted evaluations. Their insights informed this report.

Research assistance was provided by Karla Paola Avila, Agnieszka Malachowska, Jessica Jacobson and David Walker during the search phase. We would also like to thank Ella Page for coordinating the search phase and for her work during the screening process.

We recognise the useful inputs and support provided by Zoe Stephenson, Harri Lee and Clare McCrum from the Department for International Development (DFID).

We appreciate the feedback provided by the members of the reference group at various stages of the review process:

Elena Bardasi, Senior Poverty and Gender Economist, Independent Evaluation Group, World Bank

Anna Clancy, Evaluation Group, Office of Development Effectiveness, DFAT Australia

Beckie Dixon, Senior Adviser, Evaluation Department, NORAD

Nadja Dolata, former Gender and Diversity Adviser, Impact Assessment Team, Save the Children

Lindi Hlanze, Economic Adviser (economic empowerment of women and girls), Growth and Resilience Department, DFID

Naila Kabeer, Professor of Development Studies, School of Oriental and African Studies

Colleen Lowe Morna, CEO Genderlinks

Veronica Olazabal, former Evaluation and Learning, The MasterCard Foundation

Inga Sniukaite, Deputy Chief of Evaluation, UN Women

Table of contents

Acknowledgements	ii
Abbreviations	iii
Executive summary	iv
1 Introduction	1
1.1 Objective of the review	1
1.2 Definition of women and girls' economic empowerment	1
1.3 Analysis of research questions	2
2 Overview of review methodology	4
2.1 The three stages of the review process	4
3 Evaluation landscape	7
3.1 Analysis of reports included in the initial database	7
3.2 Trends and gaps in the evaluation landscape	12
3.3 Quality and coverage of the shortlisted reports	13
4 Elements of a good-quality evaluation to measure WGEE impact	15
4.1 Applying a holistic approach to women and girls' economic empowerment (WGEE)	15
4.2 Building an evaluation team with gender expertise	16
4.3 Preparing for the evaluation and understanding the context	17
4.4 Understanding the intervention – theories of change and frameworks	19
4.5 Methodology and indicators	22
4.6 Sampling	29
4.7 Tools and data collection	31
4.8 Quality of data analysis	31
4.9 Self-reflection	32
4.10 Use of findings and dissemination	33
4.11 Innovations relevant to women and girls' economic empowerment	33
5 Key WGEE findings	35
5.1 Financial services	35
5.2 Business development services	35
5.3 Trade, marketing, and collective action	36
5.4 Skills development	37
5.5 Social protection	37
5.6 Legal and regulatory frameworks and good-quality employment	38
5.7 Recommendations for designing WGEE projects	38
6 Recommendations	40
6.1 General considerations to improve WGEE evaluations	40
6.2 Building a team	40
6.3 Preparing for the evaluation and understanding the context	41
6.4 Understanding the intervention – theories of change and frameworks	41
6.5 Methodology, including indicators	42
6.6 Sampling	43
6.7 Tools and data collection	44
6.8 Analysis	44
6.9 Self-reflection	44
6.10 Report writing	45
6.11 Dissemination and use of evaluation and research study findings	45
References	46

Figures

Figure 1: Identifying which reports to include and exclude during the screening process	5
Figure 2: Methodological approaches used by the shortlisted reports	5
Figure 3: Number of reports included in the review database, by thematic area	7
Figure 4: The 10 main commissioning agencies	8
Figure 5: Number of studies using each methodology	9
Figure 6: Distribution of report methodologies by thematic area	10
Figure 7: Geographical coverage of reports included in the database by region	11
Figure 8: Share of reports with and without explicit evaluation questions	12
Figure 9: Scores for methodology used to assess Women's and Girls' economic empowerment	14
Figure 10: Theoretical framework – collective action	21

Tables

Table 1: Distribution of methodology used in the report, by commissioning agency	10
Table 2: Types of women and girls' economic empowerment outcomes identified	11

Boxes

Box 1: Interesting innovation: the use of an index based on social norms restricting women's behaviour	18
Box 2: When mixed methods work well	23
Box 3: Women's economic leadership: the story of Maminè Sanogo in Mali	36

Abbreviations

AusAID	Australian Agency for International Development
CCT	Conditional cash transfer
CT	Cash transfer
EPPI	Evidence for Policy and Practice Information
DAC	Development Assistance Committee
DCED	Donor Committee for Enterprise Development
DD	Difference in difference
DFAT	Department of Foreign Affairs and Trade
DFID	Department for International Development
FGD	Focus group discussion
GBV	Gender-based violence
IDI	In-depth interview
IDRC	International Development Research Center
IFPRI	International Food Policy Research Institute
ILO	International Labour Organization
J-PAL	Abdul Latif Jameel Poverty Action Lab
M&E	Monitoring and evaluation
MMAT	Mixed methods appraisal tool
OPT	Occupied Palestinian Territories
PSM	Propensity score matching
RCT	Randomised controlled trial
SEWA	Self-Employed Women's Association
USAID	United States Agency for International Development
WGEE	Women and girls' economic empowerment

Executive summary

1. Introduction

This review assesses the quality and effectiveness of evaluation methods and approaches used to analyse the effects of programmes or projects on women and girls' economic empowerment (WGEE). The objective of the review is to provide information to improve the design and commissioning of evaluations to deliver stronger positive impacts across the different dimensions of women and girls' economic empowerment.

1.1 Definition of women and girls' economic empowerment

For the purposes of this review we have defined economic empowerment as a process whereby women's and girls' lives are transformed from a situation where they have limited power and access to economic assets to a situation where they experience economic advancement and their power and agency is enhanced in the following areas (Veneklasen and Miller, 2002)¹ (change outcomes):

- **Power within:** the knowledge, individual capabilities, sense of entitlement, self-esteem, and self-belief to make changes in their lives, including learning skills for jobs or start an enterprise.
- **Power to:** economic decision-making power within their household, community, and local economy (including markets) not just in areas that are traditionally regarded as women's realm but extending to areas that are traditionally regarded as men's realm.
- **Power over:** access to and control over financial, physical and knowledge-based assets, including access to employment and income-generation activities.
- **Power with:** the ability to organise with others to enhance economic activity and rights.

Economic empowerment will only be possible and sustainable if there are changes at different levels: within the individual (capability, knowledge and self-esteem); in communities and institutions (including norms and behaviour); in markets and value chains; and in the wider political and legal environment (Golla et al., 2011). We define this as a 'holistic' approach to WGEE because it is not just looking at whether women and girls have increased their access to income and assets, but also whether they have more control over them and are able to use them to have greater control over other areas of their life.

1.2 Scope of the review

The review analysed evaluations that assessed some measure of women and girls' economic empowerment in one or more of the following eight thematic areas: Financial services; Business development services; Skills training; Asset provision (both financial and physical); Social protection; Unions and fair employment; Trade and access to markets; Regulatory and legal frameworks

2 Overview of methodology

The methodology drew on and adapted systematic review principles in order to ensure a comprehensive search of the literature and to identify as many relevant resources as possible. There were three stages in the review process:

- 1) **Initial literature search:** The search strategy was designed to identify evidence and evaluations from across disciplines that report outcomes for women and girls' economic empowerment. The systematic search process used consistent inclusion criteria, search terms, coding, and search locations. A total of 382 papers were uploaded into the database for review. Only the 254 that reported empirical research (excluding systematic reviews, literature reviews and theoretical papers) were subsequently analysed for quality in the next stage.

¹ The outcome areas have been adapted to be relevant to economic empowerment, not just overall empowerment.

2) **Narrowing the field of inquiry:** To narrow down the field from the 254 empirical reports identified in the first stage, the research team devised a two-stage process using (i) a 'basic scoring tool' that captured the extent to which there was information on the methodological approaches used in the report and points were given according to the level of information available – 160 papers scored 2 or more as they contained at least basic information on methods and (ii) a purpose-designed 'quality assurance tool' - 70 documents passed this quality assurance test by scoring 2 or more.

3) **Review:** Out of the 70 documents shortlisted, there was at least one in each of the eight thematic areas. They covered a range of methodological approaches, but most used quantitative (32) and mixed methods (26), with only 11 using qualitative methods alone. Shortlisted reports were reviewed by a panel of four WGEE evaluation experts, using a review template that contained questions derived from evaluation best practice, as well as key resources on women's economic empowerment. The individual reviews and the synthesis each expert wrote analysing the key methodological issues identified in the reports they reviewed provided the basis for our analysis, findings and recommendations.²

3 Evaluation landscape

254 empirical evaluations and studies were found to have an analysis (to varying levels of depth) of economic empowerment of women and girls. Out of the 254 reports of empirical research the majority (117) were of financial services projects. In many cases these were mixed with other thematic areas. The area of financial services is well evaluated. Business development services, skills training, social protection and trade and access to markets all had around the same number of evaluations (between 28 and 37 each). Only six evaluations were found to focus exclusively on unions and fair employment and one on regulatory and legal frameworks.

The ILO had commissioned the highest number of evaluations (17), with other evaluations coming from Oxfam, AusAID, DFID, World Bank, Care, USAID, Population Council, IFPRI and the Fairtrade Foundation. There were few reviews exploring WGEE found commissioned by key donors such as DFID and the World Bank. There was a fairly even spread between quantitative (77) and mixed method (71) evaluations, with a higher number of qualitative evaluations (97). South Asia had the highest representation of evaluations (35%), followed by sub-Saharan Africa (30%) and Latin America and Caribbean (11%) and East Asia and Pacific (9%).

The main gap in coverage in relation to evaluation areas was the limited number of reports looking at the economic empowerment of adolescent girls – with hardly any evaluations disaggregating data by age. In addition there were very small numbers of evaluations of projects working on legal and regulatory frameworks, unions and fair employment and asset provision. Quite a large number of evaluations were of low quality – with just under half of them lacking any evaluation questions. It is also important to note that only 70 reports were assessed as having both enough detail in their evaluation or research methodology and a minimum quality standard in the methodological approach used to make it to the review stage; that is only 28% of the reports initially identified.

4 What makes a good-quality evaluation that effectively measures WGEE impact?

This section analyses key dimensions of 'quality' in the approaches used by the 70 shortlisted evaluations or study documents that were reviewed by the expert panel. It gives examples from specific evaluations to illustrate strengths or weaknesses in relation to the quality criteria defined during the review phase. It draws on information from the expert reviews, and builds on insights derived from the meta-analysis of the set of reviews.

² The four experts on the panel have significant experience of issues of gender equity and women's empowerment, as well as robust knowledge of quantitative and qualitative research methodologies. Nevertheless, when assessing the quality of a report, panel members were influenced by their individual expertise, knowledge base and methodological preference. This review presents a balanced analysis based on careful reading of all the individual reviews and the analytical syntheses prepared by each expert. For more details on reflections about the review process, please see Annex 3.

Only around 30% of evaluations applied a holistic approach to women and girls' economic empowerment (WGEE). Many of the reviewed evaluations had a partial approach to WGEE and looked at change in terms of economic advancement only.

Evaluations that demonstrated a holistic approach to women and girls' economic empowerment included the following: Ahmed, et al (2009), Garikipati (2008), Arif et al. (2010), Baden (2013), Bureau of Applied Research in Anthropology (2013) and Jones and Shaheen (2013). However, there are varying levels of sophistication of the WGEE frameworks used. Evaluations assessed as 'good' measured changes in the individual's decision-making power within the household as a measure of women's agency. The **measurement of change in norms and attitudes** was generally not done at institutional level or at the level of the legal and policy enabling environment.

Fewer than 25% of the reports reviewed undertook a thorough contextual gender analysis. Gender differences, gender-related norms and behaviours, gender roles and gender relations differ in every context, and assumptions about these should be scrutinized when designing an evaluation. In some cases, an evaluation did aim to measure women's economic empowerment but there was a lack of understanding of the importance of the multiple gender inequality dimensions present, and how they related to the intervention's achievements, and its evaluation.

The reviews identified that evaluation teams and researchers often make underlying assumptions and use stereotypes of women's activities, which can limit the level of analysis. The type of questions that are asked of women and men can be gender specific – such as type of expenditure, type of decisions and sector for economic activities.

Evaluations that scored well used a multidimensional range of indicators to measure women and girls' economic advancement and changes in their power and agency. The most innovative studies used unique variables to measure power that captured aspects of economic empowerment that one might not normally consider, such as whether a young girl was less likely to have unwanted sex (e.g. Bandiera et al., 2012).

44 % of the 70 reviewed evaluations had a high quality theory of change. Most of the highest-scoring evaluations and studies had a well-articulated theoretical framework and/or theory of change. This enabled the evaluators to demonstrate where and how change would happen as a result of the intervention – and thereby to frame their evaluation around that change process. The best theories of change showed dynamic and multifaceted change processes that reflected transformational changes in women's agency and economic advancement, and the relationship between them.

Mixed methods evaluations were the most effective in terms of measuring change and outcomes, and also provided good material for learning and improving future project design. They were effective in various ways: they captured change in more diverse ways, exploring not only what change occurred but why and how it occurred; they also added understanding to quantification on complex gender dynamics and added confidence about causality. A good mixed methods approach means that qualitative and quantitative data can complement each other, so that the whole is more than the sum of its parts. It also allows for effective triangulation of data. In low-resource cases, qualitative or quantitative data alone would be useful as long as the evaluation design was robust. From the mixed methods evaluations, those that scored higher used both the quantitative and qualitative data to complement and reinforce each other, and each component of data collection was effectively sequenced.

A large number of reviewed qualitative evaluations had limited scope and reach, with small samples. This suggests a lack of value and funding attached to qualitative methods. In many of the studies reviewed, it is unlikely that the point of data saturation was reached.

In many cases, evaluations were measuring progress against a pre-agreed set of indicators that appeared to be at output level only, with no analysis of unexpected or wider changes that might have taken place. Measurement of change at outcome or impact level of the log frame would have required a set of

indicators that (given the context) adequately measured transformational change on women and girls' economic advancement and power and agency.

Hardly any of the reviewed documents had included the complete set of evaluation tools used, either in the main report or annexes; only in some cases, sets of questions were included in footnotes (e.g. Bandiera et al., 2012). Some reports with a robust methodological approach at the conceptual level had poor design and implementation of instruments, which led to poor-quality data and weak analysis. Data collection methods sections in the reports reviewed rarely discussed gender issues with respect to methodology and enumerators/interviewers, for example, explaining how the sex of the interviewer could influence responses in certain contexts.

Many of the evaluations reviewed did not conduct interviews, surveys or focus group discussions with men and/or young men unless the project was specifically targeted at men as well. It is difficult to measure women's economic empowerment unless men's attitudes and behaviours are also taken into account. This was despite the fact that the engagement of men was generally seen as critical in most programmes/evaluations for gaining acceptance of the strategies chosen to improve women's economic empowerment.

There were very few evaluations of interventions targeting girls or young people's economic empowerment. Reports were found for some such programmes but were excluded because they were either poor quality and/or lacked detail. In the other evaluations, lack of age-disaggregated data meant that impact on women and girls was not separated out, rather just categorised as 'married' or 'unmarried' even though there are both young women and girls who are married with different experiences depending on their age group. There was limited evidence of differentiated approaches being used for targeting women/girls of different ages in programmes many programmes and/or their evaluations.

Though women's voices were captured in most evaluations, data analysis and presentation of quantitative evaluations hardly ever reported verbatim women's voices, which is not only useful evidence, but also conveys findings more powerfully.

5 Key findings on WGEE from the evaluations

There were a number of relevant findings from the shortlisted evaluations. Overall, the most interesting finding is that joint/mixed or integrated interventions providing economic skills and services alongside life skills and other training services (including gender training) appear to deliver the best results for women's economic empowerment.

Financial services

Reports looking at microfinance present a similar picture: microfinance for women increases access to credit for businesses but has no effect on social outcomes. Access to credit enables businesses to grow – leading to higher levels of assets and profits (in some cases). Furthermore, households that have access to microfinance are able to smooth consumption. However, the majority of studies looking at microfinance as a single intervention found that it has no effect on female bargaining power in the household, expenditure on female goods or on children, or even on female health outcomes like more knowledge of how to avoid sexually transmitted infections (STIs), or a decrease in unwanted pregnancies. In many cases, though, microfinance is not offered as a stand-alone intervention. Other components, such as business training are built into it and are often not considered within the evaluation.

Business development services

Most of the evaluations reviewed combine business development with microfinance or other activities such as marketing and collective action. Findings in this area were diverse given the very distinct programmes evaluated. For example, an Oxfam evaluation (2012) of a project providing livelihood (agriculture) and trade support showed that there had been an impact on women's ability to influence decisions made by producer associations; improvements in attitudes towards women's economic roles; increases in productivity and

improvements in farming practice. However, there was no increase in women's decision-making power at the household level, or in their income. An interesting model of 'microfinance plus' business development is the WORTH programme in Nepal (Valley Research Group and Mayoux, 2008), which combines business development services, banking and literacy. The evaluation demonstrated high levels of loans and increases in the size of village banks; improvements in self-confidence among women; a decrease in gender-based violence and an increase in the value and quality of social capital. Positive effects on women's economic empowerment were largely dependent on how robustly the different components of these interventions were implemented together.

Trade, marketing, and collective action

There were several interesting evaluations looking at trade (including fair trade) and collective action. At least three of them showed important effects on WGEE, although few of them had WGEE as an objective. For example, Nelson and Smith's (2011) evaluation of Fairtrade cotton projects showed a positive impact on women's economic empowerment, including greater control over cotton income, more women's representation in producer organisations, and more cotton produced by women farmers. Bacon's (2010) evaluation of the Nicaraguan Fair Trade (FT) coffee cooperatives found that the FT/organic (male and female) co-op reported higher empowerment levels than the conventional co-op, but lower than the women-only FT co-op; they attributed this to men occupying all the leadership positions in the FT/organic co-op, while fewer women in this co-op held land titles. However, although members of the co-ops demonstrated a strong sense of collective empowerment, individual interviews revealed uneven gendered empowerment processes in the households and more diverse gender relationships.

Skills development

Skills development consisted of courses that participants had to attend for a period of time –from one to five days a week. Women would receive vocational training (such as sewing skills) or lessons on life skills (e.g. negotiating safe sex). These life skills components prove very important and can be replicated in many countries with appropriate adjustments to reflect contextual realities. However, the type of vocational programmes aimed to give women specific skills to make money and lift themselves out of poverty (e.g. sewing or picking rice, for instance) will vary across contexts; the efficacy of the programme will depend on the appropriateness of the skills taught. Overall, the training programmes seem to work and have a positive effect on women's economic empowerment. When the two components – vocational and life skills – are packaged together, there can be strong results in both areas (labour market and life skill outcomes). These types of programmes can also be cost effective and have the potential to play a crucial role in WGEE.

Social protection

Most of the social protection evaluations reviewed explored conditional and unconditional cash transfer (CT) programmes. Reports reviewed suggest that approaches focused on individuals or households are unlikely to have significant community-level effects unless there are also community-level interventions – especially around countering discriminatory gendered social norms. This will make social protection programmes more responsive to gendered needs. Some of the evaluations explored changes in women's decision making power in the household as a result of becoming programme recipients, and what this means for better targeting in practice. In the case of social protection programmes, it is important to think about transfer amounts and balance of inputs, including complementary services, as these can have a major bearing on programme impacts.

Legal and regulatory frameworks

Out of the two evaluations of interventions focusing on legal and regulatory frameworks reviewed, one pertains particularly to how new property and inheritance rights increased access to land for widows and children (Uwayezu and Mugiraneza, 2011). Another study, by the World Bank and the IFC (2011), demonstrates how legal and regulatory frameworks for Special Economic Zones affect women's employment and rights in the workplace. However, the study also found that women's participation in business ownership and

entrepreneurship in these zones was limited due to legal and cultural constraints, and women still experienced high levels of harassment in the workplace.

Recommendations for designing WGEE projects

Women and girls' economic empowerment should result in women attaining better decision-making power over the economic areas of their lives – and so must include processes that free women from control, coercion and violence. Likewise, better incomes and economic stability/security can enable women to live independently from abusive men. However, evidence from this review suggests that economic interventions alone are not sufficient to economically empower women (though they may increase their income). This implies that an integrated approach to programming may be very important. It is also important to reflect more about 'the complexities of people's lives' – for instance, female-headed households may still be beholden to or dependent on male partners in various ways, positive and negative.

6 Recommendations

The main justification for allocating funds to evaluation from implementation of a WGEE intervention is to yield practical recommendations for increasing its contribution to women's economic empowerment and for learning within the sector. With this in mind recommendations have been made here to improve evaluation quality and learning potential and to encourage the uptake of evaluation findings to inform the design and implementation of programmes:

Gender analysis

Ensure there is a theory of change that describes transformational change in women and girls' economic advancement and power and agency in the short term, medium term and long term and at the level of individuals, communities, institutions, and the legal and policy environment.

Build a team that has sufficient gender expertise and involve this from the beginning at evaluation design and context analysis. Evaluations should aim to include experts with solid experience and expertise in researching and evaluating gender and economic empowerment/enterprise.

Undertake context analysis that has a full gender analysis and that covers different levels of analysis – including the economic, political, institutional and technical context and the status of economic empowerment in relation to social norms and attitudes.

Research methods

Base indicators and tools to be used in the evaluation on local knowledge. Conduct prior and follow-up participatory and/or qualitative research to identify appropriate local indicators and ways of wording questions so that they are likely to be clearly understood and taken seriously by interviewees and thus encourage the most reliable responses. Below are some suggested measurements that could be captured by indicators:

- transformational change in women's economic advancement and agency
- attitudes, norms and behaviours of women and men
- change in gender mainstreaming and attitudes in institutions
- gender differences in empowerment between women and men to increase local relevance
- changes in men and women's social capital, economic autonomy and political participation
- effects of the wider market, value chain and employment, as all of these will affect women's economic advancement and empowerment

Use mixed methods effectively and ensure that indicators adequately measure transformational change. Ensure that the quantitative and qualitative components are well designed and able to complement each other, either through a sequenced approach in which findings from one inform the design of the other, or by ensuring that each approach collects complementary data. The evaluation should aim to establish whether change took place and what caused it, so quantitative methodology might be required. However, qualitative information will

be needed to understand the details of the process of change, and how implementation has been experienced by individuals, as well as to uncover any unintended impacts or changes and other factors that may have influenced change. This is especially important for WGEE projects as changes in norms, attitudes and behaviours are difficult to fully understand with quantitative data alone.

Design appropriate tools for different study participants and provide them in full in the research report.

The team should design appropriate tools and data collection methods for young people and different population groups in the sample. This is particularly important when collecting sensitive data from adolescent girls and boys. Participatory tools can work well with this cohort. In order to demonstrate effect and especially impact in the absence of baseline data, there should be greater use of **historical tools** – e.g. life histories, generational comparisons, community histories.

Disaggregate data fully by age, sex and other relevant groups. Include women and men, and, where appropriate, girls and boys in the sample – even if the intervention is targeted at women and girls only. Evaluators and researchers should analyse the difference, the co-dependence between the groups, the positive outcomes, and relationships between men and women – not just the outcomes for women. Ensure differentiation between women of different ages and their stages in the life cycle, as well as women from different backgrounds.

Provide a self-analysis of the quality of the research and the resultant data and analysis in the evaluation. Include in the report details on the composition of the evaluation and data collection team and how this may have influenced results. There should be an explanation of what worked well and what did not work well and how political, economic or natural situations and crises may have influenced the findings of the evaluation.

For commissioning agencies

Design evaluation frameworks before project implementation begins in order to include a baseline and maximise the adequacy of the methodology and the use of the findings.

Commission and undertake more quality evaluations that explore impacts on WGEE across all the thematic areas. In particular, it is important to commission more evaluations of projects in areas that have few evaluations currently – such as projects focusing on adolescent girls, projects that use employment and unions and/or the legal and regulatory environment to increase women and girls' economic empowerment.

Qualitative methods should be included more frequently in evaluation commissioning processes, and funding for qualitative evaluations, particularly as part of mixed methods research, should increase to enable greater quality and more robust evaluation and research findings.

The ultimate aim of an evaluation is for findings to be used by decision makers. **Provide concrete programme or policy recommendations in the report**; this could be included in the TOR for the evaluation. The report's recommendations could provide clear guidance for how to enhance impact on women and girls' economic empowerment – and this should include policy implications.

1 Introduction

1.1 Objective of the review

The objective of this review is to improve the design and commissioning of evaluations of interventions (programmes or projects) on women and girls' economic empowerment (WGEE). It is based on an extensive analysis of evaluation approaches for interventions that have direct or indirect effects on WGEE (the dissemination strategy and target audience for the review are described in Annex 1). The review aims to contribute to more effective programming that delivers stronger positive impacts across the different dimensions of women and girls' economic empowerment. The main methods used to conduct this review included:

- an extensive search to identify evaluations across a range of WGEE technical areas (see Section 1.4)
- statistical analysis of the range of evaluations and identifying gaps in evaluation coverage, in terms of thematic and geographical areas, and evaluation questions and methods
- an analysis of the rigour and appropriateness, and the strengths and weaknesses, of the approaches and methods of selected WGEE project evaluations against pre-established criteria based on monitoring and evaluation (M&E) good practice
- presentation of key findings from selected good-quality evaluations across a range of thematic areas.

1.2 Definition of women and girls' economic empowerment

The report adopts the following definition of WGEE: Economic empowerment is a process whereby women and girls' lives are transformed from a situation where they have limited power to a situation where their power is enhanced in the following areas (VeneKlasen and Miller, 2002)³ (change outcomes):

(1) The knowledge, individual capabilities (e.g. business skills or financial literacy), sense of entitlement (understanding and recognition of their own subordination and their own sense of rights), self-esteem, and self-belief to make changes in their lives, including learning skills to get a job or start an enterprise (**power within**).

(2) Economic decision-making power within their household, community, and local economy (including markets) not just in areas that are traditionally regarded as women's realm but extending to areas that are traditionally regarded as men's realm (**power to**).

(3) Access to and control over financial, physical and knowledge-based assets, including access to employment and income-generation activities (**power over**).

(4) The ability to organise with others to enhance economic activity and rights – including cooperatives, unions, group-based financial services, marketing and learning opportunities (**power with**).

Empowerment is thus a process of change that transforms women and girls' lives in these four areas and interacts with resources (pre-conditions), agency (process), and achievements (outcomes) (Kabeer, 1999). Economic empowerment will only be possible and sustainable if there are changes at different levels: within the individual (capability, knowledge and self-esteem); in communities and institutions (including norms and behaviour); in available resources and economic opportunities; and in the wider political and legal environment (Golla et al., 2011). We define this as a 'holistic' approach to WGEE because it is not just looking at whether they have increased their access to income and assets, but also whether they have more control over them and are able to use them to have greater control over other areas of their life.

³ The outcome areas have been adapted to be relevant to economic empowerment, not just overall empowerment.

Scope of the review

We undertook an extensive search of evaluations that included some measure of women and girls' economic empowerment and covered one or more of the following thematic areas:

- 1) **Financial services:** including business finance (credit, equity, guarantees), microfinance, micro-insurance, savings – including interventions using new technologies such as mobile phones.
- 2) **Business development services** e.g. accountancy, human resources, marketing.
- 3) **Skills training:** including financial literacy, job and industry-related skills development.
- 4) **Asset provision:**
Financial assets: including but not limited to grants, financial guarantees, remittances
Non-financial assets: including but not limited to income from work, goods, land and property, productive assets from which an income can be generated (including livestock, technology, and information and communications technology (ICT)), consumer durables, IT equipment, energy-saving devices, etc.
- 5) **Social protection:** including cash transfers, public works, childcare services.
- 6) **Unions:** such as the Self-Employed Women's Association (SEWA) in India, and other group-based activities to promote access to fair employment.
- 7) **Trade and access to markets:** fair trade, ethical trade, cross-border trade support, trade regulations and agreements.
- 8) **Regulatory and legal frameworks:** including for asset and business ownership and control, as well as gender equity regulatory frameworks to support an enabling environment for women's economic activities.

The search covered evaluations and research reports that either focused on women and girls' economic empowerment or included some measure of women and girls' economic advancement and/or agency. It did not cover thematic areas in relation to infrastructure (e.g. roads and travel), water and sanitation, education, or health – even though these areas have a substantial impact on economic activity. The aim was to focus on projects that explicitly aim to increase women's and/or girls' (and men's) economic and financial security, including income-earning potential and initiatives, enterprise, trade, enterprise-related collective action, and financial safety nets.

A desk review of WGEE theoretical frameworks was used to develop a sample theory of change (see Annex 2) and to develop a common understanding of WGEE among the review experts and the rest of the team. This forms the basis of much of our analysis. It was combined with a further literature review of general M&E good practice and guidelines in order to develop an understanding of what a 'good' WGEE evaluation might look like.

1.3 Analysis of research questions

The research questions used in this review are detailed below, and Section 2 describes the methodology used to answer them. We used a series of intermediary questions to screen evaluations, developing codes for each of the sub-questions, which were then analysed using EPPI Reviewer.⁴

1. How rigorous and appropriate are the evaluation methodologies for addressing the main evaluation questions?
2. What are the strengths and weaknesses of the different evaluation approaches?
Sample sub-questions:
 - Does the evaluation measure change women and/or girls' economic advancement and/or agency?
 - Is there sufficient information in the document about theory of change, evaluation questions, methodology, analysis, and use of findings?
 - Is the evaluation 'good quality' in relation to defined good practice criteria?
 - Is the evaluation using a comprehensive definition of WGEE that goes beyond economic advancement to include transformational dimensions, including changes in women's agency?
3. What are the gaps and trends in terms of types of programmes evaluated, geographical coverage, evaluation questions and methods?
Sample sub-questions:

⁴ Evidence for Policy and Practice Information (EPPI) Reviewer database management software www.eppi.ioe.ac.uk.

-
- Which country (or countries) and/or region(s) are covered by this evaluation?
 - Which thematic area is the evaluation covering?
 - What target group(s) is the evaluation covering or missing?
4. What are the key findings of the evaluations reviewed, including findings on the economic and social impacts of the interventions?

2 Overview of review methodology

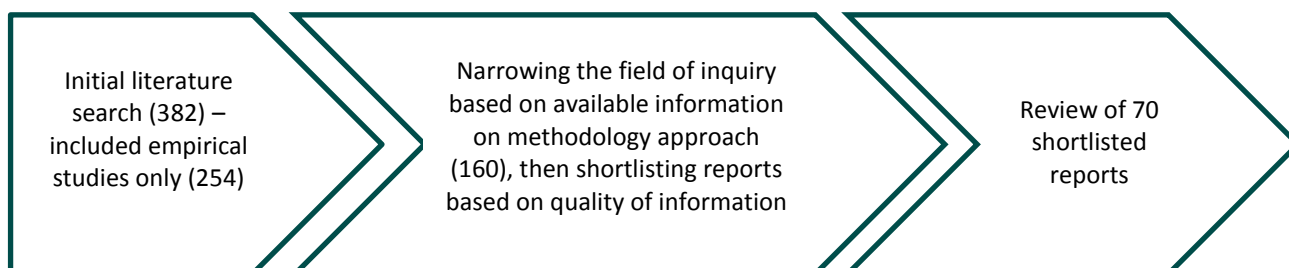
The methodology has drawn on adapted systematic review principles in order to ensure a comprehensive search of the literature and to identify as many relevant resources as possible. Full details of the methodology can be found in Annex 3.

Since the objective of the review was to explore evaluation methodologies and approaches, rather than findings (which are generally the focus of reviews), the tools used at every stage were adapted to capture the information needed to answer the research questions.

The review uses a broad definition of ‘evaluation’ and includes research designs that may not explicitly be described as evaluations but aim to show the impact or effects of a project, programme or policy (an ‘intervention’⁵) on women’s and/or girls’ economic empowerment. As such, in addition to evaluations, a number of research reports, appraisals and assessments were included in the database. These used a variety of methodological approaches, including but not limited to: qualitative and participatory evaluations; ex-ante and post-hoc evaluations; quasi-experimental project designs; and randomised controlled trials (RCTs). The reason for adopting this broader definition was to find a range of interesting and innovative methodologies that are being used to assess the impacts of interventions on WGEE.

2.1 The three stages of the review process

The review process consisted of three stages:



- 4) **Initial literature search:** The search strategy was designed to identify evidence and evaluations from across disciplines that report outcomes for women and girls’ economic empowerment. This includes evaluations that did not explicitly target women and girls’ economic empowerment, but had wider economic empowerment and wealth creation objectives. The search was largely web-based and designed to cover both grey and academic literature, without privileging any particular discipline. The systematic search process is based on the principles of transparency and replicability. It used consistent inclusion criteria, search terms, coding, and search locations, as detailed in Annex 3.

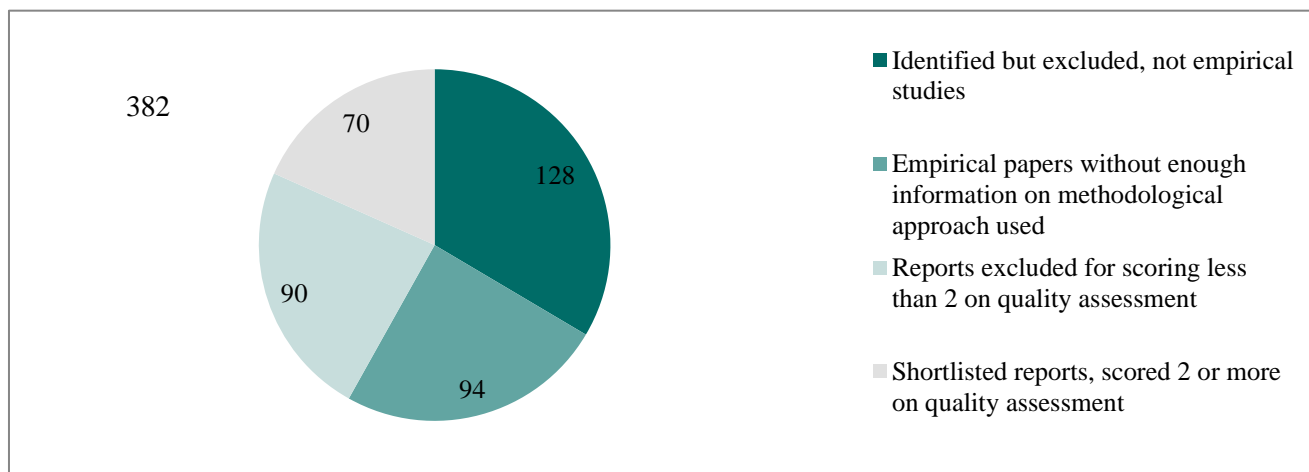
A total of 382 papers were uploaded into the database for review. Only the 254 that reported empirical research (excluding systematic reviews, literature reviews and theoretical papers) were subsequently analysed for quality in the next stage. The team reached ‘saturation point’ based on the search methodology, drawing on web searches as well as liaising with experts and relevant contacts who sent useful, often unpublished evaluations.

- 5) **Narrowing the field of inquiry:** To narrow down the field from the 254 empirical reports identified in the first stage, the research team devised a two-stage process using two different tools. The first ‘basic scoring tool’ captured the extent to which there was information on the methodological approaches used in the report (see Annex 4). Reports that scored 2 or more points on that tool (160 of them) were then assessed using a purpose-designed ‘quality assurance tool’. This was based on the principles of the Mixed Methods Appraisal

⁵ Other agencies may use the term ‘intervention’ differently, but we use it here to include policies, programmes and projects for simplicity.

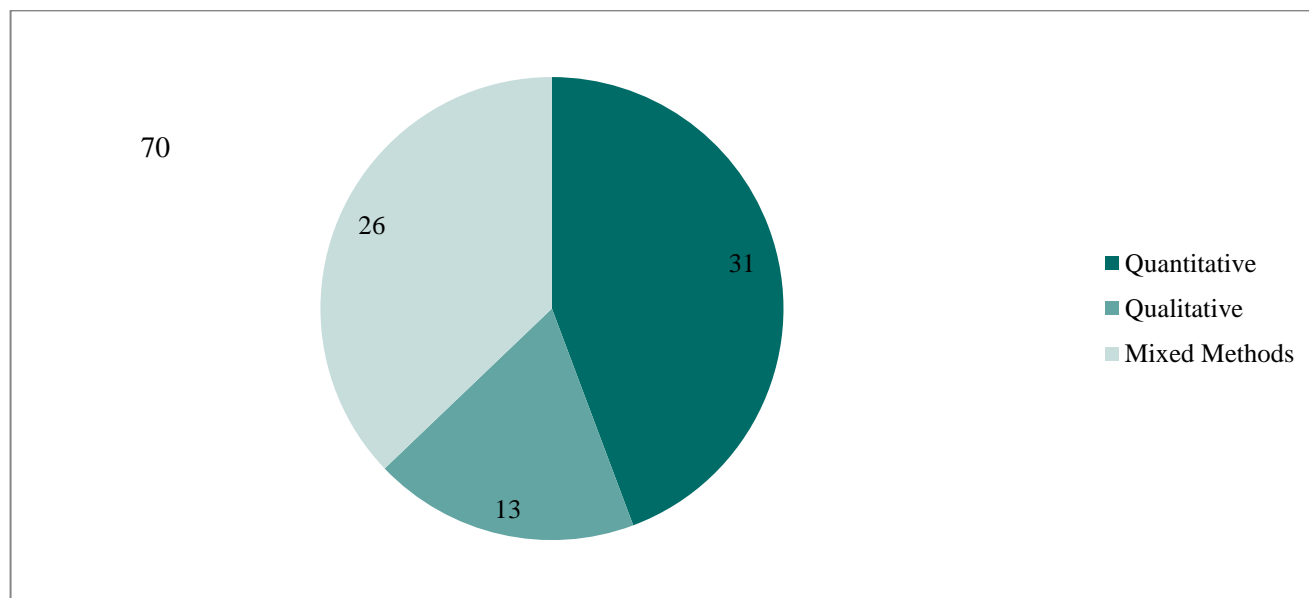
Tool (MMAT) developed by Pluye and colleagues (2011), which is designed for the appraisal stage of systematic reviews that include qualitative, quantitative and mixed methods studies. This tool gives a rapid assessment of quality based on basic criteria set out for each methodological approach used. It provided us with a light touch analysis through a rapid review of the documents, enabling further in-depth screening (see Annex 5 for details). Reports that scored 2 or more points on this second tool were considered for the final stage of the review. In total, 70 documents 'passed' the quality assurance test.

Figure 1: Identifying which reports to include and exclude during the screening process



6) **Review:** Out of the 70 documents shortlisted, there was at least one in each of the eight thematic areas. They covered a range of methodological approaches, but most used quantitative (31) and mixed methods (26), with only 13 using qualitative methods alone.

Figure 2: Methodological approaches used by the shortlisted reports



The lower representation of qualitative methods among the shortlisted reports (despite a more balanced representation in the overall database) might reflect the fact that qualitative evaluations and studies tend to have less detail about their research methodology than quantitative reports, and thus might have been excluded in stage two of the screening process on this basis.

Shortlisted reports were reviewed by a panel of four experts.⁶ (profiles of the experts can be found in Annex 6), using a review ‘template’ (see Annex 7) that contained questions derived from evaluation best practice, as well as key resources on women’s economic empowerment – for example, the Development Assistance Committee (DAC) criteria for evaluating development assistance,⁷ Bond’s quality of evaluation checklist⁸, DFID’s evaluation policy (DFID, 2013), and Kabeer’s (1999) paper, on reflections on the measurement of women’s empowerment. The purpose of the review template was to enable the expert panel to make a robust analysis of the quality of methodologies used, particularly in relation to their ability to capture relevant information about the effects of the intervention on women and girls’ economic empowerment. The template required experts to ‘score’ each report’s methodology using a scale of 1 (highest) to 5 (lowest), providing a separate score for the methodology used to assess economic empowerment (if this was different). Reviewers were also required to analyse each report’s content in terms of the strengths, weaknesses, and appropriateness of the intervention, as well as its understanding of economic empowerment of women and/or girls, and its contribution to the overall literature. Given resource constraints, it was only possible for the reports to be analysed by one reviewer, according to her/his thematic and methodological area of expertise⁹. The individual reviews and the synthesis provided by each expert form the basis for our analysis and findings (sections 4 and 5) and recommendations (Section 6)

The next section looks at the evaluation ‘landscape’, with information on the types of programmes evaluated, thematic areas covered, range of commissioning agencies, and other key factors.

⁶ The four experts on the panel have significant experience of issues of gender equity and women’s empowerment, as well as robust knowledge of quantitative and qualitative research methodologies. Nevertheless, when assessing the quality of a report, panel members were influenced by their individual expertise, knowledge base and methodological preference. This review presents a balanced analysis based on careful reading of all the individual reviews and the analytical syntheses prepared by each expert. For more details on reflections about the review process, please see Annex 3.

⁷ ‘DAC Criteria for Evaluating Development Assistance’ (www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm).

⁸ <http://www.bond.org.uk/effectiveness/principles>

⁹ With more resources, it would have been ideal to have two reviewers look at each report to have two perspectives on the scores.

3 Evaluation landscape

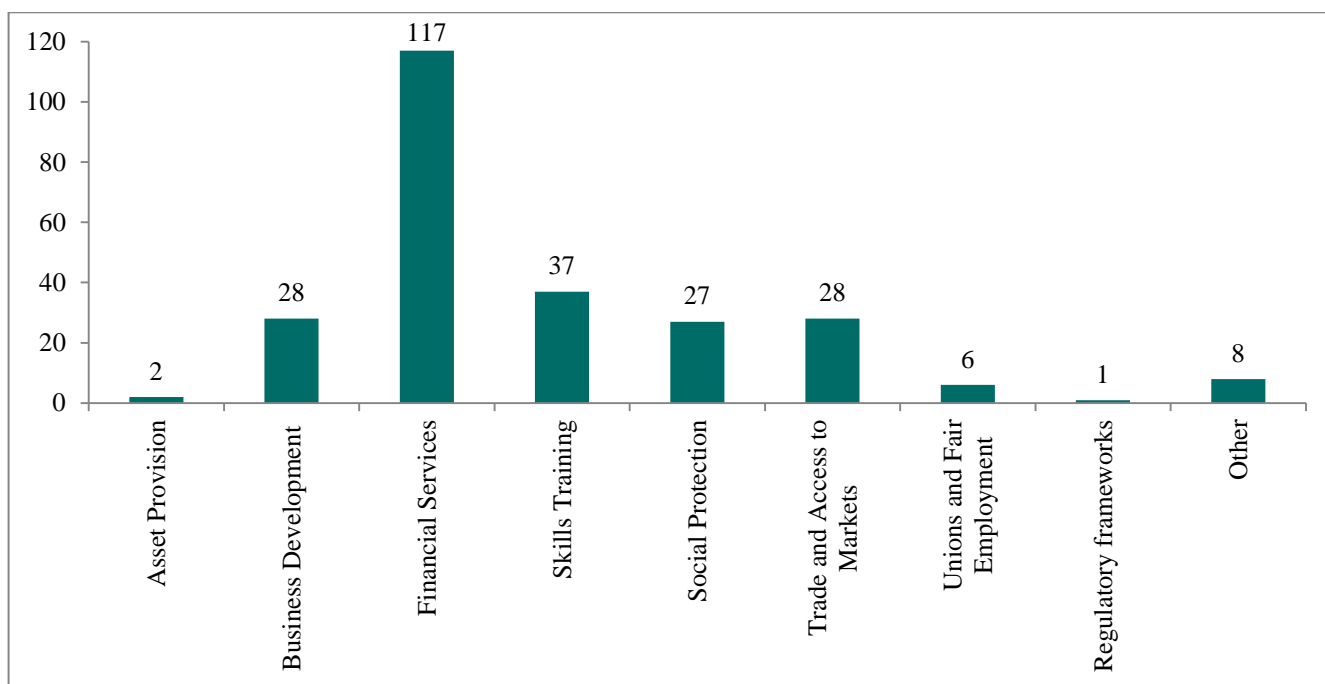
As explained in Section 2, the systematic search conducted in stage one of the review process resulted in 254 empirical evaluations and studies being included in the database. The analysis below presents the key trends and gaps found across these reports to assess the 'landscape' for evaluations of women and girls' economic empowerment (WGEE).

3.1 Analysis of reports included in the initial database

3.1.1 Types of programmes evaluated

The reports covered interventions that did not always fit into just one of the eight thematic areas chosen for the review. Figure 3 (below) shows how many reports came under each thematic area according to the main component of the intervention. However, more than half of the reports (54%) featured interventions that spanned more than one component – for example, 'financial services' interventions often also had elements of skills training; similarly, 'business development' interventions generally combined elements of financial services, trade and access to markets, and so on. Eight reports were classified as 'other', including those researching access to ICTs as a form of economic empowerment and livelihood strengthening in the context of climate change.

Figure 3: Number of reports included in the review database, by thematic area



Just over half of the reports (134) evaluated financial services interventions (either as a primary or secondary theme), including microfinance, micro-insurance, savings, and self-help groups. Of these reports, 49 looked exclusively at financial services interventions while the remaining 85 included elements from other thematic areas, with microfinance most commonly paired with skills training. Despite the high number of reports covering financial services interventions, there were many similarities in the methodological approaches used, particularly in quantitative evaluations.

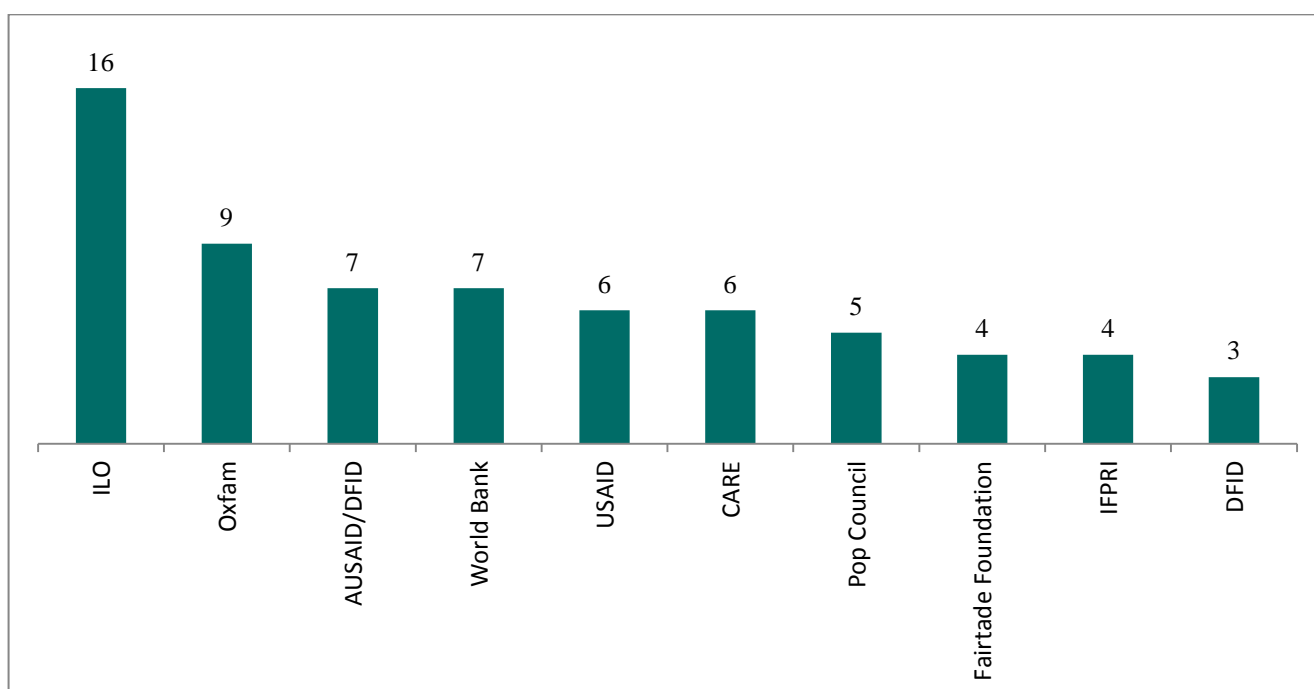
Some of the other thematic areas were well represented in the database, with 37 reports on skills training, 28 on business development, 28 on trade and access to markets, and 27 on social protection. This suggests there are a good number of evaluations and studies in these fields that explore, directly or indirectly, the economic empowerment of women and girls. But the three remaining areas (asset provision, unions and fair employment, and regulatory frameworks) were under-represented. This is partly because interventions in these technical

areas are usually implemented in parallel to others, and thus relevant reports might be included in the database but under a different thematic area as the main component. For example, none of the evaluations included in the database focused exclusively on the effects of regulatory frameworks, while there were only two reports that focused on regulatory frameworks as the main intervention; however, regulatory frameworks were a complementary theme in 16 of the studies included in the database.

3.1.2 Who is commissioning evaluations on women and girls’ economic empowerment?

The reports included in the review were commissioned by a number of different agencies (see Figure 4). A substantial number (71) did not specify the commissioning agency, probably because they were internally rather than externally commissioned. Twenty four reports were linked to academic institutions, including reports by independent researchers who conducted their own studies as well as studies commissioned by or affiliated to specific universities. From those that did specify affiliation, no one academic institution had commissioned more than one of the papers included.

Figure 4: The 10 main commissioning agencies



Some commissioning agencies were donors; others were non-government organisations (NGOs). The International Labour Organization (ILO) had commissioned most of the evaluations that included dimensions of women and girls’ economic empowerment. With regard to the Department for International Development (DFID), seven reports that were part of a study on gender and social protection co-funded by DFID and the Australian Agency for International Development (AusAID) were identified through the search criteria. In addition to those, only three other evaluations that met the inclusion criteria had been commissioned by DFID. The authors expected more evaluations and studies by major donors such as DFID and the World Bank exploring WGEE specifically to be identified and shortlisted.

Reports by research institutes such as the Abdul Latif Jameel Poverty Action Lab (J-PAL) and the International Initiative for Impact Evaluation (3ie), as well as NGOs (Concern, Save the Children) were also included in the database, but only one or two from each met the screening criteria. This suggests that the agencies that commission evaluations are paying limited attention to exploring the impacts of interventions on women and girls’ economic empowerment.

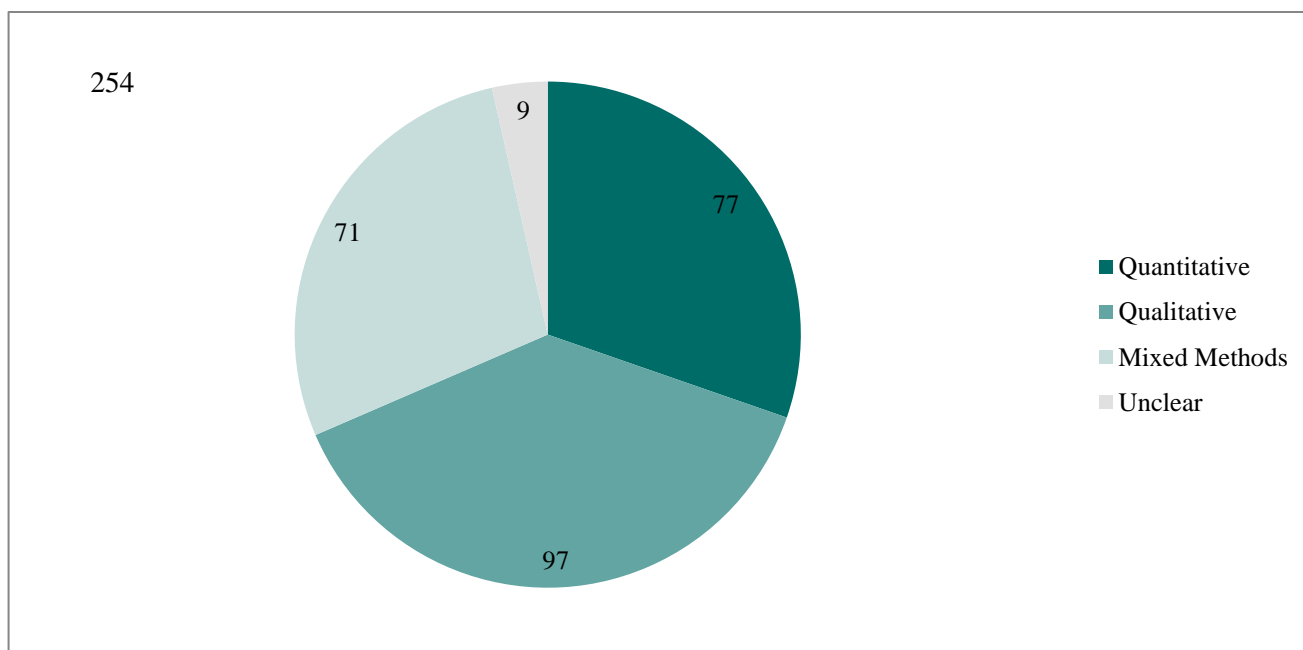
While the initial search and screening processes were done systematically, there might have been biases in the identification of resources as a result of the techniques used. For example, snowballing from certain documents might have led to specific journal articles or webpages and not to others. However, as key donors and

individuals in the main evaluation centres were contacted, we can be confident that there is good representation of commissioning agencies in relation to evaluation reports that explore economic empowerment of women and girls.

3.1.3 Types of methodologies used

Most reports included in the database used exclusively qualitative methods (97), followed by those that used exclusively quantitative methods (77), and those that used mixed methods (71). In the case of 9 reports, the types of methods used were not specified (unclear). This distribution is in contrast to the distribution of the shortlisted reports (see Section 2), of which 32 used solely quantitative methods, 26 used mixed methods, and 11 used qualitative methods. This difference can be explained by the fact that qualitative reports tend to have a very weak description of the methodology used, and so many of them were excluded during stage two. In contrast, more detailed explanation of the methodology used is a standard feature of quantitative reports.

Figure 5: Number of studies using each methodology



The qualitative methods used in the reports were standard: mainly focus group discussions and in-depth interviews. Of the 97 qualitative evaluations and studies included, only 24 used more complex participatory methods, ethnography, case studies, etc. Of the 76 quantitative evaluations and studies included, there was some variation in the tools used, which ranged from RCTs (37), quasi-experimental design (49) (with propensity score matching (PSM)/difference in difference (DD)), analysis of secondary quantitative data (19), and longitudinal studies (6), while some studies used more than one quantitative tool simultaneously (for example, a longitudinal RCT).

There was significant variation in how well these methods were applied, as well as the level of detail provided in each report about the methods used, which raises an important point about the need to set minimum standards for reporting on evaluation methodologies used. Qualitative methodologies in particular are reported with much less rigour and detail than quantitative methodologies. (This is explained in greater detail in Section 4.)

The use of mixed methods was common (by 71 out of the 254 studies). However, the weight of the quantitative and qualitative components in these evaluations varied. Some were essentially quantitative reports that had included a few focus group discussions and in-depth interviews to ‘tick the box’ of having included a qualitative component, while others were essentially qualitative reports that had a small quantitative component that was not sufficient to provide robust quantitative data findings.¹⁰ Some of the best examples of documents identified –

¹⁰ Because we only conducted an in-depth analysis of the 70 shortlisted reports, we cannot quantify how many of the 71 mixed methods reports had an imbalanced used of quantitative vs. qualitative methods, but this was observed during the screening process in the next stage.

many of which were shortlisted – had a well-integrated mixed methods design, planned the sequencing of each method so as to maximise their combined value, and then triangulated data for more robust analysis.

Figure 6 shows the distribution of methodologies used in the reports included in the database by thematic area. ‘Financial services’ has the most quantitative evaluations (40), with a relatively balanced number of mixed methods (39) and qualitative (37) studies. There is a similar balance found in most of the thematic areas, except for ‘unions and fair employment’ and ‘trade and access to markets’, where there are far fewer quantitative evaluations.

Figure 6: Distribution of report methodologies by thematic area

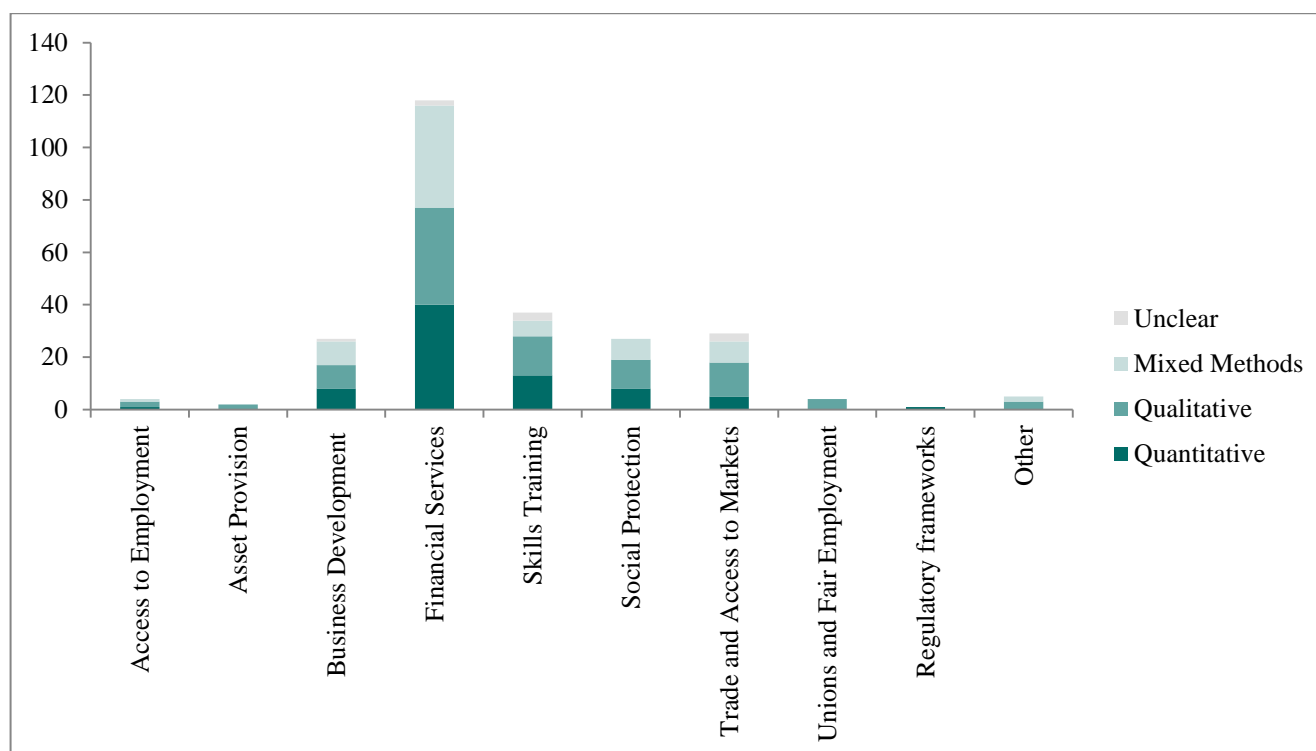


Table 1 shows distribution of methodologies used to evaluate interventions for women and girls’ economic empowerment by commissioning agency. It illustrates that the World Bank clearly has a preference for quantitative evaluations; the United States Agency for International Development (USAID) seems to commission more mixed methods evaluations and studies; while DFID evaluations seem to be mainly qualitative – though as this is a very small sample in relation to other evaluations commissioned, the findings cannot be generalised. There was a relative balance between the methods used in reports commissioned by other agencies.

Table 1: Distribution of methodology used in the report, by commissioning agency*

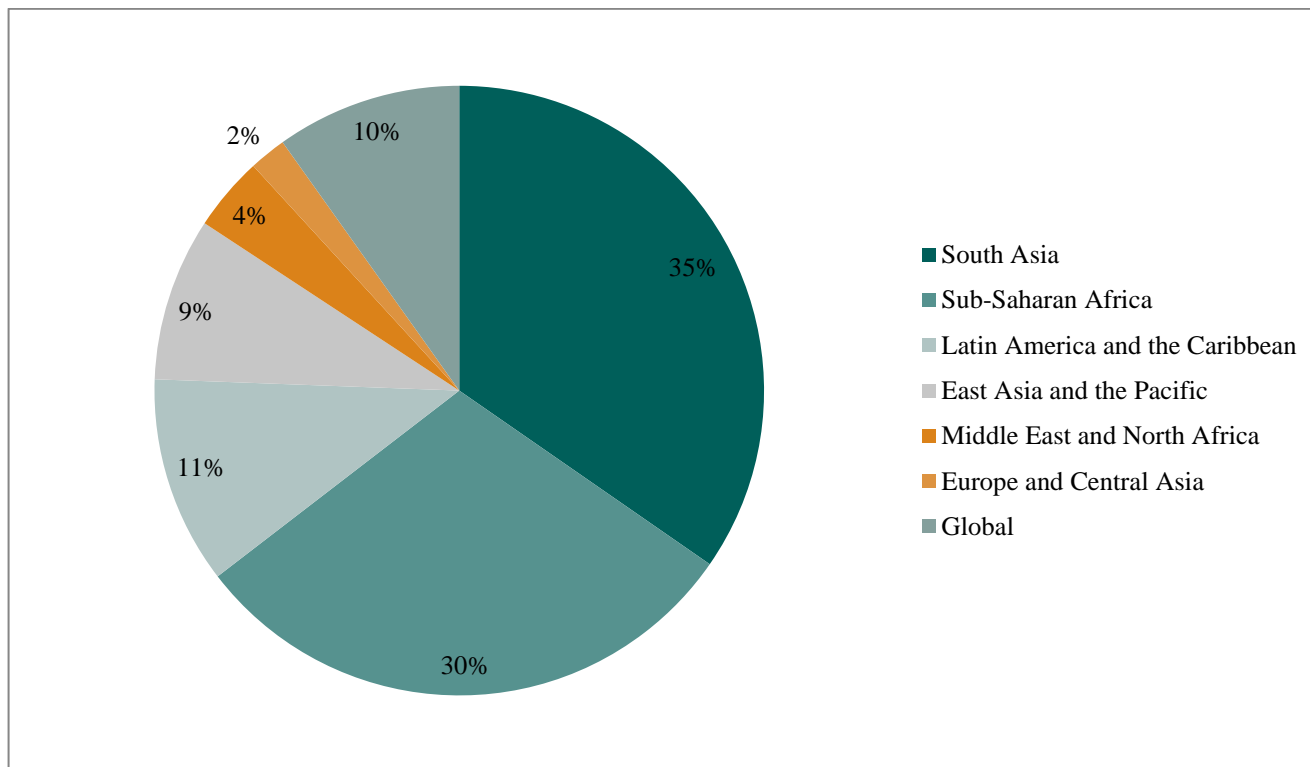
Theme	Quantitative	Qualitative	Mixed methods	Not clear
ILO	0	9	4	3
Oxfam	1	3	4	1
AusAID/DFID	0	3	4	0
World Bank	5	1	1	0
USAID	0	0	6	0
Care	1	2	3	0
Population Council	2	0	3	0
Fairtrade Foundation	0	4	0	0
IFPRI	2	1	0	0
DFID	0	2	0	1

* The table only includes the 10 agencies with most commissioned evaluations in the database, so there are fewer reports considered.

3.1.4 Geographical coverage

Reports identified during the search and included in the database primarily covered interventions on women and girls' economic empowerment in South Asia (35%) and sub-Saharan Africa (30%).

Figure 7: Geographical coverage of reports included in the database by region



Within South Asia, most of the evaluations took place in India (56) and Bangladesh (33), focusing on financial services. Studies in sub-Saharan Africa were distributed across the region, with most reports looking at interventions in Uganda (29), Kenya (20), Tanzania (15), Ethiopia (14), Ghana (10), and Malawi (10). Several studies covered more than one country, and there was also a broader thematic variation in the interventions being evaluated. In East Asia, Vietnam was the country with the most evaluations (12).

3.1.5 Women and girls' economic empowerment outcomes

The reports in the database give fairly rich information about the types of women and girls' economic empowerment outcomes that were measured (see Table 2). Many projects addressed a range of economic empowerment issues, tackling enterprise-related and market-related barriers alongside those around gender inequality in institutions and communities. Other evaluations concentrated on developing a better understanding of the economic advancement aspects of women's economic empowerment.

Table 2: Types of women and girls' economic empowerment outcomes identified

WGEE outcomes	Number of documents coded under this WGEE * outcome**
Economic advancement	142
Individual capabilities	134
Economic decision-making power	104
Ability to organise with others	64
Access to finances	45
Access to markets	25
Improved regulatory frameworks	15

* From the framework 'Theory of change for women and girls' economic empowerment', Annex 2

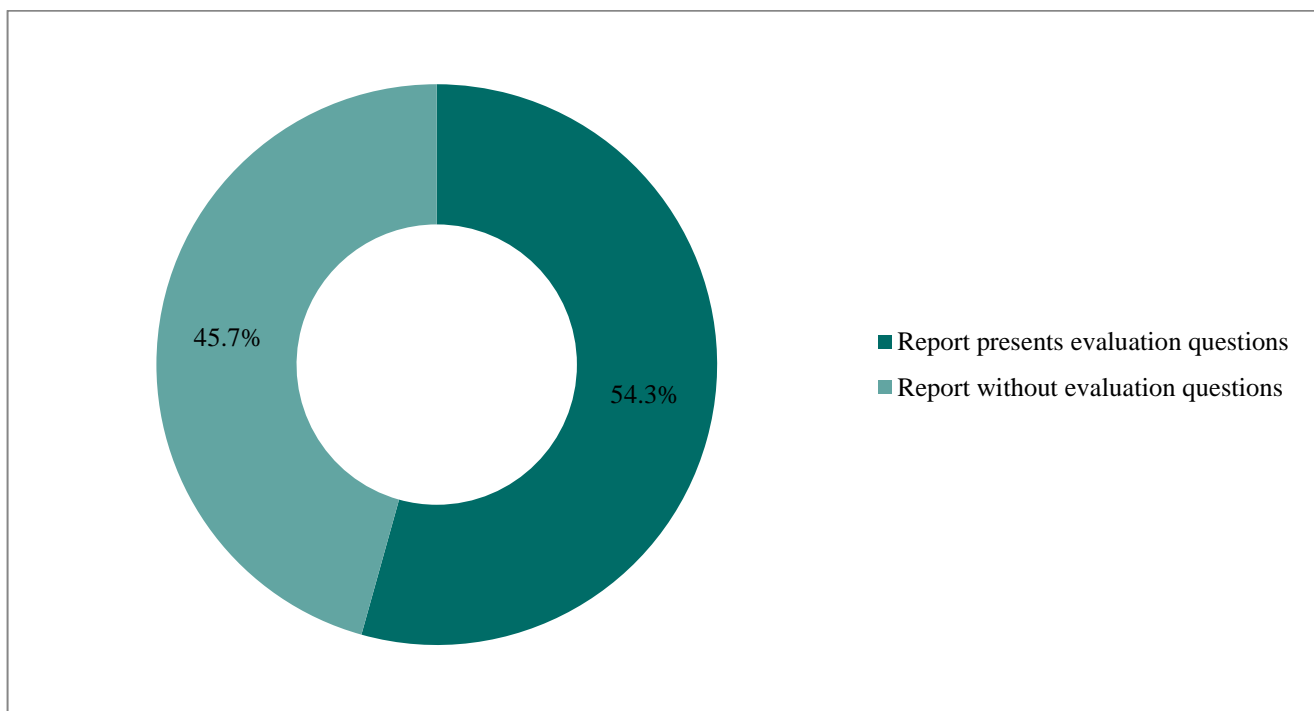
** Note: the sum of the number of reports coded for these WGEE totals more than the 254 documents included in the database; this is because most of them discuss more than one type of WGEE outcome.

The reports analysed, which assess 'economic empowerment' according to one or several of the outcome areas indicated in the table, collectively capture a range of indicators in one or more of the four areas of economic empowerment outlined in Section 1.2: power within (agency and capability); power to (agency and institutional environment); power over (economic advancement); and power with (social relations, networks) (VeneKlasen and Miller, 2002).

3.1.6 Evaluation questions

The first question on the basic scoring tool used to screen reports at the start of stage two was whether the report had clear research questions or research objectives; 138 of the 254 documents in the database (54%) did (see Figure 8). However, even where explicit evaluation questions were reported, not all of them referred specifically to women and girls' economic empowerment; this depends on whether the intervention aimed to change the situation of women, or if it was a mainstream intervention where WGEE was only one feature. In the latter case, it was rare to find evaluation questions specifically focused on outcomes for women and girls' economic empowerment. In order for the evaluation to explore dimensions of WGEE, specific questions in this area are needed to guide the research and analysis.

Figure 8: Share of reports with and without explicit evaluation questions



3.2 Trends and gaps in the evaluation landscape

Based on the studies included in our database, we can identify certain trends and gaps, both in the types of interventions evaluated and the methodologies used.

Trends

- Many of the evaluations and studies (46%) focus primarily on financial services and their effects on women and girls' economic empowerment. Within this thematic area, interventions supporting micro-credit and self-help groups (including savings and loans schemes) are commonly evaluated, but other areas such as micro-

insurance and the use of new technologies (such as mobile phones) are less well explored, particularly in relation to their impacts on women.

- While interventions (particularly successful ones) promoting WGEE generally combine services across different thematic areas, those focusing on business development were often complemented by services such as microfinance, access to markets, and skills development.
- There was a relatively even balance in terms of the methods used, with 30% using quantitative methods, 38% qualitative methods, and 28% mixed methods. A relatively small proportion (4%) used methods that were unclear. Among quantitative evaluations, the use of randomised control trials (RCT) and quasi-experimental approaches was most common.
- The most common donors for evaluations and studies of women and girls' economic empowerment (provided funding directly to the commissioning or implementing agency or through secondary channels) were the Bill and Melinda Gates Foundation, DFID, the Economic and Social Research Council (ESRC), and the International Development Research Centre (IDRC). However, figures on the sources of funds may not be precise, given that this was not always explicitly stated in the reports. The portfolios on WGEE evaluations and studies of donors and commissioning agencies were small overall.
- In geographical terms, most of the reports covered interventions in South Asia (35%) and sub-Saharan Africa (30%).
- Only 54% of the evaluations and studies included in our database had explicit research or evaluation questions.

Gaps

- Relatively few of the evaluations and studies included in our database (55, or 21%) explored economic empowerment outcomes for adolescent girls. While these kinds of interventions have been increasing in recent years with initiatives such as the Girl Hub¹¹, there are still few documented evaluations of their impacts.
- Most of the reports lacked age-disaggregated data, therefore failing to recognise the different experiences of girls and women during different stages of the life cycle (adolescence, early marriage, older married women and elderly women).
- In terms of geographical coverage, relatively few evaluations and studies were included on WGEE interventions in the following regions: Latin America and the Caribbean (11%); East Asia and the Pacific (9%); Middle East and North Africa (10%); and Europe and Central Asia (5%). In the case of Latin America, many relevant evaluations might have been missed because only English language reports were included. In other regions, possible reasons why we have not identified more reports are that there might be fewer interventions to promote WGEE or these might be insufficiently documented.
- In terms of thematic areas, there were relatively few reports in three areas (i.e., with each as the main component): legal and regulatory frameworks (2); unions and fair employment (6); asset provision (4). This might be because few interventions in these areas have targeted women, reflecting a gender blind approach; or it could be because a dearth of gender-disaggregated data in these areas does not allow for a differentiated analysis of the effects of these interventions on women and men's economic empowerment. In the case of 'asset provision', some relevant evaluations and studies might have been classified under a different thematic area such as social protection. Eight reports were classed as 'other' in terms of thematic area; these included interventions focused on information and communications technology (ICT) and on changing social norms.

3.3 Quality and coverage of the shortlisted reports

The scores given to the 70 shortlisted evaluations and studies by each of the four expert panel members are detailed in Annex 8. As noted in Section 2, the quality of the methodology was scored using a scale of 1

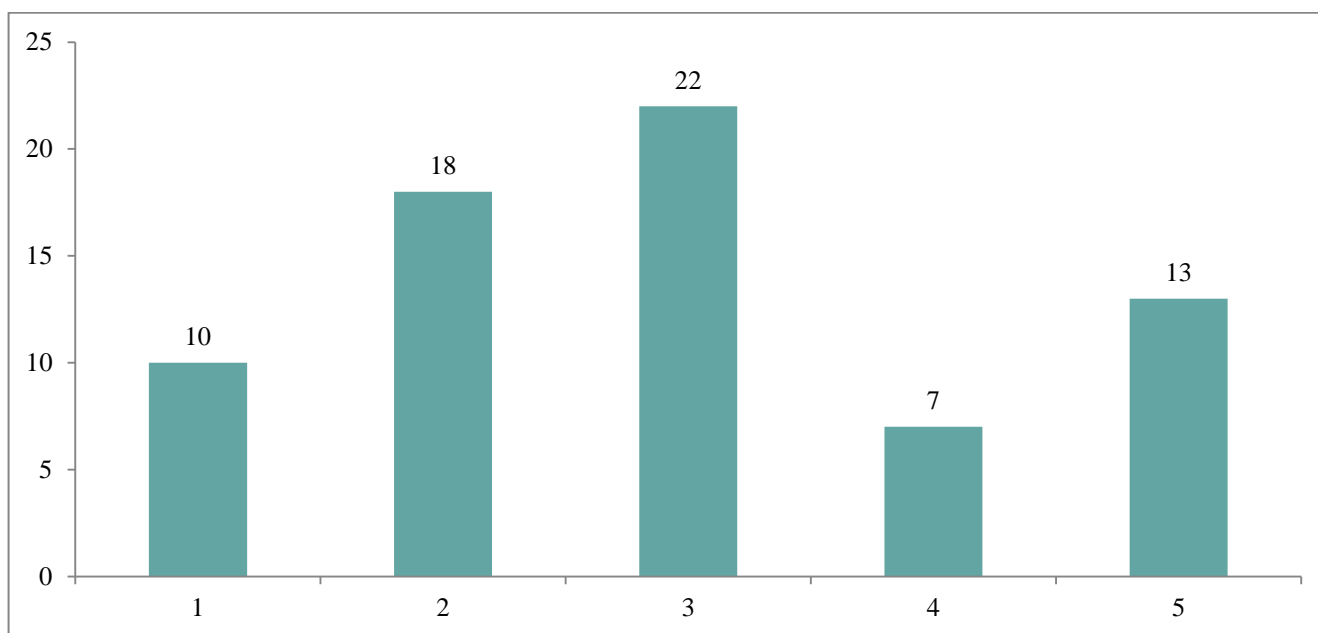
¹¹ A joint initiative between DFID and Nike Foundation.

(highest) to 5 (lowest), and the evaluations received scores across all five grades in the scale. There is no consistent pattern in terms of which evaluations scored the highest – those that scored 1 included mixed methods (e.g. Ahmed, et al 2009), quantitative (e.g. Karlan et al., 2007) and qualitative evaluations (Jones and Shaheen, 2012). Only 9 out of the 70 scored 1 for the quality of their WGEE methodology (which in most cases was the same as the overall methodology used).

In terms of thematic areas, despite the number of ‘financial services’ evaluations, relatively few scored highly on quality. Banerjee et al. (2013) and Karlan et al. (2007) – both quantitative evaluations – scored 1 for their robust, well-conducted methodologies that captured WGEE impacts. Roxin (2010) was one of the few qualitative evaluations in the financial services thematic area that scored highly (2). Evaluations in the social protection thematic area generally scored well, including those using qualitative methods (Jones and Shaheen, 2012), mixed methods (Devereux et al., 2007) and quantitative methods (Oxfam, 2012). Despite the systematic process for selecting evaluations to be reviewed, 10 of the final 70 still scored the lowest grade (5) because their methodological approaches to assessing impacts on women and girls’ economic empowerment were deemed very poor quality.

Figure 9 illustrates the distribution of scores given to the 70 evaluations and studies reviewed based on the methodology used to assess women’s and girls’ economic empowerment, with 1 as the highest and 5 the lowest scores. It is worth noting that even at this stage of the review, 13 reports were scored as having very weak methodologies to assess WGEE, and only 10 were scored to have an excellent methodology. There were 22 reports that scored ‘satisfactory’ on this measure, the score that occurred most often.

Figure 9: Scores for methodology used to assess Women’s and Girls’ economic empowerment



* Note: On the horizontal axis, 1 is the highest score, 5 is the lowest

Section 4 analyses the quality and success factors of the evaluation methodological approaches used in the 70 shortlisted documents.

4 Elements of a good-quality evaluation to measure WGEE impact

This section analyses key dimensions of ‘quality’ in the approaches used by the 70 shortlisted evaluations or study documents that were reviewed by the expert panel. It gives examples from specific evaluations to illustrate strengths or weaknesses in relation to the quality criteria defined during the review phase and captured on the review templates¹² (see Section 2). It draws largely on information from the expert reviews, and builds on insights derived from the meta-analysis of the set of reviews. Examples of evaluations or parts of evaluations are used throughout this section to illustrate important issues such as good practice of evaluations on WGEE; elements of methodology that are useful (even where the overall evaluation is not); and weaknesses that other interventions or evaluations can learn from.

As noted in the introduction, for the purpose of this study empowerment is defined as a process of change that transforms women and girls’ lives in these four areas and interacts with resources (pre-conditions), agency (process), and achievements (outcomes) (Kabeer, 1999). Economic empowerment will only be possible and sustainable if there are changes at different levels: within the individual (capability, knowledge and self-esteem); in communities and institutions (including norms and behaviour); in available resources and economic opportunities; and in the wider political and legal environment (Golla et al., 2011). We define this as a ‘holistic’ approach to WGEE because it is not just looking at whether they have increased their access to income and assets, but also whether they have more control over them and are able to use them to have greater control over other areas of their life.

Findings are presented under the following headings, which correspond to the evaluation process:

1. Applying a holistic approach
2. Building a team
3. Preparing for the evaluation and understanding the context
4. Understanding the intervention – including theoretical frameworks and theory of change
5. Methodology – including indicators
6. Sampling
7. Tools and data collection
8. Analysis
9. Self-reflection
10. Dissemination and use of findings
11. Innovations

4.1 Applying a holistic approach to women and girls’ economic empowerment (WGEE)

- Not all of the 70 shortlisted documents reflected this holistic approach. Many had a partial approach to WGEE and looked at change in terms of economic advancement only (11 out of 70, or 16%).
- Where change in women’s power and agency was assessed, it was common to limit this to decision-making power within the household.
- There are varying levels of sophistication of the WGEE frameworks used. At the very minimum, the evaluation would measure changes in the individual’s decision-making power within the household as a measure of women’s agency.
- The lack of measurement of change in norms and attitudes carried through to all levels of analysis.

¹² The questions and criteria in the template were based on a document review of several evaluation best practice documents, as well as on key resources on women’s economic empowerment – for example, the Development Assistance Committee (DAC) criteria for evaluating development assistance

¹², the Bond quality of evaluation checklist, DFID’s evaluation policy (DFID, 2013), and Kabeer’s (1999) paper on Reflections on Women’s Economic Empowerment (more details in annex 3).

The evaluations reviewed by the expert panel covered interventions that were either mainstream interventions that did not have a particular focus on women and girls, or interventions that did have a specific focus on women and/or girls (the latter accounted for most of the evaluations reviewed). Many of them did not explicitly aim to 'economically empower' women and/or girls, but the evaluations may have looked at economic empowerment results or possibly just at economic advancement.

According to the definition of women and girls' economic empowerment used for this review, it is necessary for the intervention to result in some kind of transformational change on *women and girls' economic advancement and women and girls' power and agency*. As noted in Section 1, the empowerment process should show change in four dimensions – power to, power over, power within, and power with – in relation to women and girls' economic activity and their ability to control assets. We define this as a 'holistic' approach to WGEE (see Annex 2), because it is not just looking at whether they have increased their access to income and assets, but also whether they have more control over them and are able to use them to have greater control over other areas of their life.

Not all of the 70 shortlisted documents reflected this holistic approach. Many had a partial approach to WGEE and looked at change in terms of economic advancement only (11 out of 70, or 16%). Where change in women's power and agency was assessed, it was common to limit this to decision-making power within the household. Interestingly, the quantitative studies were less likely to have a multidimensional and holistic view of women's economic empowerment, possibly because the types of indicators used tend to be more limited to changes in economic status, and do not always capture how and why changes happened to women and/or girls.

Evaluations that demonstrated a holistic approach to women and girls' economic empowerment included the following: Ahmed, et al (2009), Garikipati (2008), Arif et al. (2010), Baden (2013), Bureau of Applied Research in Anthropology (2013) and Jones and Shaheen (2013). However, there are varying levels of sophistication of the WGEE frameworks used. At the very minimum, the evaluation would measure changes in the individual's decision-making power within the household as a measure of women's agency. However, this is not sufficient to demonstrate economic empowerment, as it would not show the potential for long-term and lasting change. For example, one evaluation had demonstrated an improvement in women's decision-making power within the household, but this did not follow through into improvements in community norms and attitudes. Potentially, a backlash at this level could have limited any longer-term changes that might have been possible.

Targeted evaluations or evaluations of targeted interventions seem to have fared better in terms of capturing women's 'voices' than those with general population targets. However, these did not always evaluate how men's attitudes and behaviours might have changed, or how the intervention may have led to changes in men's economic empowerment in relation to women's economic empowerment. In 38 evaluations men's views were elicited as part of the assessment of changes in women's economic empowerment. However, this was not as common as expected by the reviewers, the information was not always used in the analysis and overall there was very little measurement of men's attitudes and behaviours in relation to women's economic empowerment.

The lack of measurement of change in norms and attitudes carried through to all levels of analysis. It is generally not done at the institutional level and at the level of the legal and policy enabling environment. Only two evaluations considered gender mainstreaming in institutions – and these were limited to the fair trade and collective action thematic areas.

4.2 Building an evaluation team with gender expertise

- Evaluation teams with gender experts or expertise presented better quality analysis in relation to women and girls' economic empowerment.
- Gender expertise generally contributed to the integration of gender into research methodology.
- Evaluation reports that included information on team members' profiles were typically of higher quality and more transparent.

The 70 shortlisted evaluations and studies gave very little information about the expertise of evaluation team members. The review process did, however, reveal that gender expertise was an important aspect of a successful evaluation. Evaluation teams that included a gender expert, or where gender was a mainstreamed skill across the team, tended to have a better integration of gender concerns throughout the evaluation process

and performed better on various gender-related aspects. Weak understanding of gender dynamics and knowledge of women's empowerment issues was highlighted as an important limitation in a substantial number (42%) of reports, reflecting insufficient or inadequate analysis of women's economic empowerment.

4.3 Preparing for the evaluation and understanding the context

- Approximately 59% of the evaluations presented a good-quality context analysis, but very few of these included a thorough analysis of gender differences and the underlying context of gender inequality.
- Good-quality context analysis usually includes desk-based research and qualitative data collection.
- Good context analysis presented market and economic context, political and institutional context, gender differences, and processes of change at household, community and institutional levels.
- While it is fairly common for context analysis to look at gender differences, there are few examples of analysis of gender norms and the attitudes of men and women, as well as of gender relations and gender roles.

A good analysis of the social and economic context, including a strong gender analysis, was consistent with good-quality evaluations. If context analysis is done sufficiently well, it will enable a relevant theory of change to be developed and a more profound understanding of gender dynamics and how the intervention might affect them, at different levels. Further, initial context scoping, when designing the evaluation, should also help in the design of relevant sampling strategies and tools. The expert reviewers scored the documents highly for context analysis where it included a process of (1) desk-based research into background information, previous relevant research, and available data sets for the country and local communities; and (2) qualitative primary data collection.

Examples of good context analysis and of gaps

Bacon, 2010 (*A Spot of Coffee in Crisis*), *Nicaraguan Smallholder Cooperatives, Fair Trade Networks, and Gendered Empowerment* gives a strong context analysis that includes the (coffee) market, the political situation, relevant empowerment processes, and the relevance of cooperative and union institutions. The Saving for Change impact assessment (Bureau of Applied Research, 2013) included a set of interviews for context analysis to support design of the methodology. The following quote demonstrates the range of business and social topics covered by initial interviews, but also the insufficient depth of gender analysis in relation to attitudes and norms, and gender relations: '*Key informant interviews within the region also provided important context on local access to credit, markets, and other business opportunities available to women. Qualitative interviews with program beneficiaries and community leaders informed the design of the quantitative survey instruments, and the rate of expansion across villages informed statistical power calculations for the sampling strategy of the quantitative study.*'

4.3.1 Gender analysis of context

Evaluations of interventions designed to improve women and girls' economic empowerment require a thorough gender analysis of the context. Economic empowerment is a process of change, and so any evaluation will need to measure whether change has happened, and how. Gender differences, gender-related norms and behaviours, gender roles and gender relations differ in every context, and assumptions should not be made when designing an evaluation. For example, one evaluation measured the level of empowerment based on whether men and women spent money on what they wanted. However, this made assumptions about what women typically spend money on (for example, clothes), without accompanying evidence of what women actually spend money on and why. Making these kinds of assumptions could lead to inconclusive results.

Fewer than 25% of the evaluations and studies reviewed undertook a thorough gender analysis. This may have been because the intervention did not aim to improve women and girls' economic empowerment, but was focused on men and women's economic empowerment or advancement generally. In some cases, an evaluation did aim to measure women's economic empowerment but there was a lack of understanding of the importance of the multiple gender inequality dimensions present, and how they related to the intervention's achievements, and its evaluation. This can happen when assumptions are made about gender differences (as in the example in the previous paragraph); when the views of men and boys are not taken into account; or when cultural and social norms are not examined fully.

Examples of good gender context analysis

An example of good gender context analysis can be found in Arif, et al, 2009, which explores the following aspects of the gender context in Indonesia:

- women's exclusion from decision-making roles in community level committees
- limited intra-household decision-making and bargaining power based on age and/or gender
- time poverty as a result of unpaid productive work responsibilities and/or familial care work.
- prevalence of gender-based violence

The expert reviewers considered that a thorough context analysis was more likely to lead to a good theory of change (see next section) and appropriate methodology and tools for measuring impact on WGEE.

The World Bank and the IFC (2011) examined the opportunity for Special Economic Zones (SEZs) to promote women's economic empowerment and boost zone and enterprise competitiveness in eight developing countries. The context analysis focused on the underlying conditions that affect women's employment. (Note that it does not include information about household norms and behaviours in terms of decision-making.) It is a useful example because it shows the range of issues that were considered for assessing women's employment, not just their skills and education (which is what studies typically focus on). The context section also includes a good description of Special Economic Zones and the advantages and disadvantages of this kind of employment for women.

Box 1: Interesting innovation: the use of an index based on social norms restricting women's behaviour

Field et al.'s (2010) analysis of how traditional institutions constrain female entrepreneurship, based on a field experiment on business training in India, included an interesting innovation: the creation of an index based on restrictions faced by women.

The study sample was comprised of the 597 women surveyed successfully in a follow-up round of data collection and who could be categorised into sub-castes based on surname. They were categorised into three broad social groups: Muslims, Hindu Scheduled Castes (SCs), and Hindu Upper Castes (UCs) (non-scheduled castes including other backward castes). In addition, the authors scored how restrictive each Hindu and Muslim sub-caste was with regard to five norms governing women's behaviour: ability to socialise alone, requirements to cover the face or wear a veil, ability to speak directly to elders, ability to leave the house or neighbourhood alone, and ability to remarry.

They created an index ranging from 0 to 5 equal to the number of norms for which the sub-caste was highly restrictive. According to this index, SC women face the fewest restrictions, followed by UC and Muslim women.

Context analysis in Special Economic Zones – barriers to women's employment:

- Cultural norms and expectations influence women's role in society and the economic opportunities available to women.
- Lack of access to education limits women's employability and potential for upward mobility once employed.
- Inadequate knowledge of women's health issues and limited access to women's health services pose serious obstacles for women that impact absenteeism and productivity.
- Lack of access to quality and affordable childcare has been noted as a significant challenge, even in countries where provision of childcare is mandated by law.
- Substandard or unsafe living conditions and transportation to the workplace make women workers more vulnerable.

Source: *Fostering Women's Economic Empowerment through Special Economic Zones* – WB and IFC, 2011

UC women were 50% more likely to face severe social restrictions than SCs, and Muslims, in turn, had more than twice the rate of severe restrictions as UCs.

Source: Field et al., 2010

4.3.2 Use of wider literature and international knowledge

A few of the evaluations and studies shortlisted did not use or refer to the wider body of evidence in the field, which meant that they were not able to use this information to strengthen the evaluation design and analysis. This sometimes happened because the evaluation team did not draw on wider international knowledge. Approximately half of the reports did convey good knowledge of the relevant literature and built on this to deliver a more innovative or value added approach to generate new findings. This was the case for Rubalcava et al. (2002), which built on the extensive evaluation literature of the Progresas / Oportunidades conditional cash transfer (CCT) programme in Mexico. This evaluation used a standard methodology to identify the effects of the CCT on household expenditure, focusing on whether women's bargaining power had increased. The approach led to careful selection of comparison groups in order to avoid potential unobserved heterogeneity.

4.4 Understanding the intervention – theories of change and frameworks

- 71% of the evaluations gave a clear description of the intervention being evaluated.
- Only 44% presented a clear, explicit theory of change; 26% had no theory of change or theoretical framework at all.
- A good theory of change was correlated with overall good-quality evaluation.
- A good theory of change will: identify causal relationships and multiple relationships; identify underlying assumptions; show change over different timeframes; show transformative change in women and girls' economic advancement and agency; and include changes in norms and institutions.

Most of the highest-scoring evaluations and studies had a well-articulated theoretical framework and/or theory of change. This enabled the evaluators to demonstrate where and how change would happen as a result of the intervention – and thereby to frame their evaluation around that change process. The best theories of change showed dynamic and multi-faceted change processes that reflected transformational changes in women's agency and economic advancement, and the relationship between them. However, 26% of the reports reviewed had no theory of change or theoretical framework. A further 13% only had partial theories of change within the evaluation text, but no explicit framework. Where there was a theory of change, several only referred to change in economic advancement – with no transformation in women and girls' agency.

Examples of theories of change and frameworks

A useful theory of change can be seen in Slegh et al.'s (2013) evaluation of a pilot project in Rwanda that targeted male partners of women involved in a CARE-funded microfinance scheme so as to counter gender norms, power dynamics, and gender-based violence (GBV). The theory of change employed situates the pilot project within broader literature on the multiple spill-over effects of women's economic empowerment on their individual social wellbeing and on familial wellbeing. But they note that this positive relationship cannot be assumed and is, in part, dependent on men's reactions to women's changing roles and status:

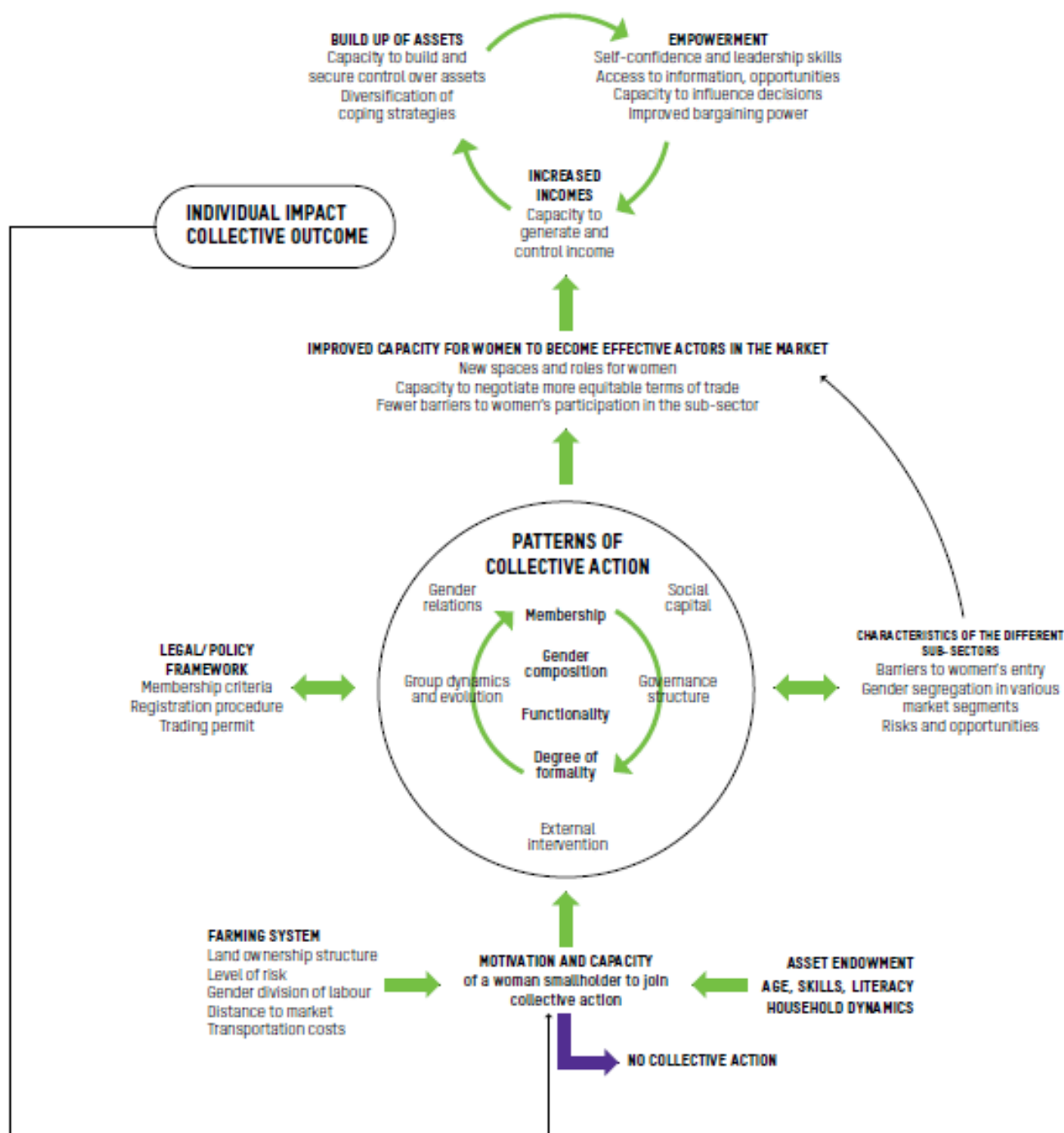
'Some of this research has highlighted the fact that men react in very diverse ways when their wives are beneficiaries of microcredit programmes (e.g. Ahmed 2008). Men do not uniformly "fall into line" and accept gender equality when their female partners gain more income.' (p 17)

A study on the Special Economic Zones (SEZs), women's employment, and the legal and regulatory environment (WB and IFC, 2011) did not present a framework for women and girls' economic empowerment – though it was looking at SEZs across eight countries, rather than evaluating a specific project. It focused mostly on women's rights within the workplace, economic advancement, and the benefits of women's employment to companies and national economies. However, it does provide some useful information and recommendations for developing a theory of change for this type of interventions. It also has a useful table on M&E processes, with extensive exploration of indicators at impact, outcome and output levels.

The theory of change employed by Baden (2013) (see Figure 10) shows the multiple levels at which decisions and changes take place. It shows how the context (shown at the bottom of the picture) affects the extent of women's participation in collective action, and its impact on patterns of collective action (in the middle of the

picture). This is a useful way of presenting the change process because it encompasses the conditions for various changes to happen. This is one way of highlighting assumptions (and risks) that may apply, but it could be argued that any framework should be more explicit about the assumptions underlying the change and then test these in the evaluation

Figure 10: Theoretical framework – collective action



Source: Baden, 2013

Additional examples of good theories of change can be found in Annex 9.

4.5 Methodology and indicators

- Mixed methods evaluations were the most effective in terms of measuring change and outcomes, and also provided good material for learning and improving future project design.
- Few evaluations had a baseline to measure progress against.
- Evaluations that used qualitative and participatory methodologies gave a more in-depth analysis of how and why change happened.
- Indicators to measure change in women and girl's economic power and agency were better quality if multi-dimensional across individual, family, community and institutions.

4.5.1 Quantitative, qualitative or mixed methods?

Overall, the experts took the view that mixed methods are optimal for evaluating the impact of interventions on women and girls' economic empowerment. As long as mixed methods evaluations are of good quality, they can provide robust evidence on the impact of the intervention through the indicators identified, as well as providing insights about why and how change happens, which is particularly important in understanding changes in WGEE. A good mixed methods approach means that qualitative and quantitative data can complement each other, so that the whole is more than the sum of its parts. It also allows for effective triangulation of data. In low-resource cases, qualitative or quantitative data alone would be useful as long as the evaluation design was robust. The expert review panel observed the complementary roles of both kinds of methods:

- **Quantitative** methodologies tended to be used to demonstrate that: (1) change had taken place; (2) the intervention caused the change to take place (*causality*); and (3) the findings can be generalised across a population group (in the case of representative samples).
- **Qualitative** methodologies were most commonly used for: (1) context analysis to design sampling and tools; (2) conducting participatory activities to identify indicators of change; (3) establishing how the change takes place; (4) understanding *why* change happens (or does not happen); (5) researching how people understand and describe that change; and (6) identifying unintended changes or impacts.

Sequencing of quantitative and qualitative data collection was relative to the particular role each method played in the overall evaluation. However, it was rare to see a good-quality explanation of why mixed methods had been chosen and how a particular sequence had been decided. Evaluations tended to use qualitative methodologies either before or after the application of a quantitative survey, but most did not explain why they had sequenced the data collection in a particular way.

Multiple sources of data were important for triangulation in cases where there was no baseline. Evaluations were generally scored higher if the analysis in a mixed methods approach used both the quantitative and qualitative data to complement and reinforce each other. However, thorough triangulation in mixed methods evaluations was rare; many had separate and distinct sections for the qualitative and quantitative aspects, which in some cases did not reference each other.

Members of the expert panel suggested that **funding** appeared to be lower for qualitative evaluations. A large number of qualitative evaluations in the shortlist reviewed had limited scope and reach, with small samples. This suggests a lack of value and funding attached to qualitative methods. In many of the studies reviewed, it is unlikely that the point of data saturation was reached.

Examples of good mixed methods approaches to measuring impact of WGEE interventions

Examples of good data triangulation include Hashemi et al. (1996), a study of the impact of rural credit programmes on women's empowerment in Bangladesh, which used several methods (quantitative surveys, informal and qualitative interviews, observation and case studies) and exhibited thorough triangulation of data during the analysis. Two other examples are given in Box 2.

Holvoet (2005), in a study on the impact of microfinance on decision-making agency in South India, used qualitative interviews to inform the survey design. The focus was on comparing current and retrospective responses about the decisions individual women made, rather than their perceptions of decision-making, thereby making the need for baseline data less relevant.

In the case of Andrews et al. (2003) (a study of the EMPOWER programme addressing food security and gender relationships among small-holders in Ethiopia), despite some limitations of the data provided by programme staff, concerns were openly discussed by the evaluators who sought ways to triangulate and verify data and participants' experiences as described through stories.

Box 2: When mixed methods work well

The **Bureau of Applied Research in Anthropology (BARA)/Innovations for Poverty Action (IPA) 2013** impact evaluation of the Saving for Change (SfC) programme in Mali was also scored as excellent by the expert reviewers. It is a good example of longitudinal and mixed methods, with the use of a randomised controlled trial (RCT) and interesting complementary methods. A good context analysis feeds into the methodology and data collection tools chosen. BARA undertook the data collection for the qualitative study and IPA for the quantitative study. The results were presented in different sections of the report, but another section analysed joint findings. In the analysis, there is good triangulation of quantitative and qualitative findings, to analyse complex and multifaceted topics. Questions are approached from several different angles, which enriches the evaluation overall.

The quantitative element of the evaluation comprised of an RCT with a baseline and endline survey, a social network census, and a series of high-frequency surveys to a panel of treatment and control group respondents (this is discussed in more detail in Section 4.10, on innovations). There was a good point-by-point analysis of the differences or similarities between the treatment and control groups, particularly in relation to: savings and borrowing; vulnerability, risk coping and security; investment, income and assets; health and education; empowerment and social capital; and village characteristics.

The qualitative study methodology consisted of focus group discussions with men and women, key informant interviews, and household interviews in a sample of villages (most of which were involved in the RCT). *'The methodology was based on the Household Livelihood Systems Approach (HHLS), which analyses local household dynamics in a holistic socioeconomic framework within the broader community and region. Throughout the research process, an emphasis was placed on community participation in identifying and analysing household dynamics.'* Discussion topics included the history of the SfC group, its objectives, criteria for membership, relationships among members, and group activities. The groups also discussed SfC's impact on women's lives – on their status, their role in the household and in the community. Finally, they discussed its contribution to increasing the resilience of the entire household.

4.5.2 Randomised controlled trials (RCTs)

17% (12) of the evaluations and studies reviewed used RCTs to evaluate the interventions.

The level of randomisation and the ability to follow up with the control group are key factors in determining whether an RCT is used effectively. This was not found to be done in all of the RCTs reviewed. Of the 12 reports using RCTs, 4 were scored as excellent quality, 4 as good and 4 were scored as satisfactory quality.

With the RCT-based evaluations that scored 'satisfactory' there were several issues. When collecting data at two points in time (baseline and endline), some RCTs faced high levels of attrition – that is, large amounts of the sample had moved away from the area and there was no mechanism in place for follow-up. RCTs that worked well had developed plans *ex ante* to determine how they would follow up with a very mobile group of people (mindful that people in developing countries are very likely to move for economic reasons).

Another important quality aspect for RCTs is that in order for them to work, the sample needs to be large enough for results that are likely to be significant at hypothesized levels of the effect and the researcher can test whether the control and treatment groups are balanced. For example, at the community level, one would want a large enough sample that any difference in the observables could be tested for significance. This is relevant in case of evaluations and studies on WGEE since some of them were found to have limited resources and thus the small sample did not allow for a Robust RCT. Adequate resourcing and robust sample and study design are needed to increase the number of good quality RCTs looking at WGEE.

Example of a good-quality RCT measuring traditional institutions

Field et al.'s (2010) analysis of how traditional institutions constrain female entrepreneurship provides a good example of an effective and well-implemented RCT. The authors used an RCT to help them develop a unique index to examine the role that traditional barriers (religion, social norms, etc.) play in restricting the business growth and profitability of female-run enterprises. The RCT allowed for a causal interpretation of the programme. Ideally, besides randomly assigning females to the training programme, one would also randomly assign each person with a different level of traditional barriers. Then one could look at the effect of barriers independent of what those barriers are usually correlated with – such as wealth and ethnicity. However, as in all cases where one is looking at social norms, this cannot be done. According to the expert reviewer, the RCT was the best approach to adopt in this case, especially since the authors deal with the potential correlation by carefully selecting the sample on which to do the randomisation. The authors took steps to make sure that as many key observables as possible were balanced the same for each group identified by the index, as if the social group were randomly assigned. Thus, in this case, women of homogenous socio-economic backgrounds were the only ones used in this study

4.5.3 Other quantitative methods

Of the evaluations and studies that used quantitative methods (54 used quantitative methods, including 31 purely quantitative evaluations and 23 as part of mixed methods¹³), 21 used quasi-experimental design and 14 were descriptive or correlational. Quasi-experimental design is used when evaluators seek to determine impact or causality but where randomisation is not possible, drawing on techniques such as propensity score matching (PSM) or difference in difference (DD). These methods were successfully used in some cases.

However, some of the evaluations that used quantitative methods did not have enough analysis and justification to ensure robustness – including, for example, not having the required checks or a discussion of the potential biases that can occur from the un-testable assumptions. Twenty-four of the evaluations that used quantitative methods were scored 3 or lower by the expert reviewers.

Examples of studies that used quasi-experimental design well

Ahmed et al. (2009) used PSM to compare a treatment and control group at one point of time. This was the most appropriate approach given that the programmes had already been implemented when the household survey was carried out. The authors note, in their background research, that PSM is reliable in this context: *'Through comparisons with experimental estimators, Heckman, Ichimura, and Todd (1997, 1998) and Heckman et al. (1998) have shown that PSM provides reliable, low bias estimates of program impact provided that (1) the same data source is used for participants and nonparticipants, (2) participants and nonparticipants have access to the same markets, and (3) the data include meaningful explanatory variables capable of identifying program participation'* (p 29). The authors used a mixed methods approach, complementing a quantitative household survey with qualitative interviews and focus group discussions. This evaluation had used good methodologies well, and provided disaggregated data, although it could have given a stronger analysis of issues that are relevant to women's economic empowerment.

Another example of how quasi-experimental design was well utilised is Oxfam GB's (2012) evaluation of Guatemala's Highland Value Chain Development Alliance. They had an ex-post set up and asked recall questions so that they had baseline data they could use for the PSM technique to get an adequate control group. The PSM technique was well implemented, and the variables used to elicit information about women and girls' economic empowerment were innovative and useful – particularly those on decision-making and self-efficacy, which were created using principal component analysis. The evaluation uses the 'Oxfam GB global indicator for women's empowerment', a measure which is positive when a woman has both greater involvement in household decision-making and greater ability to influence affairs at the community level. The use of a good, well-implemented methodology coupled with good indicators makes this study important to understanding how interventions aimed at increasing food and agricultural security can be complemented with economically empowering women.

¹³ 3 mixed methods evaluations used another mix of methodologies, not including quantitative methods.

4.5.4 Qualitative methods

Of the evaluations and studies reviewed, 39 used qualitative methods (13 exclusively and 26 as part of a mixed methods approach). Standard tools such as in-depth interviews (IDs) and focus group discussions (FGD) were used most, although it was not always clear why the specific qualitative research methodology was chosen as the methods were generally not described in detail. Those that scored highest were transparent about why the qualitative research design chosen was used. In this sense, including detailed discussion of both qualitative and participatory methods – rather than referring vaguely to ‘outcomes from FGDs’ as if these are unproblematic – can strengthen the credibility of qualitative findings.

A common problem with qualitative evaluations was the small size of the sample. Although qualitative evaluations are generally not representative of a population, they serve to shed light on the change process that is occurring, so it is important that the sample selection is explained and findings contextualised accordingly. Qualitative studies are thus well complemented by a quantitative component.

One of the strengths of the evaluations that used qualitative methods was the value of quotes from IDs and more in-depth case studies in text boxes, which add legitimacy to the findings and help to bring the issues alive for the reader.

Example of a good qualitative evaluation that includes strong analysis of WGEE

Jones and Shaheen’s (2012) evaluation of the effects of a cash transfer on female-headed households in the Occupied Palestinian Territories (OPT) has a detailed methodology section that provides a clear overview of the research objectives, research questions, tools and sample. There is a concise and helpful rationale for the use of the multiple instruments that make up the qualitative study. The authors also reflect on capacity-building considerations and research ethics, which indicates transparency on the part of the research team and awareness of dynamics and context, which are particularly important when collecting qualitative data. While the report goes beyond a gender analysis, the analytical framework (which includes a focus on economic and social vulnerabilities) helps to ensure that gender considerations are adequately reflected in the research questions. The qualitative information is interwoven throughout the analysis, while remaining engaged with the broader theoretical literature and a rigorous conceptual framework. Given the focus on beneficiary and community perceptions, the final analysis provides rich insights into the complex vulnerabilities and coping repertoires of impoverished women in the West Bank. However, more could have been done methodologically to tease out the gender dimensions of the political economy/accountability sections.

4.5.5 Participatory methods

The few evaluations or studies that had used participatory methods had a richer set of tools and a more complex set of results. The use of participatory methods was particularly important in capturing the voices of women and girls, and men and boys, to inform both the development of indicators as well as the evaluation findings.

Examples of how participatory methods can improve quality

Bacon’s (2010) research into the empowerment of women involved in coffee cooperatives used participatory methodology to develop data collection tools and analyse data. Data collection was done by young people from the coffee-growing communities, and gender promoters. The data from an initial social development and agronomic survey at household level were presented and co-analysed with research participants using a popular education approach (a participatory methodology commonly used in Latin America). This analysis was followed by FGDs and interviews with a subsample. As well as household-level empowerment, gender equality and empowerment processes were assessed at the cooperative level. The following example of the findings shows both how the participatory work produced strong detail about the cooperative structures and how the cooperative may have influenced women’s behaviour:

‘Although members of the cooperative demonstrated a strong sense of collective empowerment, individual interviews revealed uneven gendered empowerment processes. When gender promoters conducted their first workshops with co-op members, men often disrupted the meetings. The co-op had only two active female members, but key informants said that the level of machismo had decreased since 2004. Diverse gender relationships have developed over time, many in accordance with specific personalities and intra-household dynamics outside of the cooperative structure. Some households affiliated with this cooperative have maintained

more gender equity in their relationships, and some women have regularly left their houses to participate in natural-medicine workshops, community meetings, and other activities' (Bacon, 2010, p. 59).

The participatory methodology used in Baden's 2013 research on women's collective action was an end in itself, with the aim of enhancing credibility and ownership. It was accompanied at different stages by stakeholder dialogues with the following groups: women small producer leaders, representatives of formal producer organisations on the ground, and NGO practitioners at local and national levels. Sub-sector focus was determined by working with local stakeholders in each region in order to target sectors where women are economically active and where collective action is happening at different stages of the value chain. Local stakeholders also identified key informants. '*Women producer leaders and development actors supporting WCA (women's collective action) in the regions provided feedback on preliminary findings, and on which questions which are relevant to them to inform the research as it develops, at each stage.*'¹⁴ The research study applied a mixed methodology of quantitative and qualitative tools. Participatory tools and feedback meetings enabled sharing of analysis of results. This range of methods contributed to triangulation and gave the findings and lessons their strength and credibility.

4.5.6 Ethnographic methods

Ethnographic research is a useful and underused qualitative methodology that relies on long-term data collection in the community. It gives the chance to unpack complex dynamics in a thoughtful way over time, which is particularly important to capture medium and long-term change dynamics resulting from transformational economic empowerment. It seems unlikely that these sorts of insights could be captured through more qualitative snapshot approaches, which are the norm. Although by themselves, the findings generated by ethnographic research are likely to be limited, they could have considerable value as part of a mixed methods approach. This method is, however, resource and time intensive, part of the reason why it is seldom used.

Example of a good ethnographic evaluation

Corboz (2013) is a useful example, as using this method has enabled a much more in-depth analysis of the impact of conditional cash transfers in Uruguay. It goes beyond personal and community capabilities to explore the power relations between women and men in their community and with the state, which provides the cash transfer. The ethnographic method picked up complex issues around empowerment/agency or lack of it – particularly in relation to violence and lack of security. The research explores how violence, crime and unequal gender relations limit 'women's ability to comply with their social and civic duty'. However, the expert review of this study suggested that the methodology could have been enhanced by using and triangulating more data collection methods, including a quantitative survey or other qualitative methodologies such as community mapping, focus group discussions, and key informant interviews with programme implementers and others.

4.5.7 Time frame of an evaluation – demonstrating long-term change

To uncover empowerment as a process that results from an intervention, particularly through qualitative data, it is important to have instruments that delve into change over time, such as life histories (LH), intergenerational (IG) comparisons, community maps, etc. This is particularly important in the absence of baseline data against which to measure changes over time. Although recall responses may not always be sufficiently rigorous measures of impact, they provide rich information about how individuals perceive their lives to be affected by an intervention.

Where baseline data are available, there is more scope for quantitative and qualitative evaluations to explore causality and attribute changes more clearly to the intervention.

Evaluations can be conducted some time after an intervention has ended and still show evidence of long-term impact – for example Blattman et al. (2013) conducted an RCT of a cash transfer programme in Uganda, surveying a panel of 2,675 young adults from the start of the programme and then two and four years after receiving grants. As such, they show change over a four-year period, highlighting how a cash transfer can be more effective than finance in terms of long-term effects.

¹⁴ For more information on the stakeholder dialogues, see: <http://womenscollectiveaction.com/Stakeholder+Dialogues>.

4.5.8 Types of indicators and how indicators are chosen

The choice of dependent variables for quantitative analysis is vitally important for a good-quality assessment of change in women and girls' economic empowerment. In many cases, evaluations were measuring progress against a pre-agreed set of indicators that appeared to be at output level only, with no analysis of unexpected or wider changes that might have taken place. Measurement of change at outcome or impact level of the logframe required a set of indicators that (given the context) adequately measured transformational change on women and girls' economic advancement and power and agency.

Some evaluations of outcome and impact measured economic advancement alone (for example, Blattman, 2013). Others used indicators to demonstrate change in both women and girls' economic advancement and their power and agency. However, measures of change in power and agency could be one-dimensional – focusing on decision-making power within the household only or, in the case of Bandiera et al. (2012), looking only at changes at the individual level (girls) rather than more broadly, in the family and wider community.

The reviews identified that evaluation teams and researchers often make underlying assumptions and use stereotypes of women's activities, which can limit the level of analysis. These types of assumptions should be avoided. The following examples highlight how the type of questions that are asked of women and men can be gender specific:

- *Type of expenditure* – women are only asked about expenditure on household items or child-related expenditure whereas men are asked about recreation and other activities. This reinforces the stereotype that women are supposed to be responsible for household and child-related expenditure, and may limit investigation into changes that could be taking place. It also negates the possibility that men may also have a responsibility for household expenditure.
- *Type of decisions* – women are typically asked about who makes decisions about issues that are relevant to the division of household labour (who decides how the woman and children's time is allocated), but seldom asked about other critical areas of life such as decisions about who participates in community or civic events, how women's time is allocated to activities outside the household (labour or social), etc. Only rarely are questions asked about joint decision-making, with researchers tending to frame their questions as 'either/or' (for example, are household decisions made by men or women), rather than asking if some or all decisions are made jointly. In many cases, joint decision-making might actually denote a better positioning of women in the household.
- *Type of economic activities* – assumptions may be made about women being home-based, or economically active only in certain sectors, so that their involvement in other sectors is not considered.

Evaluations that scored well used a multidimensional range of indicators to measure women and girls' economic advancement and changes in their power and agency (see the example of Garikipati, below). The best evaluations had a clear theory of change or theoretical framework from which indicators could be developed. As mentioned earlier, a thorough gender context analysis and participatory approach to defining indicators produce a better quality evaluation and more relevant indicators. The most innovative studies used unique power variables that captured concepts of economic empowerment that one might not normally consider, such as whether a young girl was less likely to have unwanted sex (e.g. Bandiera et al., 2012). The choice of what variables to use will obviously be context-specific; but, of the reports reviewed by the expert panel.

Examples of reports that used rich and multidimensional indicator sets

BARA/IPA's (2013) evaluation of the Saving for Change programme in Mali used indicators for the quantitative research that were comprehensive and appropriate. They used three indexes: one on intra-household decision-making; one on community action; and one on social integration. Questions used to elicit decision-making power in the household were based on the following: whether the woman makes her own decisions about food expenses, education expenses, health expenses, and business decisions.

Swain and Wallentin's (2007) study of whether microfinance empowers self-help groups in India used indicators that included: women's primary activity; access to independent savings; hypothetical response to possible verbal, physical, and emotional abuse; awareness of rights; and political activity. These indicators go beyond

traditional 'decision-making' indicators and encompass other areas of women's control over their lives that are appropriate for the economic empowerment topic and context under evaluation.

Hashemi et al.'s (1996) study on women's rural credit programmes in Bangladesh used eight indicators (with sub-variables and weighting for each) that were defined during extensive pre-study discussions with participants and programme managers, and pre-testing in the field. The indicators selected were: mobility; economic security; ability to make small purchases; ability to make larger purchases; involvement in major decisions; relative freedom from domination by the family; political and legal awareness; and participation in public protests and political campaigning. There was also a composite empowerment indicator by which a woman was classified as empowered if she had a positive score on five or more of the main eight indicators.

Garikipati (2008)'s evaluation of the impact of microcredit on women's empowerment is another good example. The study developed indicators through an extensive literature review and a participatory context analysis – and consequently produced a multidimensional set of indicators for the evaluation. A series of questions were developed to assess 'vulnerability' and 'empowerment'. Scores for each were assigned by question and then a cut-off point was defined to establish either 'vulnerable' or 'not vulnerable' and 'empowered' or 'not empowered'. Vulnerability indicators were used as a proxy for income.

(See Annex 10 for additional examples of empowerment indicators from reports included in the database.)

4.5.9 Microfinance evaluations: weaknesses around indicators

Several of the microfinance evaluations reviewed by the expert panel had measured outcomes of the microfinance component of the intervention (usually savings or loans) only, and did not take into account other components. In one case, this meant that the effects of credit inputs were measured, but not the effects of women forming groups to manage savings. Many microfinance projects involve a range of different 'services' and activities beyond the savings and loans elements, the most common of which is the self-help group or savings group. Project elements that work alongside or as part of the microfinance component are likely to be an important aspect of women and girls' economic empowerment (sometimes called microfinance plus). These can include: social activities, social mobilisation, networking, the development of social capital and mentors, linked services (such as health, training, financial literacy, education), gender training, and building negotiating power with men. An important issue to note given the number of microfinance evaluations included, none of the microfinance interventions reviewed had assessed its impact in combination with other activities (microfinance plus, or linked services), or how any impact could be differentiated between microfinance input only, and microfinance plus.

Garikipati (2008) investigated this issue through literature review and impact evaluation: '*Overall the impact evaluation literature that is emerging from India once again reiterates the central paradox that we attempt to investigate in this paper; while the economic benefits of microcredit at the household level are somewhat predictable (if not guaranteed), the benefits for women clients are much more ambiguous and may depend on other factors exogenous to lending*' (p. 2624). This theme runs through nearly all of the microfinance evaluations reviewed by the expert panel.

Self Help Group (SHG) evaluations were common, especially in India. One such evaluation (Deininger and Liu, 2009; Deininger and Liu, 2013) appeared to have used comprehensive indices of women's empowerment and social capital, with creative use of a political participation variable. However, a closer look at the way questions were asked revealed that women were asked about health, consumption, and empowerment, while information about agricultural production was obtained from men. Also, the political empowerment question was about frequency of attendance at village meetings, with a lack of more in-depth analysis of issues such as the quality of women's participation. There is no discussion of the external factors that might have contributed (or not) to empowerment, so the analysis of the data obtained is incomplete and, in the opinion of the expert reviewers, inconclusive. The findings cannot therefore be generalised or assumed to be caused by the SHG programme. The evaluation does, however, provide a useful example of how to use PSM to examine whether unobservable factors are playing a role.

4.5.10 Indicators used in social protection evaluations, including cash transfer and food subsidy measures

Several evaluations of social protection interventions had used innovative and interesting indicators to measure women and girls' economic empowerment. Others, however, made underlying assumptions that weakened the indicators used. For example, in one evaluation, rather than using a direct measure of female power, it looked at a proxy (receipt of the transfer) that assumes an increase in female bargaining power; however, this does not necessarily result in changes in empowerment paradigms within the household.

Examples of evaluations using multidimensional indicators

Arif et al.'s (2010) study used multidimensional indicators that covered the individual, household, community, and policy/programme levels. The authors considered economic risks and vulnerabilities (wage equity, access to employment, ownership of assets) and social risks and vulnerabilities (including intra-household power dynamics and food distribution, social capital, and voice in community meetings), as well as gendered coping strategies. Transformational change was measured in terms of women's ability to meet their practical gender needs (providing for their household's basic needs) as well as investigating the extent to which there has been a shift in intra-household and community-level gender dynamics. In addition, the report usefully reflects on the extent to which women and girls' agency is considered in programme design, and M&E.

Jones and Shaheen's (2012) analysis of beneficiary perceptions of the Palestinian National Cash Transfer Programme provides a comprehensive picture of the effects of the cash transfer on the individual, household, community, and polity-level dimensions of gender empowerment, which is valuable given that many evaluations tend to be more narrowly focused. What is perhaps most innovative is the relatively detailed attention to the psycho-emotional dimensions of empowerment – the importance of dignity, respect, of confidantes/non-judgemental listeners, of freedom from male surveillance, and of having access to social support.

4.5.11 Gender mainstreaming in institutions

Hardly any of the evaluations included an assessment of gender mainstreaming in institutions or gender integration throughout project activities. Report authors and reviewers consider this to be an important aspect both of project implementation and of successful gender outcomes, and should therefore be part of the assessment of efficiency and effectiveness in an evaluation. It is currently not systematically considered part of the evaluation criteria to assess WGEE.

4.6 Sampling

- Information about and justification of sample size and sampling strategy is essential for a good evaluation.
- Quantitative evaluations had better explanations of sampling than qualitative evaluations.
- Very few evaluations included men and boys in samples.
- There were only six reviewed evaluations of projects focusing on adolescent girls.

Many of the evaluations did not explain how they had established their sample and sample size. It may have been that they had a rigorous method, but a report that does not explain the sampling method does not appear to be so rigorous. It was more common for qualitative evaluations to omit an explanation of sampling, presumably because this did not appear to be an important part of the methodology. However, it is essential for all evaluations to explain the sampling strategy chosen and to justify the sample size. For example, in some mixed methods evaluations, there was a very small sample for the qualitative part of the research, with no explanation of how this had been selected or why.

Quantitative evaluations were more rigorous about explaining the sample size and, in the case of experimental and quasi-experimental designs, the process used for selection of both the treatment and control groups. Evaluations that included a full analysis of similarities and differences between control and treatment groups and any areas for contamination or other bias that might creep in usually scored better than those that did not include this level of information.

Lack of rigour in choosing the sample and explaining it can leave the evaluation results open to confusion. In one micro-credit evaluation, for example, it was not clear whether the women who were sampled and given credit were already empowered to use credit. The evaluation measured how well these women's businesses performed. If the group of women were already empowered, it may be that this was the reason their business

did well, and not the access to credit. The methodology and sampling made it difficult to ascertain whether the programme and the credit provided were the reasons why these women's businesses performed well.

4.6.1 Involvement of men

Many of the evaluations reviewed did not conduct interviews, surveys or focus group discussions with men and/or young men unless the project was specifically targeted at men as well. It could be argued that it is impossible to measure women's economic empowerment unless men's attitudes and behaviours are also taken into account. In project evaluations that were scored highly by the expert reviewers, the engagement of men was seen as critical in gaining acceptance for the strategies chosen to improve women's economic empowerment. Some of the evaluations had a general sample with men and women, but with no disaggregation of results, so women's empowerment was not easy to identify.

Examples of evaluations that included men in samples

In Karlan et al.'s (2007) assessment of the impact of a commitment savings product in the Philippines, men were also offered access to the saving product. The findings revealed that there was only impact on women's economic empowerment and not men's, which makes the study more powerful.

One of the strengths of the evaluation by Montgomery et al. (2011), on whether commercially-oriented microfinance can help meet the Millennium Development Goals (MDGs), was that it included interviews with both 'men and women heads' in each household. This ensured that it covered wives of male borrowers and husbands of female borrowers, providing a richer understanding of household dynamics in relation to the use of microfinance.

4.6.2 Girls and young women

The expert reviewers found virtually no high-quality evaluations of interventions targeting girls or young people's economic empowerment. Of the 55 evaluations identified in the original search, only six were shortlisted. Reports were excluded because they were either poor quality and/or lacked detail. In the other evaluations, lack of age-disaggregated data meant that impact on women and girls was not separated out, even though there are many young women and girls who are married and may form an important part of a sample. There was no evidence of differentiated approaches being used.

Examples of evaluations of interventions targeting adolescent girls

One interesting example is Bandiera et al. (2012), who evaluated the impacts of a programme designed to empower adolescent girls in Uganda through the simultaneous provision of: (1) life skills to build knowledge and reduce risky behaviours; and (2) vocational training, enabling girls to establish small-scale enterprises. The paper represents a high-quality research methodology relating to the economic empowerment of girls. It uses an RCT to elicit the causal effect of a standard intervention that has been implemented in Uganda, and uses innovative outcome variables to explore the effects of the programme on female empowerment. Its main weakness, however, is that the methodology does not allow the individual components to be tested – that is, we cannot see the causal effect of the vocational or life skills components separately.

Dunbar et al. (2010), on the other hand, focus particularly on adolescent girls in their assessment of a micro-credit and life skills HIV prevention intervention to reduce risk among adolescent female orphans in Zimbabwe. However, even when qualitative research was conducted directly with girls, there was a very limited gender focus. The authors assume that by focusing on adolescent female orphans, the analysis will be gender sensitive. In fact, the only gender dimension reported in the qualitative data was as follows: 'Girls' vulnerability, evidenced through theft or confiscation of goods and threats to personal safety, was also a barrier to success'. Still, the expert reviewers felt that the evaluation findings were very interesting.

4.7 Tools and data collection

- Hardly any of the reviewed documents had included the complete set of evaluation tools used, either in the main report or annexes.
- Some reviews with good-quality methodology had poor design and implementation of instruments, which led to poor-quality data and weak analysis.
- Data collection methods rarely discussed gender issues with respect to methodology and enumerators/interviewers.

4.7.1 Inadequacies in the survey or interview methodology

The expert review process highlighted a number of weaknesses with data collection, as follows.

- There was very little information about the way qualitative methodologies were implemented, with virtually no information about data collection or gender and other issues in relation to interview and focus group techniques. It is not possible, therefore, to see how gender-aware each methodology was in practice (e.g. place of interviews, sex of enumerators, etc.).
- There is a lack of explanation about how questions are asked in surveys or interviews.
- Empowerment questions are sometimes based on recall, with insufficient triangulation to back up the findings. Most of the expert reviewers agreed that recall is subjective and can be unreliable, so other data sources are needed.
- There is no information about or discussion of who the interviewers are and how their gender or social identity might affect the results or discussions.

A small number of evaluations relied in part on data from project record-keeping. Often, the methods and rigour for this kind of record-keeping are poor and the data are unreliable. Data samples can be verified to establish how useful the data are. Also, triangulation with other sources of data can be done. However, this has to be planned for at the data collection stage.

Example of an evaluation with good data collection methodology

WB and IFC (2011), in a study of women's empowerment in Special Economic Zones, show how gender considerations were integrated into data collection. *'It is important to take into account cultural and religious considerations when approaching women for conversations. For example, women may need permission from family members to speak to a stranger or to a man. In addition, women may be very shy about discussing certain issues, particularly health, as this is rarely discussed even within the family in many of the countries visited.'*

Some evaluation reports included information on data collection tools in annexes, and these could be of some use to those doing similar evaluations. However, context and project aims will inevitably change the way evaluation tools are designed and used.

4.8 Quality of data analysis

- Hardly any of the evaluations disaggregated data by age, and lack of gender disaggregation in interventions that were not systematically targeting women.
- Triangulation of multiple sources of data, as well as different types of data (i.e. quantitative and qualitative), can generate considerably more robust findings
- Limited use of women's voices in the analysis

4.8.1 Lack of data disaggregation

Hardly any of the evaluations disaggregated data by age, which made it difficult to see any relevance of most of the evaluations to girls and youth – unless this was a particular target population. There was also a surprising lack of gender disaggregation in interventions that were not systematically targeting women.

4.8.2 Use of multiple sources of data and triangulation

Triangulation of multiple sources of data, as well as different types of data (i.e. quantitative and qualitative), can generate considerably more robust findings. For example, Korenblum (2009), in an evaluation of the Cotton

Textile Supply Chain Project in India for Oxfam GB, used triangulation effectively, supplemented by stories for change and a good understanding of context. When there was no baseline, the research relied on personal recall and triangulation by interviewing people who were similar to the respondents nearby (though they did not comprise a control group). Triangulation is also used to overcome the data validity problems.

4.8.3 Women's voices

Interviewing women does not necessarily mean that there is any analysis of women's economic empowerment – or indeed that women's voices are captured in the analysis and presentation of results. For example, most purely quantitative evaluations survey women, but while women's responses are being captured, if women being researched had no involvement in the development of the indicators, or the issues women face in that particular context were not well understood and reflected in the questionnaire, women's responses will not be adequately capturing their 'voices'. Some evaluations, however, made particular efforts to capture women's voices, which was generally achieved through using qualitative methods – for example, in Jones and Shaheen (2012), and Nelson and Smith (2011). Including quotes from women and girls to illustrate findings helps bring the analysis alive, making women's voices more visible.

4.8.4 Cost effectiveness and cost benefit

Hardly any of the evaluations did a thorough analysis of cost benefit or cost effectiveness. There were two useful examples though, which are described in Annex 10. BARA/IPA (2013) analysis of the 'Saving for Change' Programme (SfC) looks at the cost of the programme, and explain that although it is methodologically challenging to determine statistically whether SfC's benefits outweigh the costs, the cost-benefit analysis highlights that the modest impacts stemming from SfC were achieved through a very inexpensive program.

Another interesting example can be found in Ahmed (2009: 112) who calculated the cost effectiveness of cash transfers by using the findings of impact and effectiveness in terms of the economic results accruing to participants.

4.9 Self-reflection

- It is important for an evaluation to discuss limitations and caveats of the study for credibility purposes.
- Evaluations must be transparent about weaknesses in methodology or implementation, challenges in the process, conflict of interest and other key variables that can affect the quality of the evaluation.

It is important for an evaluation to discuss limitations and caveats of the study for credibility purposes. In this sense, the expert reviewers and documentation on evaluation quality all point to the importance of an evaluator's self-reflection and criticism. This should be part of the analysis and should ideally consist of the following aspects, together with discussion on how the evaluators sought to overcome or mitigate these risks, and the potential or real impact the limitations had on the analysis and findings:

- weakness in the design of the methodology
- sampling difficulties and inconsistencies
- difficulties in implementation of the research and data collection (particularly the impact of contextual situation)
- position of the researcher and data collection team and their influence on outcomes of the research
- difficulties or mistakes in analysis
- lack of funds, which may have limited the scope and extent of the evaluation
- conflict of interest issues (such as the evaluators also being responsible for project design and/or implementation).

The BARA/IPA (2013) study of the Saving for Change programme in Mali recognised that there was an unusual level of political and economic stress in addition to drought and flooding – and that this was particularly problematic during the evaluation period. The researchers discussed how their results had been affected by these conditions.

Some evaluation studies with poor-quality methodologies and /or important flaws do not acknowledge these limitations, which 'misguides' those reading them. A strong peer review, in-built as part of the evaluation

process, could help strengthen the overall evaluation report. While not many evaluations currently do this, independence and analysis of influence on the process can be really important to strengthen an evaluation output, and to help others learn lessons for conducting future evaluations.

4.10 Use of findings and dissemination

Hardly any of the evaluations reviewed by the expert panel described whether or how the evaluation would be used. In some cases the text mentioned that the work was to be used for future policy and programming. To ensure policy relevance, a good-quality evaluation report will have drawn from both the evaluation findings and from evidence in the field, and should reflect current understanding of the issues involved and relevant lessons from other interventions around the world.

In general, the evaluations that were shortlisted for review did not include feedback from project implementers or a response to the evaluation. Neither was there any plan for addressing evaluation findings or for planning similar interventions in the future. Including this information would give a better idea of how the evaluation findings will be used, particularly important for encouraging learning, and uptake of the findings through policy and programming. Using women's voice in the dissemination of evaluation findings is important, as is feeding back information from the research team to the participant communities in a way that can aid further learning and discussion.

Examples of evaluation reports with good presentation and strong policy relevance

The expert reviewers considered Oxfam's (2012) review of the Guatemala Highlands Value Chain Development Alliance to be excellently presented, and could be used as an example of how to write a good evaluation report.

Nelson and Smith (2012; 2013) include a management response from the Fairtrade Foundation for the recommendations about Fairtrade cotton. This provides an extensive description of how issues are already being addressed as well as plans for the future. It is clear that recommendations of this sort can also support institutional change processes by pushing change forward through external pressure. By making the report public (and publishing), the organisation is making a sincere commitment to addressing the issues raised in the evaluation or research. Ahmed et al. (2009) also includes a useful section on policy relevance of findings.

4.11 Innovations relevant to women and girls' economic empowerment

- Methodologies used in the shortlisted evaluations were fairly standard and there were few innovations in terms of approach.
- Innovation was most likely to be seen in the mix of methods and the development of useful indicators for measuring women's economic empowerment.

The methodologies used in the shortlisted evaluations were fairly standard and there were few innovations in terms of approach. Innovation was most likely to be seen in the mix of methods and the development of useful indicators for measuring women's economic empowerment. A number of these have already been mentioned. But there were specific innovations in the following evaluations:

Nelson and Smith's *Synthesis Report on FT Cotton: Assessing Impact in Mali, Senegal, Cameroon and India* was notable for the interesting set of recommendations it contains for conducting future impact assessments of a global trade sector (see Section 2.9 of the report). These recommendations contain some suggestions pertinent for this assignment about best methods and practice in conducting evaluations and impact studies related to women and girls' economic empowerment.

Smith (2010)'s report, Fairtrade Bananas, was innovative in its level of collection and use of sex-disaggregated data and the depth of gender analysis within institutions, which was beyond the remit of the commissioned study. When the data were disaggregated and interrogated, it became clear that there were gender differences (e.g. impact of lack of land ownership for women producers, sexual harassment experienced by women workers on plantations), albeit these only affected a small number of the total workers in the sector. Smith then used these findings to reflect on the lack of gender awareness among producers, and make recommendations about the need for Fairtrade International (FLO) to strengthen gender perspectives in fairtrade include developing

specific criteria in the FLOs producer standards, going beyond non-discriminatory clauses, to direct promotion of gender equity. This report represents a lesson, therefore, in why it is so important to disaggregate data by gender.

BARA/IPA (2013)'s report on the Savings for Change project in Mali used innovative methodologies for data collection. These included social network census, high-frequency surveys and the use of global positioning system (GPS). These innovative methods were combined with a randomised control trial and a range of qualitative methods to obtain more robust findings.

5 Key WGEE findings

There were a number of interesting findings from the review of the shortlisted evaluations that should be relevant for those designing interventions aimed at increasing women and girls' economic empowerment. Given that the majority of the evaluations reviewed were of financial service interventions, most of the findings relate to that sector. However, given that many of these financial services evaluations were not situated in the relevant literature/developments in that sector, many reports seemed to be repeating the same work and did not move the agenda forward in terms of evaluating different ways and combinations of delivering services. There were useful findings for the fair trade, collective action, and social protection thematic areas. Findings for business development services were almost always linked to those of another thematic area.

It is important to bear in mind that interesting evaluation findings can emerge despite low-quality methodology. This section provides a synthesis of the shortlisted evaluation findings. Annex 8 shows the score that each evaluation received, which gives an idea of the quality of its findings.

Overall, the most interesting finding is that joint/mixed or integrated interventions providing economic skills and services alongside life skills and other training services (including gender training) appear to deliver the best results for women's economic empowerment.

5.1 Financial services

The set of reports looking at microfinance present a clear picture: microfinance for women increases access to credit for businesses but has no effect on social outcomes. Access to credit enables businesses to grow – leading to higher levels of assets and profits (in some cases) – because businesses can then invest in capital. Furthermore, households that have access to microfinance are able to smooth consumption and are thus less likely to suffer from food shortages. However, the overwhelming majority of studies looking at microfinance as a single intervention found that it has no effect on female bargaining power in the household, expenditure on female goods or on children, or even on female health outcomes like later age at marriage, more knowledge of how to avoid sexually transmitted infections (STIs), or a decrease in unwanted pregnancies.

In many cases, though, microfinance is not offered as a stand-alone or one-off intervention; for instance, sometimes women are also required to attend business classes. Therefore, the social changes that policy-makers want to attribute to microfinance are probably due to the programmes it is coupled with. However, when microfinance has been paired with another service such as business training (as in Karlan, 2011), the effects, while positive, have been minimal. According to the report, part of the reason for this can be explained by the fact that the women who participate in microfinance are already a self-selected group that may be more empowered than other women in their community.

Banerjee, et al (2013) and several other evaluations show that microfinance alone does not result in women's economic empowerment. A more comprehensive approach that provides support in other areas such as skills training and life skills is needed for transformations to occur.

5.2 Business development services

There were not many examples of interventions providing business development services alone. Most of them were combined with microfinance or other activities such as marketing and collective action. So this section overlaps significantly with the other sections. One example of business training in India given to women who were also customers of a microfinance programme demonstrated how the women who were most subject to cultural restrictions (Upper Caste) were more likely to experience changes in economic empowerment than women from the Scheduled Castes or Muslim communities in the study.

An Oxfam evaluation (2012) of a project providing livelihood (agriculture) and trade support (the Guatemala Highlands Value Chain Development Alliance) showed that there had been an impact on women's ability to influence decisions made by producer associations; there had also been improvements in attitudes towards

women's economic roles, and increases in productivity and improvements in farming practice. However, there was no increase in women's decision-making power at the household level, nor any increase in their income. Interestingly, evidence of impact was found to be greatest for the wives and partners of the male beneficiaries, even though there was some evidence that the intervention had bolstered the self-efficacy of females who benefited directly as well.

An interesting model of 'microfinance plus' business development is the WORTH programme in Nepal, which was evaluated by the Valley Research Group and Mayoux (2008). WORTH is a women's empowerment programme that combines business, banking and literacy – in which women become social activists, social entrepreneurs, and effective leaders who bring about change in their communities. The evaluation demonstrated many results: high levels of loans and increases in the size of village banks; improvements in self-confidence among women; a decrease in violence (gender-based); an increase in community participation and literacy; and an increase in the value and quality of social capital in the form of the WORTH groups. Most of the groups and village banks had also continued to function long after the project had ended.

5.3 Trade, marketing, and collective action

Nelson and Smith's (2011) evaluation of Fairtrade cotton projects showed a positive impact on women's economic empowerment, including an increase in control over cotton income, more women's representation in producer organisations, and more cotton produced by women farmers.

Bacon's (2010) evaluation of the Nicaraguan Fair Trade (FT) coffee cooperatives found that FT sales prices were substantially higher than those of conventional co-ops, and that the FT cooperatives had a 'stronger sense of empowerment'. It also found that the FT/organic (male and female) co-op reported higher empowerment levels than the conventional co-op, but lower than the women-only FT co-op; they attributed this to men occupying all the leadership positions in the FT/organic co-op, while fewer women in this co-op held land titles. However, although members of the co-ops demonstrated a strong sense of collective empowerment, interviews with individuals within households revealed uneven gendered empowerment processes, and more diverse gender relationships. Some households had maintained, or carried forward, the equity achieved in the co-op to their intra-household relationships, but others had not – that is, gender inequities occurred at the household level even when they were not perceived at the co-op level.

Baden's (2013) evaluation of women's collective action (focusing on women's trading associations) found an impact on women's decision-making power in a few areas, as well as an increase in income, access to and use of credit, and access to market information. Also, training and technology was found to improve product quality and productivity (see the box below for one woman's story).

Box 3: Women's economic leadership: the story of Maminè Sanogo in Mali

Known as Maïmouna in her village, Maminè Sanogo, 50, was born, lived and married, in Kaniko. She went to school and got involved in community development activities from a young age. She is now a member of several women's organisations at the regional, national and international levels. Her village went through tough times, with social conflicts between opposing groups. Maïmouna became convinced that women could bring peace back into the village:

'I thought that if women of my village managed to unite within a single development organisation, they would contribute to re-establishing peace in the village and would gain respect from their husbands and other men... I pondered and realised that the only economic resource that was exclusively under women's control in Minyanka society was Shea. I also knew, from trainings, that improved butter was more remunerative. So I mobilised the co-operative members (women) and convinced them to get trained to produce improved Shea butter. Today, every woman from the co-operative that is involved in improved Shea butter production says that she derives important revenues from it.'

According to Maïmouna, a group leader needs to exert caution, to be strong-minded, and good at making decisions (including unpopular decisions that ultimately benefit the group). In 2009, the Uyelo Co-operative earned 1.5 million FCFA at a commercial fair in Dakar, Senegal. Every woman in the co-operative was keen to share the dividends among its members. *'But I refused'*, says Maïmouna. A few months later, a local development agent offered to build an improved Shea butter processing unit, provided the co-operative could

contribute 2 million FCFA towards it. Maimouna presented the 1.5 million FCFA she had put aside, and the members provided the remaining 500,000, and the group succeeded in getting its own processing unit. All the women thanked her for her vision.

Maimouna was also able to make men in her village and in other villages aware of the necessity to give plots of land to women so that they could plant Shea trees. Wherever she went, she talked to men, they listened to her, and agreed to allocate two to three hectares to women's groups for planting Shea trees. She has also noticed that many men have begun to plant Shea trees in their own fields. 'This,' she says, 'is one of my greatest satisfactions in life!' Men in her village now support the women's Shea co-operative in various ways. The village chief says that the future is in women's hands.

Source: Baden (2013: 54)

5.4 Skills development

Vocational training and life lessons: These interventions were typically courses that participants had to attend for a period of time – anywhere from one to five days a week. In the training, the participants (women) would receive vocational training (such as sewing skills) or lessons on life skills (e.g., learning about their legal right to say 'no' to sex, about how HIV is transmitted, and about the health and other benefits of delayed marriage, etc.). The life skills interventions can be replicated in many countries with appropriate adjustments to reflect contextual realities as women in the developing world face similar problems. However, the vocational programmes aimed to give women specific skills they could use to make money and lift themselves out of poverty. The training (sewing or picking rice, for instance) that is appropriate will vary from country to country, or even region to region. Therefore, the potential external validity of the results of evaluations of vocational programmes is weaker because the efficacy of the programme will depend on the appropriateness of the skills taught.

Overall, the training programmes seem to work – they have a positive effect on women's economic empowerment. The vocational courses tend to have some impact on the probability of being employed (in part because women typically make the effort to become self-employed); however, life skills courses had a small impact on sexual practices and health, thus leading to weaker results. When the two components – vocational and life skills – are packaged together though, there can be strong results in both areas (labour market outcomes and life skill outcomes). An economic explanation for this is that this is due to the fact that the underlying problem causing a woman to partake in risky sexual behaviour, for instance, is poverty, and if one only teaches a woman life lessons and she is still poor, then her choices are unlikely to change. Therefore, creating a joint programme that aims to lift women out of poverty (with vocational training) and provide women with knowledge of their rights and how they can best benefit from education (with the life skills component) can have a huge effect. Indeed, the Bandiera et al. (2012) report illustrates this argument well.

The same report also shows that joint vocational training and life lessons programmes are extremely cost effective – even more so than many other options. This was one of the few evaluations that focused on adolescent girls. The joint training programme resulted in improvements in the girls' sexual and reproductive health, an increase in their entrepreneurial skills and income-generation activities, and a relaxation of social norms allowing girls to earn money.

Furthermore, these types of programmes can be offered to all women – not a subset that just wants to take out a loan, as in the microfinance situation for example. Thus, based on the reports reviewed, the results produced by the joint vocational and life skills training programme were most substantial, and seem to have the potential to play a crucial role in the economic empowerment of women and girls.

5.5 Social protection

Most of the social protection evaluations reviewed explored cash transfer (CT) programmes. Two types of CTs were considered: conditional and unconditional.

The reports reviewed suggest that approaches focused on individuals or households are unlikely to have significant community-level effects unless there are also community-level interventions – especially around countering discriminatory gendered social norms.

Social protection that challenges social norms by providing work in a sector that is not traditionally associated with women, and by providing assets to women, has a positive impact on women's decision-making and mobility (Ahmed et al., 2009). Devereux et al.'s (2007) evaluation of the NGO-run Dowa Emergency Cash Transfer Project in Malawi finds that the mechanism gives women access to formal financial services, which increases the potential for savings and wider economic engagement and their demand for financial services. However, the evaluation also found that the project was increasing household tensions as men felt challenged by women's newfound assertiveness.

Camfed's Seed Money Scheme (SMS) in Zimbabwe (Mak, et al, 2010) combined financial training, peer mentorship, non-repayable grants, and loans to enable women to achieve economic independence. The findings of the scheme's evaluation suggest that on the whole, respondents have managed to improve their situation, whether in terms of greater income or through increased ability and resilience. Their achievement is even more exceptional when their background and levels of household poverty are considered. In addition, they have managed to overcome severe challenges in the country, for which no training could have adequately prepared them.

Jones and Shaheen (2012) evaluated a cash transfer programme in the Occupied Palestinian Territories (OPT), focusing on beneficiary and community perceptions, and found an increase in education and economic independence for women, reduced reliance on patriarchal societal arrangements, more loan opportunities, and better child support arrangements.

Rubalcava et al. (2002) analysed the expenditure habits of men and women, and found that men's expenditure habits did not change when the household received additional money in the form of the cash transfer, but women's did.

5.6 Legal and regulatory frameworks and good-quality employment

There were very few shortlisted evaluations of interventions focusing on legal and regulatory frameworks. One study on property and inheritance rights in Rwanda (Uwayezu and Mugiraneza, 2011) demonstrated that a new law was being implemented and that widows and children were beginning to obtain access to land that had previously been grabbed. This suggests that women's rights were being protected at the same time as enhancing their potential for economic advancement.

Another study, by the World Bank and the IFC (2011), demonstrates how legal and regulatory frameworks for Special Economic Zones affect women's employment and rights in the workplace. There are a range of interesting results and stories from the eight countries that were studied. For example, in Kenya, several companies were providing services (such as access to credit) and facilitating workers' representation through gender committees. However, the study also found that women's participation in business ownership and entrepreneurship in these zones was limited due to legal and cultural constraints, and women still experienced high levels of harassment in the workplace.

5.7 Recommendations for designing WGEE projects

In many countries and contexts, power differences between men and women inevitably lead to situations of psychological and physical control, including physical and sexual violence. Women's economic empowerment should result in women attaining better decision-making power over the economic areas of their lives – and so must include processes that free women from control, coercion and violence. Likewise, better incomes and economic stability/security can enable women to live independently from abusive men. Economic empowerment can therefore protect women from violence.

However, evidence from this review suggests that economic interventions alone are not sufficient to economically empower women (though they may increase their income). Evidence of this is particularly strong in the microfinance thematic area. But it is also clear that high fertility rates, health problems, and lack of education are holding women back from achieving more in business. If family planning, health, and education services are not made available, the impact of an intervention on women's economic empowerment is unlikely to be maximised. This implies that an integrated approach to programming may be very important.

In the case of social protection programmes, it is important to think about transfer amounts and balance of inputs, including complementary services, as these can have a major bearing on programme impacts.

It is also important to reflect more about 'the complexities of people's lives' – for instance, female-headed households may still be beholden to or dependent on male partners in various ways, positive and negative. Similarly, researchers should not assume that households are either male-headed or female-headed – there are often grey zones in between that need to be explored.

6 Recommendations: How to strengthen the quality and value of future evaluations exploring women and girls' economic empowerment

The recommendations are presented in the same structure as section 4. Annex 11 presents a comprehensive set of recommendations for microfinance evaluations in relation to WGEE, developed by Linda Mayoux.

6.1 General considerations to improve WGEE evaluations

Although many of the recommendations here apply to designing evaluations, they are relevant for the project or programme design stage too. For example, gender context analysis would need to be an essential part of the design of any intervention on women's economic empowerment project. Section 6.2 provides recommendations that are specific to the project implementation stage.

(1) Ideally, *evaluations should be designed at the project design stage and include a methodology and indicators that are relevant to the project's theory of change*. However, in practice, many projects do not have a theory of change or evaluation design – and hence either no baseline or one that is not sufficient for a good-quality evaluation. Recommendations are applicable to both cases, 'but where an evaluation has not been designed at the project design stage, it will be more challenging to develop a good-quality evaluation approach.

(2) The main justification for diverting funds to evaluation from implementation of a WGEE intervention is to yield practical recommendations for increasing its contribution to women's economic empowerment (that is, objective 2 in the DAC guidelines). This purpose has a large influence on the methodology used in evaluations. *Clarity of purpose and a strategy for how the evaluation findings will be used are essential from the outset*.

(3) Different types of interventions may require different evaluation approaches (for example, a project that is focused on women and girls' economic empowerment as opposed to a mainstream project that needs to assess impact on women and girls in addition to other aspects of the intervention). However, care should be taken to avoid 'ticking the gender box' with an approach and analysis that is not sufficiently in-depth. In general, it is important to ensure that projects are built around a full understanding of the gender context, even where a gender component appears to be a bolt-on or afterthought.

(4) When designing mixed methods evaluations, it is important to ensure that both the quantitative and qualitative components are well designed and able to complement each other, either through a sequenced approach in which findings from one inform the design of the other, or by ensuring that each approach collects complementary data. Often, evaluations claim to have used mixed methods when they have in fact focused on one method, with only weak data from the other, which adds little to either the analysis or the robustness of its findings.

(5) The review identified very few evaluations and studies that looked at the economic empowerment of adolescent girls, despite a growing interest in this field. This suggests not only a need for more of such interventions, but also for stronger evaluations and research in this area. Such interventions provide an opportunity to engage adolescent girls, and thus also present an opportunity to assess impact on this important age group, whose experiences have so far been largely absent from many evaluations.

6.2 Building a team

(1) Evaluations would benefit from including experts with international and national experience and expertise in gender and economic empowerment/enterprise. Ideally, gender experts would know the technical area of the intervention being evaluated. If evaluators have expertise in a particular technical area (such as fair trade or

microfinance or cash transfers) but are not gender experts, the team's skills could be complemented with at least one gender expert, to be contracted for the evaluation design stage but ideally also the data analysis stage.

(2) Data collection with women, particularly on intra-household dimensions of economic empowerment, is best done by other women wherever possible in order to facilitate open discussions of constraints to women's agency. Where this is not possible due to the composition of the fieldwork team, the evaluation or study report would be more transparent if it states that this was the case and note any possible response biases this might have caused.

6.3 Preparing for the evaluation and understanding the context

(1) Exploring women's economic empowerment (and the impact of any intervention) relies entirely on a thorough context and problem analysis. The evaluation's context analysis should therefore cover gender equality and the economic, political and technical context to better understand the constraints faced by women to achieve gender equality. This could include any details that describe current economic empowerment processes and the status that women hold in economic and political life, as well as women's vulnerability and resilience. It might also reference any environmental factors that might affect women's economic advancement.

(2) The context analysis would benefit from having enough information about gender differences to enable an evaluation design with appropriate tools and data collection methodology. Evaluation teams should not make assumptions about gender behaviour across contexts or allow underlying cultural and gender stereotypes (held by the evaluation team, the data collection teams or national partners) to act as a barrier to good data collection and analysis. It is important to ensure that men and women's spending, income, debts, etc. are analysed fully, in order to understand whether some constraints are gender-based or similar for men and women in any given context. Men as well as women could be asked questions about decision-making power in the household so that researchers and evaluators can understand the paradigm that they operate in – rather than relying on assumptions about how decision-making takes place. Data on joint decision-making in the household ought to be collected; women's economic empowerment might not only be manifest through their ability to make their own decisions, but by the ability to discuss and negotiate decisions with male members of the household. Making assumptions about gender behaviour may ultimately limit the validity of the evaluation findings.

(3) The context and background analysis would be enhanced through a good-quality literature review (with bibliography) with international and national sources. This would help improve evaluation design and identify where it adds value to the literature.

(4) Methodologies to enable researchers and evaluators to fully understand the gender context should include participatory techniques and interviews, not just desk work.

6.4 Understanding the intervention – theories of change and frameworks

(1) Evaluations of interventions on women and girls' economic empowerment would be more robust if they are based on a theory of change. Even if the evaluation is not theory-based, it is essential to have a theory of change that depicts the expected change processes, and guides the research questions and formation of indicators. This should include how transformational change is expected to take place in the short term, medium term and long term, and could cover changes in women's economic situation and agency.

(2) Depending on the context analysis and the scope of the intervention being evaluated, the theory of change would show gender-related change expected at different levels: individuals, communities, institutions, and the legal and policy environment. For a thorough understanding of transformational change, it is important to consider changes in norms and behaviours (of individuals, in communities and in institutions) alongside more functional changes such as access to services (e.g. financial services, business development services, etc.).

(3) The theory of change would be enhanced through the inclusion of an exploration of assumptions that underpin the expected change process, from inputs to outputs, from outputs to short-term, medium-term and long-term change (outcomes), and from outcomes to impact. It could also include a risk analysis that explains how and why these risks could affect the results of the intervention. Many programme designs fail to take

account of the risks that might be associated with either implementation and/or outcomes. Weighing up the risks of a WGEE intervention is particularly important as it could affect women and girls' safety and security – for example, if there is a backlash (in the family or community) against women and girls taking a more assertive role; or if increased business activity and hence physical mobility makes women and girls more vulnerable in areas of conflict. Contextual analyses tend to focus on the positives – the opportunities that an intervention could open up; but the main principle should be 'do no harm'. And that requires a good understanding and analysis of the risks and potential harms.

6.5 Methodology, including indicators

The review looked for innovation and good practice in methodologies used to evaluate interventions on women and girls' economic empowerment. Unfortunately, there were few good examples of innovation and fewer examples of good practice than expected by the reviewers – with good evaluations often losing points at some stage in the review process. In the first part of the review process many reports lost points due to poor documentation of their methodology or lack of transparency. During the review process, however, this was mostly the result of inadequate gender analysis or weak use of a methodological approach. This was a result of . Experts suggest that a set of minimum evaluation methodology standards would be a good way of ensuring quality. These could include some or all of the following.

(1) Evaluations need to provide a full explanation of the methodology and indicators used, including information about why certain choices were made and the limitations of the chosen methodology.

(2) Where possible, mixed methods evaluations should be used for WGEE interventions. The evaluation would aim to establish whether change took place and what caused it (hopefully the intervention), so quantitative methodology might be required. However, qualitative information will be needed to understand other important aspects such as the details of the process of change, and how implementation has been experienced by individuals, as well as to uncover any unintended impacts or changes and other factors that may have influenced change. This is especially important for WGEE projects as changes in norms, attitudes and behaviours are difficult to fully understand with quantitative data alone. In most evaluations, a mixed method approach also enables triangulation, which is again especially important when there is no baseline to measure change against.

(3) RCTs are generally considered the 'gold standard' of research methodology for evaluating impact, proving causality and generalising findings. However, when resources or technical expertise are insufficient to do a comprehensive and statistically sound quantitative evaluation, when it is not important to prove causality or to generalise findings throughout the country, or when another method best suits the question asked, the evaluation can avoid complicated statistical models. This is because the range of variables and questions is too great and the information can be much more cheaply (and just as reliably) collected through use of good qualitative or participatory methods, together with a simple survey focusing on specific issues arising. If the desired methodological approach is quantitative, it is best to use quasi-experimental design, drawing on propensity score matching (PSM) or difference in difference (DD) when an RCT may be too expensive or ethically questionable.

(4) Evaluations would be improved by developing methodologies that allow for multiple sources of data (both qualitative and quantitative) in order to enable data triangulation. Discussions and analyses should make full use of all the sources of data, triangulate data effectively, and cross-reference between the different data sources.

(5) When using qualitative methods, authors would enhance transparency by providing more details about the specific methodology chosen as well as how data were analysed (e.g. using software, thematic synthesis, etc.). Similarly, it is helpful to include quotes from in-depth interviews and case studies as these add legitimacy to the findings and help to bring the issues alive for the reader. Their value should not therefore be underestimated.

(6) Interventions can embed the collection of ethnographic data into the project and evaluation design, where time and resources permit. This kind of data can be very valuable in highlighting complex intra-household dynamics that can change even during the course of the project. It seems unlikely that these kinds of insights could be captured through more qualitative snapshot approaches, which are currently the norm in development research. By themselves, the findings generated by ethnographic data are likely to be limited, but as part of

mixed methods evaluations they could hold much more value. Funders need to think creatively about how to embed ethnographic data collection in project design and evaluation, and may need to consider longer project timelines to make this feasible in practice.

(7) Indicators and tools to be used in the evaluation can be based on local knowledge. Conduct prior and follow-up participatory and/or qualitative research to identify appropriate local indicators and ways of wording questions so that they are likely to be clearly understood and taken seriously by interviewees and thus encourage the most reliable responses. Focus on concrete facts where people can give reliable answers – in these situations, recall can be as reliable as baseline data. Consider local priorities when weighting indicators.

(8) Indicators would be strengthened if they include measurement of the following:

- transformational change in women's economic advancement and agency
- attitudes, norms and behaviours of women and men
- change in gender mainstreaming and attitudes in institutions
- gender differences in empowerment between women and men to increase local relevance (e.g. in relation to joint decision-making)
- (for microfinance) loan use, types of economic activity, and returns to borrowing (but cannot be rigorous as people generally do not know their exact incomes), as part of the impact analysis, as both cause and outcome
- changes in men and women's social capital, economic autonomy and political participation
- rights of women in enterprise
- (in the case of employment) women's perspectives or changes in the workplace and its impact on their lives and feelings of wellbeing, security and safety, etc.
- effects of the wider market, value chain and employment, as all of these will affect women's economic advancement and empowerment.

(9) Evaluations should consider delivering innovations in the following areas:

- robust qualitative methods – using purposive sampling, direct observation, etc.
- robust participatory methods – including participatory quantification methods (see, for example, the Gender Action Learning System (GALS) Gender Justice Review tools)
- linkages between evaluation and ongoing monitoring, particularly related to social performance monitoring (SPM) and identification of relevant indicators to put into the management information system (MIS).

6.6 Sampling

(1) Evaluation reports need to provide a full explanation of sampling strategies and processes for sampling. While this is done fairly consistently in quantitative evaluations, it is seldom done in qualitative ones, making them seem less robust. There should be a justification for the sample sizes chosen for both quantitative and qualitative components. There should also be a sufficiently large sample size for qualitative components, again, with justification.

(2) The ideal sample would include women and men, and, where appropriate, girls and boys – even if the intervention is targeted at women and girls only. Adolescent girls among the targeted 'women' should be identified and interviewed, with appropriately designed methodologies for collecting data from this age group. It is important to consider the implications of project activities (e.g. microfinance) for men in terms of their empowerment and the effects of men's access to project activities on gender relations with women. Evaluators and researchers should analyse the difference, the co-dependence between the two, the positive outcomes, and relationships between men and women – not just the outcomes for women.

(3) Evaluations should differentiate between women of different ages and their stages in the life cycle, as well as women from different backgrounds (noting the implications for very poor women, women in polygamous households, and women-headed or single-headed households) – even if the conclusion is that these different groups of women need something apart from the project activities (such as microfinance); findings might yield important information about complementary services that would need to be put in place to maximise impact.

6.7 Tools and data collection

- (1) Evaluation/study tools need to be available in the final report (as an annex) for transparency and replicability.
- (2) The team needs to design appropriate tools and data collection methods for young people and different population groups in the sample. This is particularly important when collecting sensitive data from adolescent girls and boys. Participatory tools can work well with this cohort.
- (3) In order to demonstrate effect and especially impact in the absence of baseline data, there should be greater use of **historical tools** – e.g. life histories, generational comparisons, community histories.
- (4) The final report should include detailed discussion of both the qualitative and participatory methods used. Avoid vague references to ‘outcomes from focus group discussions’ as if these are uni-dimensional and unproblematic.

6.8 Analysis

- (1) Robust analysis requires disaggregation of data by sex, age, and life cycle stage, at a minimum.
- (2) The report should provide full explanation of the data analysis methods, including the software used and who undertook the analysis, as well as any limitations of the methods and / or process of analysis.
- (3) Data analysis would be enhanced by combining and cross-referencing quantitative and qualitative data in mixed methods evaluations and present it in an accessible format. Different methods should be triangulated to discuss attribution (where possible?). Researchers and evaluators should not just assume causality from statistical correlation. They should analyse and discuss implications of statistics in terms that non-statisticians can understand – that will also highlight any oversights hidden in dense text. Qualitative information from qualitative or participatory research could be selected to illustrate complexities, not just successes.
- (4) Evaluations could consider using methods that can generate quantitative data from qualitative research (e.g. participatory enumeration, etc.). However, caution needs to be exercised, because sometimes findings may be generalised even though the sample was not representative.
- (5) Researchers and evaluators must always test underlying assumptions when analysing data. For example, if the data suggest that women are contributing more money to the household, it does not necessarily hold true that they are benefiting from the additional income, controlling its use, and being given higher status in the household. Draw on all available data sources before drawing conclusions and declare any inconsistencies or lack of logic.

6.9 Self-reflection

- (1) The evaluation report can include details on the composition of the evaluation and data collection team and how this may have influenced results. Reflect on how team members’ gender knowledge, political position, nationality, and technical knowledge may affect their perspective on the evaluation. Present information on how the data collection team’s gender make-up and technical knowledge may have influenced data collection methods.
- (2) There should be an explanation of what worked well and what did not work well during the research/evaluation.
- (3) Discussions about how political, economic or natural situations and crises may have influenced the findings of the evaluation, particularly about women and girls’ economic empowerment, are important.
- (4) If possible, the report should provide a breakdown of the costs of the evaluation and compare this to previous (similar) pieces of work, reflecting on value for money.

6.10 Report writing

(1) The section of the report that describes the methodologies used must be transparent about data collection and data instruments. The inclusion of research instruments in appendices can be required when commissioning evaluations, as they quickly enable the reader to see what was and was not included. The content of research instruments is often glossed over in methodological sections that are more focused on sampling issues. Data analysis approaches are also rarely explained, particularly in qualitative evaluations, although this would provide a better understanding of how the researchers used the data to derive findings.

(2) Reports would be enhanced by including summary diagrams or tables. Excellent material and findings can be buried if they are not then translated into easily accessible formats; if researchers cannot easily understand and remember the findings, policy-makers are much less likely to do so, given competing time pressures and increasing thematic specialisation.

6.11 Dissemination and use of evaluation and research study findings

(1) Careful planning of dissemination of findings to deliver on the evaluation's purpose would be useful. If this is to include learning and better implementation of the intervention in question (and similar interventions), project and evaluation participants should have access to the findings in an accessible (and reusable?) format. In particular, the evaluation's findings should be shared with women and girls.

(2) Evaluations would benefit from having an annex on proposed dissemination.

(2) Evaluators and researchers should strive to ensure that women and girls' voices are represented in the way the findings are disseminated – either through live stories or with different forms of written and other media.

(3) The report's recommendations should provide clear guidance for how to enhance impact on women and girls' economic empowerment. For this purpose, it is important for it to include policy implications. Policy recommendations could be enriched by drawing on evaluators' broader experience and knowledge in the field, not only the findings that come directly from the specific piece of research.

References

- Ahmed, A.U., Quisumbing, A.R., Nasreen, M., Hoddinott, J.F. and Bryan, E. (2009). *Comparing Food and Cash Transfers to the Ultra-Poor in Bangladesh*. Washington DC: International Food Policy Research Institute.
- Andrews, M., Bamberger M., Tesfay, H. and Seyoum, S. (2003) *Independent External Evaluation of EMPOWER Program for USAID/Ethiopia*. The Women in Development IQC. Dev Tech systems, Inc.
- Arif, S., Syukri, M., Holmes, R. and Febriany, V. (2010) *Gendered Risks, Poverty, and Vulnerability: Case Study of the Raskin Food Subsidy Programme in Indonesia*. London: Overseas Development Institute.
- Bacon, C.M. (2010) 'A spot of coffee in crisis: Nicaraguan smallholder cooperatives, fair trade networks, and gendered empowerment', *Latin American Perspectives* 37(2): 50-71.
- Baden, S. (2013) *Women's Collective Action: Unlocking the Potential of Agricultural Markets*. Oxford: Oxfam GB for Oxfam International.
- Batliala, S. and Pittman, A. (2010) *Capturing Change in Women's Realities. A Critical Overview of Current Monitoring & Evaluation Frameworks and Approaches*. Canada: Association for Women's Rights in Development.
- Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I. and Sulaiman, M. (2012) Empowering Adolescent Girls: Evidence from a Randomized Control Trial in Uganda. London School of Economics. <http://econ.lse.ac.uk/staff/rburgess/wp/ELA.pdf>
- Banerjee, A., Duflo, E., Glennerster, R. and Kinnan, C. (2013) *The Miracle of Microfinance? Evidence from a Randomized Evaluation*. Cambridge, Massachusetts: National Bureau of Economic Research (NBER) Working Paper 18950.
- Blattman, C., Fiala, N. and Martinez, S. (2013) *Credit Constraints, Occupational Choice, and the Process of Development: Long Run Evidence from Cash Transfers in Uganda*.
- Bureau of Applied Research in Anthropology, University of Arizona, and Innovations for Poverty Action (2013) *Final Impact Evaluation of the Saving for Change Program in Mali, 2009-2012*. Oxfam America.
- Corboz, J. (2013) 'Third-way neoliberalism and conditional cash transfers: the paradoxes of empowerment, participation and self-help among poor Uruguayan women', *The Australian Journal of Anthropology* 24(1): 64-80.
- Deininger, K. and Liu, Y. (2013) 'Economic and social impacts of an innovative self-help group model in India', *World Development* vol 43: 149-63.
- Deininger, K and Liu, Y (2009). *Economic and Social Impacts of Self-Help Groups in India*. World Bank Policy Research Paper 4884. Washington DC: The World Bank Sustainable Rural and Urban Development Team, Development Research Group.
- Department for International Development (2013) 'International Development Evaluation Policy', May 2013.
- Devereux, S., Mthinda, C., Power, F., Sakala, P. and Suka, A. (2007) *An Evaluation of Concern Worldwide's Dowa Emergency Cash Transfer Project (DECT) in Malawi, 2006/07*. Lilongwe: Concern Worldwide.
- Dunbar, M.S., Maternowska, M.C., Kang, M.S.J., Laver, S.M., Mudekunya-Mahaka, I. and Padian, N.S. (2010) 'Findings from SHAZ!: a feasibility study of a microcredit and life-skills HIV prevention intervention to reduce risk among adolescent female orphans in Zimbabwe'. *Journal of Prevention & Intervention in the Community* 38(2): 147-161.
- Fafchamps, M., McKenzie, D., Quinn, S.R. and Woodruff, C. (2011) *When is Capital Enough to get Female Microenterprises Growing? Evidence from a Randomized Experiment in Ghana*. Cambridge, Massachusetts: National Bureau of Economic Research (NBER) Working Paper 17207.
- Field, E., Jayachandran, S. and Pande, R. (2010) 'Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India'. *The American Economic Review* 100(2): 125-129.
- Garikipati, S. (2008) 'The impact of lending to women on household vulnerability and women's empowerment: evidence from India'. *World Development* 36(12): 2620-2642.

- Gray, B. and Chanani, S. (2010) *Advancing Women's and Adolescent Girls' Access to Resources and Influence in Rural India, Final Evaluation Report*. India: Freedom from Hunger.
- Golla, A.M., Malhotra, A., Nanda, P. and Mehra, R. (2011). 'Understanding and Measuring Women's Economic Empowerment. Definition, Framework and Indicators'. International Centre for Research on Women 2011.
- Hashemi, S., Schuler, S. and Riley, A. (1996) 'Rural credit programs and women's empowerment in Bangladesh', *World Development* 24(4): 635-653.
- Holvoet, N. (2005) 'The impact of microfinance on decision-making agency: evidence from South India', *Development and Change* 36(1): 75-102.
- Jones, N. and Shaheen, M. (2012) *Transforming Cash Transfers: Beneficiary and Community Perspectives of the Palestinian National Cash Transfer Programme*. London: Overseas Development Institute.
- Kabeer, N (1999). 'Resources, agency, achievements: reflections on the measurement of women's empowerment'. *Development and Change* 30(3): 435-464.
- Karlan, D. and Valdivia, M. (2011) 'Teaching Entrepreneurship: Impact of Business Training On Microfinance Clients and Institutions'. *The Review of Economics and Statistics*, May 2011, 93(2): 510–527
- Karlan, D., Ashraf, N. and Yin, W. (2007) 'Female empowerment: impact of a commitment savings product in the Philippines'. *World Development* 38(3): 333-344.
- Korenblum, J. (2009) *Evaluation of Stories of Change: Cotton Textile Supply Chain Project in India*. Oxford: Oxfam GB.
- Mak, J., Vassall, A., Kiss, L., Vyas, S. and Watts, C. (2010) 'Exploring the costs and outcomes of Camfed's Seed Money Scheme (SMS) in Zimbabwe and Tanzania'. Research Report. London School of Hygiene and Tropical Medicine. November 2010
- Montgomery, H. and Weiss, J. (2011) 'Can commercially-oriented microfinance help meet the Millennium Development Goals? Evidence from Pakistan', *World Development* 39(1): 87-109.
- Nelson, V. and Smith, S. (2012) *Impact of Fairtrade Cotton. Summary and Management Response to an Independent Impact Study*. London: the Fairtrade Foundation
www.fairtrade.org.uk/includes/documents/cm_docs/2012/F/2_FTF%20Cotton%20summary%20and%20response%20May%202012.pdf.
- Nelson, V. and Smith, S. (2011) *Fairtrade Cotton: Assessing Impact in Mali, Senegal, Cameroon and India*. Synthesis report. Natural Resources Institute, University of Greenwich/Institute of Development Studies, University of Sussex.
- Oxfam (2012) *Guatemala Highlands Value Chain Development Alliance, Project Effectiveness Review*. Women's Empowerment Global Outcome Indicator. Guatemala: Oxfam.
- Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., Boardman, F., Gagnon, M.P. and Rousseau, M.C. (2011) Proposal: A mixed methods appraisal tool for systematic mixed studies reviews.
- Roxin, H., Berkmüller, H., Koller, P. J., Lawonn, J., Pooya, N. and Schappert, J. (2010) *Economic Empowerment of Women through Microcredit – The Case of the "Microfinance Investment and Technical Assistance Facility" (MITAF)*. Freetown/Berlin: Humboldt Universität zu Berlin.
- Rubalcava, L.N., Teruel, G.M. and Thomas, D. (2002) *Welfare Design, Women's Empowerment and Income Pooling*. Centro de Investigación y Docencia Económicas.
- The World Bank and the International Finance Corporation (WB and IFC) (2011) *Fostering Women's Economic Empowerment through Special Economic Zones: Comparative Analysis of Eight Countries and Implications for Governments, Zone Authorities and Businesses*. Washington DC: The International Bank for Reconstruction and Development.
- Slegh, H., Barker, G., Kimonyo A., Ndolimana, P. and Bannerman, M. "I Can Do Women's Work": Reflections on Engaging Men as Allies in Women's Economic Empowerment In Rwanda'. *Gender & Development*, 21:1, 15-30
- Smith, S. (2010) *Fairtrade Bananas: A Global Assessment of Impact*. Final report. Brighton: Institute of Development Studies.
- Swain, R.B. and Wallentin, F.Y. (2007) *Does Microfinance Empower Women? Evidence from Self Help Groups in India*. Working Paper No. 2007:24. Sweden: Uppsala University, Department of Economics.

-
- Uwayezu, E. and Mugiraneza, T. (2011) *Land Policy Reform in Rwanda and Land Tenure Security for all Citizens: Provision and Recognition of Women's Rights over Land*. Paper presented at FIG Working Week 2011, Bridging the Gap between Cultures, Marrakech, Morocco, 18-22 May. (www.fig.net/pub/fig2011/papers/ts04g/ts04g_uwayezu_mugiraneza_4914.pdf).
- Valley Research Group, and Mayoux, L (2008). *The WORTH Program in Nepal: Empowerment through Literacy, Banking and Business 1999–2007*. Kathmandu, Nepal.
- Van Hung, N., Thuy Lan, H. and Richmond, A. (2007) *Report of the Independent Final Evaluation: ILO/Japan Asian Regional Programme on Expansion of Employment Opportunities for Women (EEOW), Vietnam Chapter*. Vietnam: International Labour Organization.
- VeneKlasen, L. and Miller, V., 2002, 'A New Weave of Power, People and Politics: The Action Guide for Advocacy and Citizen Participation', Practical Action Publishing.

Annexes

Annex 1: Dissemination approach and audience

Annex 2: Theory of change for women and girls economic empowerment

Annex 3: Methodology

Annex 4: Basic scoring tool

Annex 5: Quality assurance tool (adapted 'MMAT' tool)

Annex 6: Expert panel members' profiles

Annex 7: Expert review template

Annex 8: Scores given to documents reviewed

Annex 9: Examples of good theories of change exploring WGEE

Annex 10: Examples of empowerment indicators from reports included in the database

Annex 11: Examples of cost benefit analysis

Annex 12: Recommendations for Microfinance evaluations (Linda Mayoux)



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

Our mission is to inspire and inform policy and practice which lead to the reduction of poverty, the alleviation of suffering and the achievement of sustainable livelihoods.

We do this by locking together high-quality applied research, practical policy advice and policy-focused dissemination and debate.

We work with partners in the public and private sectors, in both developing and developed countries.

Readers are encouraged to reproduce material from ODI Reports for their own publications, as long as they are not being sold commercially. As copyright holder, ODI requests due acknowledgement and a copy of the publication. For online use, we ask readers to link to the original resource on the ODI website. The views presented in this paper are those of the author(s) and do not necessarily represent the views of ODI.

© Overseas Development Institute 2014. This work is licensed under a Creative Commons Attribution-NonCommercial Licence (CC BY-NC 3.0).

ISSN: 2052-7209

Overseas Development Institute
203 Blackfriars Road
London SE1 8NJ
Tel +44 (0)20 7922 0300
Fax +44 (0)20 7922 0399



This material has been funded by UK aid from the UK Government; however the views expressed do not necessarily reflect the UK Government's official policies.