Department for International Development

Africa Centre for Evidence
Working together to make evidence-informed decision-making a reality

UNIVERSITY OF JOHANNESBURG

EPPI CENTRE

UCL Institute of Education

UKaid from the British people

WOMEN IN WAGE LABOUR:
A SYSTEMATIC REVIEW OF THE EFFECTIVENESS AND DESIGN

FEATURES OF INTERVENTIONS SUPPORTING WOMEN'S PARTICIPATION IN WAGE LABOUR IN HIGHER-GROWTH AND/OR MALE-DOMINATED SECTORS IN LOW- AND MIDDLE-INCOME COUNTRIES

June 2018

The authors of this report are:

Laurenz Langer	Africa Centre for Evidence, University of Johannesburg
Yvonne Erasmus	Africa Centre for Evidence, University of Johannesburg
Natalie Tannous	Africa Centre for Evidence, University of Johannesburg
Ekwaro Obuku	African Centre for Systematic Reviews and Knowledge Translation, Makerere University
Zafeer Ravat	Africa Centre for Evidence, University of Johannesburg
Charity Chisoro	Africa Centre for Evidence, University of Johannesburg
Mary Opondo	Africa Centre for Evidence, University of Johannesburg
Promise Nduku	Africa Centre for Evidence, University of Johannesburg
Janice Tripney	EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London
Carina van Rooyen	Africa Centre for Evidence, University of Johannesburg
Ruth Stewart	Africa Centre for Evidence, University of Johannesburg; EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London

Funding

This is an independent report commissioned and funded by the Research and Evidence Division in the Department for International Development. 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.

Acknowledgments

We would like to thank our review advisory group – Harsha Dayal, Thomas De Hoop, Sumana Hussain, Marcel Korth, Bernedette Muthien and Josephilda Nhlapo-Hlope – for their insightful comments and inputs on the review. Sincere appreciation is also expressed to Alison Bullen for leading and conducting the scientific search for academic literature. We are also grateful for the helpful feedback and support received from members of the EPPI-Centre: Mukdarut Bangpan, Sandy Oliver, Claire Stansfield, Dylan Kneale, Jeff Brunton, Sergio Graziosi and James Thomas.

Conflicts of interest

There were no conflicts of interest in the writing of this report.

Contributions

The opinions expressed in this publication are not necessarily those of the EPPI-Centre or the funders. Responsibility for the views expressed remains solely with the authors.

Citation

This report should be cited as: Langer L, Erasmus Y, Tannous N, Obuku E, Ravat Z, Chisoro C, Opondo M, Nduku P, Tripney J, van Rooyen C, Stewart R (2018) *Women in wage labour: A systematic review of the effectiveness and design features of interventions supporting women's participation in wage labour in higher-growth and/or male-dominated sectors in LMICs*. Technical report. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

ISBN: 978-1-911605-06-5

Picture

WORLD BANK PHOTO COLLECTION. PHOTO ID: LN-SL005 WORLD BANK. CC BY-NC-ND 2.0.

This systematic review is published as a full technical report and as a summary report. An open-access interactive evidence map was also published as an output of this systematic review. All publications are available at: https://africacentreforevidence.org/outputs-2/

EXECUTIVE SUMMARY

INTRODUCTION

In low- and middle-income countries (LMICs), women's participation in wage labour is significantly lower than that of men. In addition, women's participation is often clustered in particular sectors of the economy that are not well-remunerated and have lower prestige. Higher growth economic sectors such as ICT and finance are dominated by men, excluding women from lucrative opportunities for social and economic development. Even where women have equal access to wage labour opportunities, they often suffer from vertical occupational segregation, earning significantly less than men and being less likely to be promoted. This horizontal and vertical occupational segregation of wage labour markets in LMICs for women hinders both economic and social development.

In response to this challenging gendered nature of wage labour markets in LMICs, a range of interventions and policies have been proposed to increase women's employment. These interventions aim to overcome a range of barriers to women's wage labour employment in higher-growth/male-dominated sectors, such as discrimination against women by markets and work institutions, or a lack of access to credit and assets and of technical and soft skills. However, these labour market programmes to increase women's wage employment vary greatly, as do the barriers to women's wage labour participation. Different programme designs assume different pathways to support women's employment and it is not clear what programme approaches and design features are most effective.

OBJECTIVES

We conducted a systematic review of all the available impact evaluation evidence on the effects of interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors. Our systematic review aimed to meet the following three objectives:

1. To produce an interactive evidence map of research evaluating interventions aiming to overcome barriers to women's economic empowerment in LMICs.
2. To provide a rigorous synthesis of impact evaluation evidence to identify the effects of interventions supporting women's participation in wage labour in higher-growth and/or male-dominated sectors in LMICs.
3. To identify design features that influence the effects of interventions aiming to overcome barriers to women's economic empowerment in LMICs.

REVIEW APPROACH

Our systematic review followed a two-stage approach to synthesise the effects of women's wage labour interventions. We first conducted an interactive evidence map of a broader research scope in order to provide an overview of the evidence base, then decided on the final scope of the narrower

systematic review jointly with the review stakeholders. The full systematic review followed an aggregative review approach, including only rigorous quantitative impact evaluations of labour market interventions. In order to identify relevant research studies, we conducted an exhaustive search of 74 academic and Grey literature sources. Screening 16,091 citations led to the inclusion of 19 impact evaluations of interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors. The 19 included studies were subject to a detailed data extraction and critical appraisal process. We used statistical meta-analysis in order to establish the overall effects of different labour market interventions and narrative synthesis where the identified evidence base did not allow us to statistically pool studies' results. In order to investigate what design features of interventions are associated with positive intervention effects on women's wage labour outcomes, we applied Qualitative Comparative Analysis (QCA) and narrative synthesis.

IDENTIFIED EVIDENCE BASE

CHARACTERISTICS OF THE EVIDENCE BASE

The 19 included studies reported on a total of 20 interventions (Figure 4.3). The most frequently applied interventions referred to programmes combining training of women with subsequent job placement services (n=9). Two interventions provided soft skills training only, to address vertical occupational segregation, while three interventions provided job placement services only. The review also included six macro-economic interventions; these focused on the provision of national labour subsidies (n=2), changes to macro structures to support women's empowerment (n=2), changes to working hours legislation (n=1), and old age pensions (n=1).

The majority of interventions were conducted in lower-middle-income countries with lower than average economic participation of women. The interventions, by and large, focused on young women who were out of work at the start of the intervention but who had a higher level of education and skills. Interventions deliberately targeted younger and more mobile women in order to match programme participants with the labour demands in higher-growth sectors. These sectors were, in particular, the manufacturing industries, the business administration sector, engineering, and ICT and electronics. All but two interventions aimed to address horizontal occupational segregation and only two programmes targeted vertical occupational segregation.

QUALITY OF THE EVIDENCE BASE

Overall, the evidence provided by these 19 studies can be described as of low quality. A range of biases challenged the attribution of observed labour market effects to the applied interventions, and only five studies were rated as having a low risk of bias compared to ten studies with a high risk and four studies with a moderate risk. In addition, the small number of studies per intervention categories and observed inconsistencies in intervention effects led to a characterisation of the **overall quality of the evidence as low** using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework. Only the review results on the effects of combined training and placement interventions are based on moderate quality evidence and can therefore be interpreted with cautious confidence. For all other intervention categories, the

evidence is of levels of quality that do not allow us to state any conclusions or make recommendations.

Intervention category (Outcomes)	Quality of the evidence
Combined training and placements	
Wage Labour	Moderate
Income	Moderate
Empowerment	Low
Soft skills training on promotion	
Career progression	Very low
Empowerment	Very low
Job placement services only	
Wage Labour	Low
Income	Low
Empowerment	Low
National labour subsidies	
Wage Labour	Very low
Macro-level empowerment policies	
Wage Labour	Very low
Empowerment	Very low

REVIEW FINDINGS

We conducted five syntheses in our systematic review in order to examine the overall effects of different interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors. The summary results of each synthesis are presented below:

EFFECTS OF COMBINED TRAINING AND JOB PLACEMENT INTERVENTIONS

- Synthesising the effects of eight combined training and job placement interventions using meta-analysis, we identified an increase in women's formal wage employment of 0.159 standardised mean difference (SMD) (0.09, 0.23), which translates into a 7.8% greater increase in formal wage employment for women taking part in the training and placement programmes when compared to a control group who were not receiving the intervention (based on **moderate-quality evidence**).
- These interventions were further effective in increasing women's income, and our meta-analysis on income outcomes identified a positive pooled effect size of training and job placement interventions on women's income of 0.145 SMD (0.07, 0.22). This effect size

expressed a 7.2% greater increase of income for women taking part in the interventions as compared to a control group (based on **moderate-quality evidence**).

- There is insufficient evidence to investigate the aggregate effectiveness of combined training and placement interventions on women's economic empowerment. On observation, we identified anecdotal evidence that combined training and placement interventions can support women's economic empowerment if the intervention is effective in also increasing wage employment and income outcomes (based on **low-quality evidence**).
- There is insufficient evidence to investigate the cost-effectiveness of combined training and job placement interventions.

EFFECTS OF SOFT SKILLS TRAINING TO ADDRESS VERTICAL OCCUPATIONAL SEGREGATION

- There is insufficient evidence to investigate the aggregate effectiveness of soft skill programmes aiming to address vertical occupational segregation in LMICs. Upon observation, our systematic review identified two studies in the garment sector that found that two individual training programmes were effective in increasing women's promotion to managerial posts and in changing male and female perceptions of women acting as supervisors (based on **very low-quality evidence**).

EFFECTS OF JOB PLACEMENT SERVICES ONLY

- There is insufficient evidence to investigate the aggregate effectiveness of job placement services as the sole labour market programme component. On observation, our systematic review identified three studies that reported the effects of diverse job placement programmes to be mixed. This included recruitment services, a screening and matching programme, and a graduate wage voucher programme (based on **low-quality evidence**).

EFFECTS OF NATIONAL LABOUR SUBSIDIES

- There is insufficient evidence to investigate the aggregate effectiveness of national labour subsidies to support women's wage labour employment in higher-growth sectors in LMICs. On observation, our systematic review identified two studies that found that two national subsidy policies led to positive effects of women's wage labour in higher-growth sectors in the short term (based on **very low-quality evidence**).

EFFECTS OF MACRO-LEVEL POLICIES TARGETING WOMEN'S EMPOWERMENT

- There is insufficient evidence to investigate the aggregate effectiveness of macro-level women empowerment interventions on women's wage labour outcomes. On observation, our systematic review identified two studies that found mixed wage labour outcomes of two macro-level policies, but positive long-term effects on women's empowerment such as increased aspirations and economic participation (based on **very low quality evidence**).

DESIGN FEATURES OF INTERVENTIONS AIMING TO SUPPORT WOMEN'S WAGE LABOUR PARTICIPATION

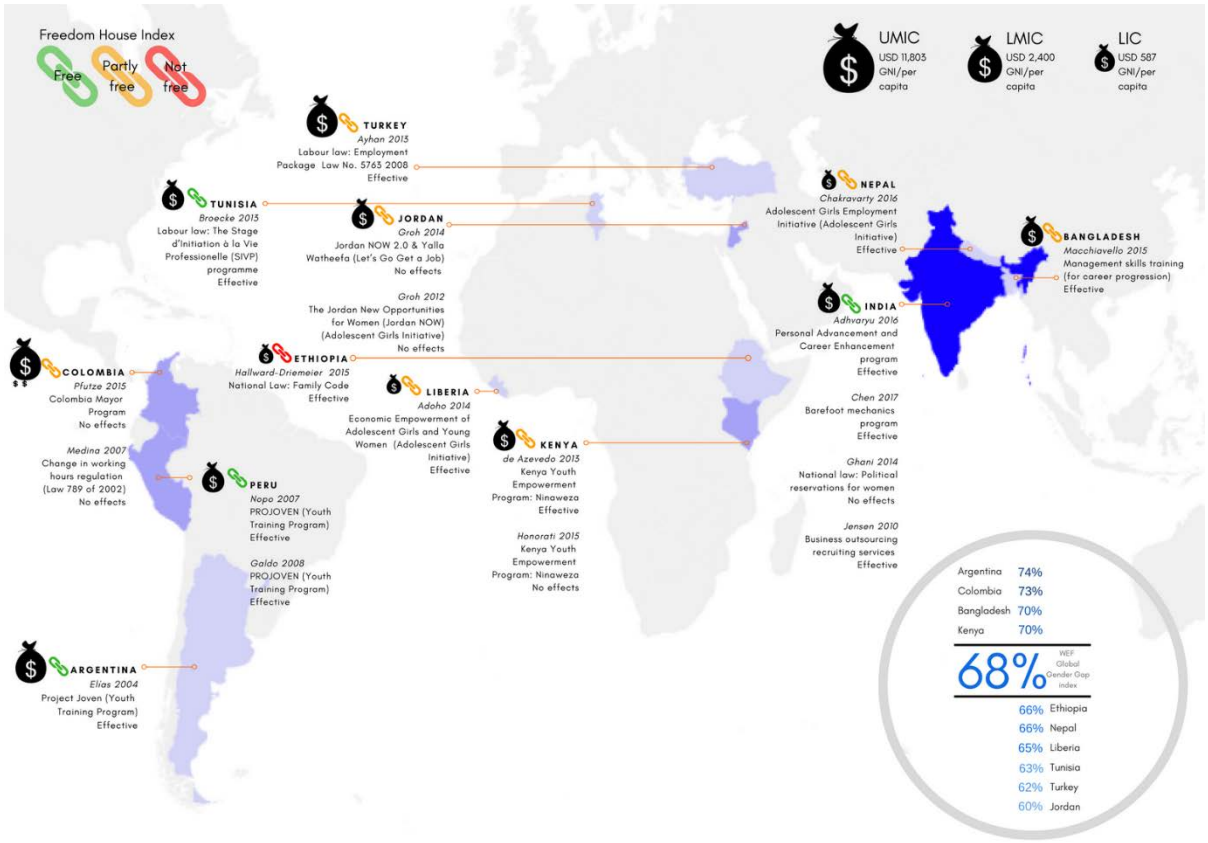
Based on a narrative synthesis and QCA of a homogeneous sub-set of interventions – combined training and job placement interventions – our systematic review identified the following seven design features that supported the effects of interventions on women's wage labour participation in higher-growth/male-dominated sectors:

Design features of interventions	**Example of the design feature**
1. Exposure to labour market participation enhancing social norms	• Conducive norms at work • Affirmative action programmes
2. Labour demand-led intervention design	• Consultation of private sector needs
3. Gender-sensitive intervention design	• Women-only programmes • Childcare facilities
4. Provision of soft/life skills and social empowerment training	• Confidence / aspiration building • Non-cognitive work skills
5. Participant profiling and targeting	• Narrow eligibility criteria • Gender-sensitive marketing
6. Clear governance structures for intervention providers	• Payment by results • Provider monitoring
7. Flexibility and responsiveness in intervention implementation and design	• Information management systems • Piloting and iteration

We further attempted to unpack the specific configurations of these seven design features and their correlation with programme effects. The subsequent QCA, however, was inconclusive and we therefore could not comment on specific configurations of design features. All in all, we concluded that the seven individual intervention design features constitute the active ingredients of combined training and placement programmes to support women's wage labour participation in higher-growth/male-dominated sectors in LMICs.

AUTHORS' CONCLUSIONS

This systematic review provided an evidence map and structured synthesis on the effects of interventions aiming to support women's participation in wage labour in higher-growth/male-dominated sectors in LMICs. We identified a small evidence base that is heterogeneous in terms of the applied labour market interventions and of low quality in terms of the methodological trustworthiness of studies and consistency of effects. Given the limitations of this evidence base, our review only identified a single intervention grouping in which we could conduct a rigorous synthesis: combined training and job placement interventions. Our meta-analysis provided cautious evidence that combined training and job placement interventions are an effective intervention approach to increase women's wage labour participation in higher-growth/male-dominated sectors in LMICs. For all other labour market interventions identified in this systematic review, the size and nature of the underlying primary research evidence base does not allow for a rigorous synthesis of effects.



1 BACKGROUND

1.1 THE IMPORTANCE OF WOMEN'S EMPOWERMENT

Women's economic empowerment is both a means and an end in international development. Following Sen's definition of Development as Freedom (Sen 1999), economic empowerment is an essential capability to live a life one has reason to value and can therefore be seen as an outcome of development in its own right. Gender equality is also one of the 17 Sustainable Development Goals (SDGs), and four of the nine targets of this goal relate directly to women's economic empowerment (UN 2015). As a mechanism for development, there is strong longitudinal evidence that increased inclusion and participation of women in the labour market supports economic growth and development (IGC 2016; Kabeer 2012; World Bank 2012). Full participation of women in labour forces for example is estimated to add multiple percentage points to most national economic growth rates, thus increasing aggregate socio-economic development (UN 2015). Similarly, the McKinsey Global Institute estimates that if all women were able to participate fully in the economy, it would contribute up to US$ 28 trillion or 26 percent of annual global gross domestic product (GDP) in 2025 (McKinsey Global Institute 2015).

The potential benefits of women's participation in the labour market underline that the economic empowerment of women is by no means a zero-sum game: society as a whole stands to gain from it. Investing in women's economic empowerment is smart economics (Revenga and Shetty 2012). Increasing women's income leads to greater household spending on education and health (World Bank 2012). It also increases women's self-esteem, aspirations, and domestic bargaining powers while reducing domestic violence and delaying early marriage and pregnancy (UN 2013; World Bank 2012). In sum, as much as women's economic empowerment has intrinsic value and benefits for individual females, so it is equally a virtuous circle that socially and economically benefits society as a whole.

1.2 THE PROBLEM

Despite the well-established positive relationship between women's economic empowerment and socio-economic development, females in all regions of the world face significant barriers to their labour market participation (ILO 2016; UN 2013; World Bank 2012). Globally, only 50% of women participate in the labour force as compared to 76% of men. This participation gap has been narrowing in most regions of the world over the last decade, albeit at a slow rate: 2.4% increase in Europe; 0.8% increase in South-Eastern Asia and the Pacific; 7.9% increase in Latin America and the Caribbean; 3.2% in sub-Saharan Africa (ILO 2016). While these positive overall trends are encouraging, there is no region that has achieved equal labour market participation for women and men. In addition, women remain at a significantly higher risk of unemployment, constitute a smaller portion of the workforce in wage employment, and face higher barriers in the transition from schooling into the job market (ILO 2016; UN 2015). This situation is not by choice, and a 2017 global survey of almost 149,000 adults found that the

majority of both men (66%) and women (70%) would prefer women to have the opportunity to seek paid employment (Gallup and ILO 2017).

What is more, the quality and nature of women's labour market participation differs from men's. Women participants earn 24% less than men do globally (UN 2013). The majority have non-standard, informal, temporary, part-time and low-paid jobs (ILO 2016). For example, globally 57% of all part-time workers are female. The same finding occurs in relation to time-related underemployment which is, in low- and middle-income countries (LMICs) particularly, significantly higher for women.

Much of this difference in the quality and nature of women's labour market participation can be explained by sectoral and occupational segregation. Globally, women in employment are overrepresented and clustered in particular professions and sectors of the economy that offer lower salaries and less lucrative employment conditions. For instance, 60% of women in low-income and lower-middle-income countries are employed in the agricultural sector, taking up poorly paid but time- and labour-intensive jobs (ILO 2016). Likewise, women are overrepresented in running informal household businesses such as tuck shops and local garment businesses with little potential for growth and high market saturation (Vaessen et al 2014). An analysis of labour market compositions across 142 countries underlines this sectoral segregation, with women consistently being overrepresented in the lowest paying professions, which explains the wage difference between women and men (ILO 2016; UN 2013).

In contrast, women are underrepresented in many higher-growth and well-paying professions. In most countries, men dominate the occupation of plant and machine operators and assemblers, law and legislation, business administration and management, finance and information and communication technologies (ICTs) (ILO 2016). This pattern is particularly acute in LMICs and hinders the use of wage employment as a pathway for the economic empowerment for women. The most common cause of this sectoral and occupational segregation is a reflection of gender stereotypes at work, in the family and in society, as well as the lack of an effective policy environment for women's empowerment (ILO 2016; UN 2013). Using data from 69 LMICs between 1980 and 2011, Borrowman and Klasen (2017) illustrate that occupational and sectoral segregation has in fact increased in more LMICs over time than it has decreased.

In addition to horizontal segregation in the labour market, women are further experiencing vertical segregation within professions. Horizontal labour market segregation refers to sectors in which female employment is significantly lower compared to men as a share of the total labour force; this has been covered above. On the other hand, vertical labour market segregation refers to sectors in which female employment might be equal to or larger than that of men as a share of the total labour force, but in which women are underrepresented in particular occupations or positions. For example, although females constitute the majority of the labour force in the garment sector, they are often underrepresented at a managerial level. Globally, only 5% of the world's largest companies are managed by a female chief executive officer (ILO 2016). Vertical segregation can present a large barrier for

women to either obtain a sufficient return on their labour or to use wage employment as a means to improve their livelihoods.

1.3 HIGHER-GROWTH/MALE-DOMINATED SECTORS

This systematic review is concerned with the participation of women in the labour force in higher-growth and/or male-dominated sectors only. Based on a review of the literature we have identified the sectors listed below to be relevant for inclusion. However, contextual differences negate a universal application of this list, as the gender composition of the labour force in economic sectors across states differs, as does the productivity rate of sectors.

Economic sectors with high or growing productivity *and/or* which are male-dominated:

- commercial agriculture: where productivity is higher than smallholdings or subsistence farming, and agriculture is linked to larger business supply chains and larger markets
- energy (mining and quarrying, electricity, gas and water supply)
- trade
- transportation
- accommodation and food
- business administration services
- finance
- electronics and ICT
- maritime services
- wood pulp and forestry
- construction
- manufacturing
- higher education/science and technology.

Economic sectors with high or growing productivity that are *not* considered to be male-dominated and are therefore excluded from the second stage of the review are:

- education
- health
- social work
- wholesale retail
- communication services
- tourism
- public administration
- garment industry
- micro- and small businesses focused on self-employment with no employees.

However, we included studies that focused on the issue of vertical segregation in female-dominated sectors. That is, even though the health sector as a whole is not considered relevant, research that focused on vertical inequalities in employment in the health sector, for example the number of females in senior positions, was considered in our review.

1.4 BARRIERS TO WOMEN'S LABOUR MARKET PARTICIPATION

A range of barriers impede women's participation in higher-growth and/or male-dominated sectors. These refer in particular to the five main barriers which are discussed below:

1. Discrimination by markets and work institutions
2. Constraints in access to credit, finance and assets
3. Constraints in employability and entrepreneurship
4. A lack of social capital and norms
5. Behavioural (social and cognitive) barriers.

First, women face discrimination by markets and work institutions. For example, many companies do not cater for women's caregiving responsibilities. The absence of paid parental leave and childcare facilities in LMICs renders many women unable to participate in the labour market. Likewise, in the public sector, procurement systems and budget allocation often overlook women's needs and hinder their ability to benefit from public sector investments. For women in work, harassment and discrimination in the workplace present an equally important barrier to meaningful and valued employment (Buvinic and O'Donnell 2016; ILO 2016; ODI 2016; Peters et al 2016; UN 2013).

Second, women face constraints in access to credit, finance and assets. In some LMICs, women are not able to own property or open a bank account on their own. Economic resources, such as loans, are often controlled by men or – if they are accessible to women – they are often too small to be effective for business development. This structural inequality impedes women's abilities to start or expand their businesses and gain from the opportunities provided by higher-growth economic sectors.

Third, women experience disadvantages in their employability and entrepreneurship. The technical and business skills required to assume more senior positions or to enter high-skill professions are often inaccessible to women. This lack of access to skills disables them from benefiting from high-return professions such as those in the ICT and engineering industries. Acquisition of labour-related soft skills and access to economic information are also often provided less frequently to women, preventing them from gaining valuable work experience. Female entrepreneurs also face additional constraints to their business success with business networks often dominated by men, and market information and opportunities being shared through informal channels rather than public ones (Buvinic and O'Donnell 2016; ILO 2016; Kabeer 2012; ODI 2016; Peters et al 2016; UN 2013).

A fourth key barrier to women's labour market participation in LMICs is restrictive social norms and a subsequent lack of social capital. Women are often expected to refrain from participating in the labour

market and when they do, socially acceptable employment opportunities are restricted to a small number of usually low-paying professions such as domestic work. Compared to men, women face greater constraints in their mobility and often lack social support to build economic aspirations. This absence of social capital leaves many females at a structural disadvantage, having to overcome restrictive norms as a first step to their labour market participation (Buvinic and O'Donnell 2016; ILO 2016; Kabeer 2012; ODI 2016; Peters et al 2016; UN 2013).

Finally, all human beings are subject to behavioural biases that influence social and economic decision making (Kahnemann 2011; Thaler and Sunstein 2008). For example, most of us are influenced by the framing of messages or struggle to implement and honour an effective saving schedule. A growing body of research provides insight into how these behavioural and cognitive barriers can be overcome with often simple techniques and small tweaks to programme designs, such as identity cues or commitment devices. The same is true in relation to female economic empowerment. In *What Works: Gender Equality by Design*, Iris Bohnet (2016), for example, lays out how reframing the wording on job adverts can increase female job applications. The Centre for Global Development refers to similar findings in the context of LMICs: their report *Women, Economic Empowerment and Smart Design* (Buvinic and O'Donnell 2016) highlights the increased attention that research has paid to overcoming women's behavioural (social and cognitive) barriers to labour market participation.

Each of these five key barriers to women's labour market participation requires the use of a deliberate intervention implemented in the form of public policies and programmes. For example, entrepreneurship training provided to females may overcome their lack of labour market relevant skills; microfinance may provide otherwise inaccessible capital; career guidance could offer access to labour market information; labour laws might include affirmative action clauses to address structural inequality; governments' provision of better childcare facilities and public investment in infrastructure might reduce women's time spent on domestic chores and increase their mobility, etc. The list of possible interventions is much longer and underlines the urgency with which women's social and economic barriers have to be addressed.

1.5 THE INTERVENTION

This review is not limited to a particular kind of intervention. Its conceptual framework (Figure 1.1) is defined by the barriers to women's wage labour market participation in higher-growth and/or male-dominated sectors. These barriers set the scope for what is considered to be a relevant intervention in our review: the review extends to any intervention likely to support women in LMICs in overcoming any of the five barriers to their labour market participation introduced above. Broad groupings for potentially relevant interventions – which may take the form of a policy, programme, strategy or other type of action – are presented in Table 1.1. Eligible interventions may vary on several different dimensions. For example, relevant interventions may be:

- complex, specialised, multidimensional programmes or much simpler interventions based on a single strategy;
- implemented in different settings;
- either routine and/or structured/tailored interventions;
- varied by type and intensity;
- delivered at various stages of the employment process (pre-employment, transition to employment, and post-employment); or
- focused primarily on something other than the objective of overcoming women's barriers to employment.

Table 1.1: Intervention categories

Categories	Examples of relevant programmes
I. Overcoming discrimination by markets and work institutions	
Interventions to balance work and family responsibilities	• Flexible working-time arrangements • Maternity and parental leave • Sick leave • Social security • Provision of care facilities (child and elderly)
Increase women's financial returns	• Salary/wage increases • Salary/wage alignments • Tax incentives
Changing business culture/practice	• Quotas and reservation approaches • Workplace gender equity programmes • Business leadership and management • Public sector practices in employment and procurement • Setting and enforcing effective laws to protect women from violence and exploitation at work • Anti-discrimination policies
Macroeconomic changes[1]	• Gender mainstreaming • Gender-responsive budgeting • Trade liberalisation/export-orientated production • Public works programmes • Local production systems • Active labour market policies • Industrialisation policies • Investment policies (domestic and foreign) • Fiscal policies • Monetary policies
Provision of infrastructure	• Roads, public transport, street lights

[1] This excludes macro-level interventions such as investment in basic to tertiary education and health care, which are known to benefit the labour market participation rates of the general population.

Categories	Examples of relevant programmes
	• Water and sanitation • Electricity and energy • Other time-saving consumer goods
II. Overcoming constraints in access to credit, finance, and assets	
Microfinance	• Micro-credit • Micro-savings • Micro-insurance • Financial inclusion
Cash transfers	• Conditional • Unconditional
Economic assets	• Access to formal loans and capital • Provision of capital in kind • Business technology
Changes to land titles, business ownership, and inheritance	• Land reform • Property rights • Inheritance laws
Bundled services/combined structural interventions	
III. Overcoming constraints in employability and entrepreneurship	
Interventions to provide education/skills	• Technical skills training • Business skills training • Literacy/numeracy skills training • Soft skills training • Financial literacy training
Interventions to provide access to economic opportunities	• Provision of economic information (e.g. business networks, peer support) • Job search assistance • Business advice and mentoring • Career guidance and counselling
Interventions to provide work experience	• Internships • Apprenticeship schemes • Vouchers and subsidies (demand-side) • Vouchers and subsidies (supply-side) • Job placements
Interventions to provide support to businesses and entrepreneurs	• Value chain services and market access • Formalisation/certification of businesses (including fair trade) • Matching grants • Innovation support • Business advice/mentoring • Micro-franchising
IV. Overcoming a lack of social capital and norms	
Social organisation	• Self-help groups • Collaboratives and collectives

Categories	Examples of relevant programmes
	• Organised labour
Changes in norms and attitudes	• Mentoring/role modelling • Gender empowerment (e.g. ambition/confidence building and autonomy) • Outreach and awareness raising, including 'policy advocacy' • Women's political participation • Mass media and public education campaigns • Youth empowerment
V. Overcoming behavioural (social and cognitive) barriers	
Gender-sensitive design	• Designing for women's time and mobility constraints • Designing for restrictive social norms (e.g. use of female implementers) • Designing for risk preference (e.g. risk aversion of women due to larger care responsibilities)
Behavioural nudges	• Commitment devices • Framing and identity cues • Micro-incentives • Reminders

Figure 1.1 outlines the conceptual relationship between the interventions and their final impact on women's economic empowerment. The first column lists the categories of interventions grouped by the barriers to female labour market participation that they aim to address. The second column indicates the intermediate changes that are assumed to result from the application of these interventions. For example, a business skills training programme might increase a female entrepreneur's managerial skills or increase the employability of a female graduate. However, these intermediate changes cannot be regarded as a proxy for women's labour market participation as they only enhance the probability of participation. Empirical measures of labour market participation are provided in the final outcome column. Broadly our review was concerned with two types of female labour market participation in higher-growth and/or male-dominated sectors: (i) wage employment in such sectors and (ii) the success of female entrepreneurs in such sectors,[2] for example growth of business and increased profits. More detailed outcome measures for these two final outcomes are provided in section 3.2.4, though neither type of labour market participation can be equated with female economic empowerment. Employment in a higher-growth sector such as construction, for instance, might not translate into meaningful improvements in women's economic empowerment if harassment at work is experienced and/or women have no control over the income they receive. As a last step in our conceptual framework, our review therefore investigated the extent to which increased labour market participation translated into meaningful and valued changes in women's economic empowerment.

[2] The outcome measure of entrepreneurial success was only relevant for the evidence map.

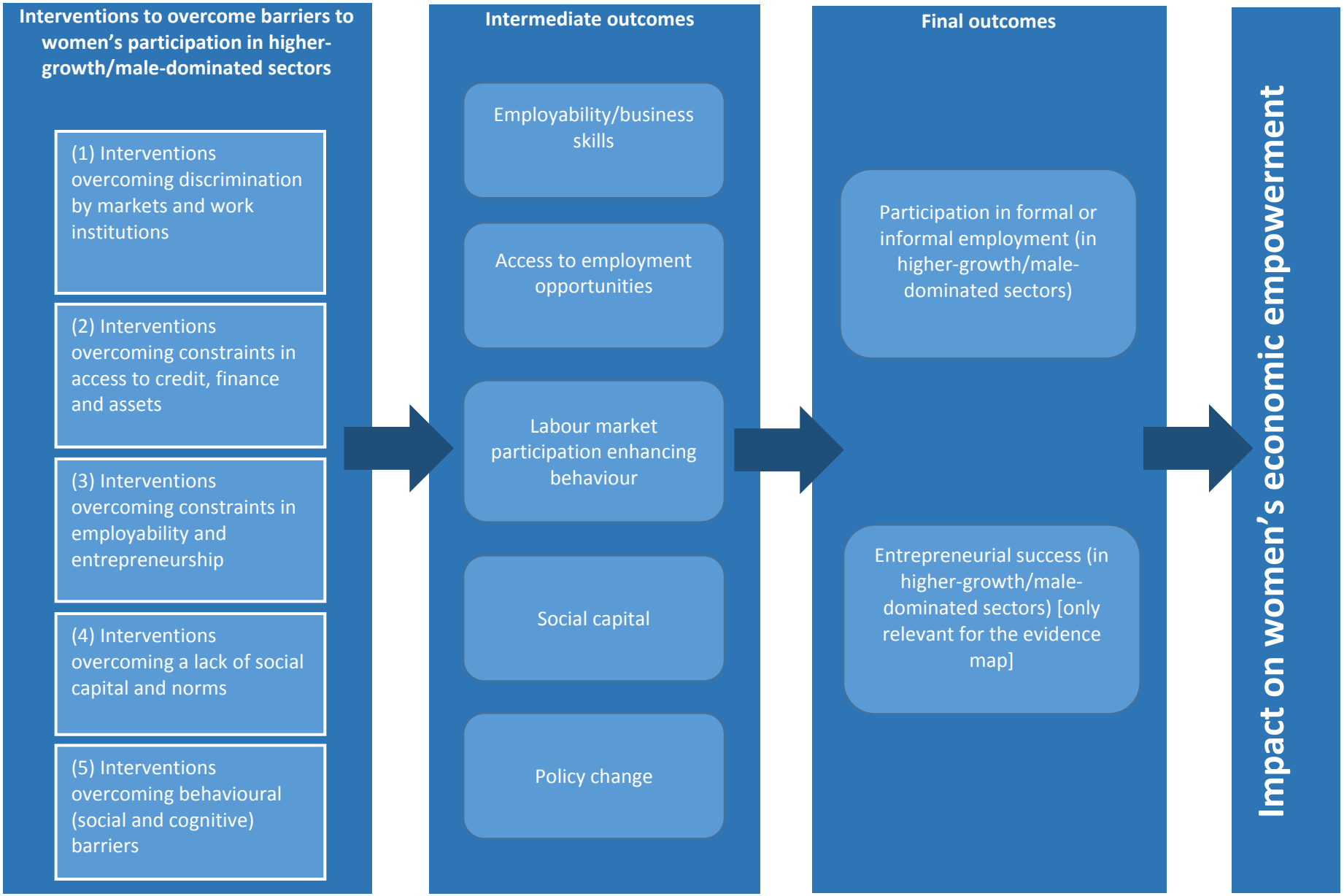


Figure 1.1: Conceptual framework

1.6 WHY THE REVIEW IS NEEDED

There are currently no systematic reviews that address the question of which interventions work to improve women's participation in the labour market in higher-growth and/or male-dominated sectors. While there is ample research evidence attesting to women's underrepresentation in such sectors, there is so far no rigorous synthesis as to which interventions can change the economic empowerment of women in such sectors. There are a number of systematic reviews on the effects of microfinance and cash transfers that also conduct sub-group analyses for impacts on women's empowerment (e.g. Gibbs et al 2012; Gopalaswamy et al 2016; Stewart et al 2012; Vaessen et al 2014; Yoong et al 2012). However, none of these reviews explicitly defines women's empowerment as participation in labour markets, let alone participation in higher-growth and/or male-dominated sectors.

The same applies to the systematic reviews on entrepreneurial training in LMICs (e.g. Cho and Honorati 2014; Kluve et al 2016; Tripney et al 2013; 2015) which are not exclusively focused on women or higher-growth sectors. Brody and colleagues' (2015) review of economic self-help group programmes to improve women's empowerment, as well as Gibbs and colleagues' (2012) synthesis of combined structural interventions for gender equality and livelihood security, each include relevant interventions, but do not focus exclusively on labour market participation in higher-growth and/or male-dominated sectors. Two recent reviews on the effects on business support (Piza et al 2016) and employment services and subsidised employment (Kluve et al 2016) are of high relevance to women's labour market participation too, but neither focuses on women exclusively or on sectors that are higher-growth and/or dominated by males.

At a policy level in international development, women's economic empowerment is a high priority. The International Labour Organization (ILO) has declared women's full participation in the labour market one of its centenary goals. Addressing women's economic empowerment is also directly mandated by SDG 5, and the UN Foundation published a major report on the state of research on women's economic empowerment in 2013 (UN 2013). This report on a roadmap to women's economic empowerment is based on a review and synthesis of 136 empirical evaluations of women's empowerment programmes and policies. The report was updated by the Center for Global Development (CGD) (Buvinic and O'Donnell 2016), which added 96 new evaluations that had been published since the launch of the 2013 report. There is thus a rich and growing body of research evidence evaluating the effect of policies and programmes on women's economic empowerment.

However, the focus of this work is not on women's economic empowerment through labour market participation per se. Rather, it investigates how women can be empowered to have greater control over their economic situation and ambitions. For example, interventions to increase women's decision-making power within the household, regardless of whether this increase is caused by participation in the labour market, are of high relevance in the UN's roadmap. This scope is slightly different from our review as it does not focus specifically on wage employment as a means to empowerment, let alone employment in higher-growth and/or male-dominated sectors. In addition, aside from the CGD and UN reports, little attention has been paid overall to the design features of women's economic empowerment interventions, which limits our understanding of why some interventions work (or do not work) and under what circumstances.

A systematic review of research evidence evaluating the effects and design features of interventions aiming to overcome barriers to women's labour market participation in higher-growth and/or male-dominated sectors is therefore a timely contribution to the ongoing policy and practice debates on what works to support women's economic empowerment. Our review rigorously synthesised a sub-set of economic empowerment interventions focused on wage labour market participation as a pathway to women's economic empowerment. The application of systematic review methodology, statistical meta-analysis, and qualitative comparative analysis (QCA) further enhanced the contribution of this review compared to existing synthesis work on women's economic empowerment. In conducting our systematic review, we drew on the above-mentioned existing reviews to ensure that we built upon, and did not duplicate, previous or ongoing efforts; this also applied to ongoing systematic reviews which overlap with some of the included interventions of our review, for example Chinen and peers (2016) and Ibanez and colleagues (2016).

2 OBJECTIVES OF THE REVIEW

Our systematic review aims to answer the following review question:

> What are the effectiveness and design features of interventions that aim to overcome the barriers to women's participation in the labour market in higher-growth and/or male-dominated sectors in low- and middle-income countries?

In doing so, we address the following review objectives:

1. To produce an interactive evidence map of research evaluating interventions aiming to overcome barriers to women's economic empowerment in LMICs.
2. To provide a rigorous synthesis of impact evaluation evidence to identify the effects of interventions supporting women's participation in wage labour in higher-growth and/or male-dominated sectors in LMICs.
3. To identify design features that influence the effects of interventions aiming to overcome barriers to women's economic empowerment in LMICs.

3 METHODS

3.1 TYPE OF REVIEW

We used a two-stage review approach consisting of a first-stage initial evidence map and a second-stage analytical review and synthesis.

3.1.1 EVIDENCE MAP

The evidence map is broader in scope than the full systematic review. It mapped evidence from impact evaluations and systematic reviews on interventions that aim to overcome barriers to women's labour market participation using an intervention-outcome matrix to highlight the size and nature of the evidence for different configurations of interventions and outcomes. The framework used for this matrix is available online.[3] The evidence map is visualised on an interactive online interface using software similar to the International Initiative for Impact Evaluation's (3ie) evidence gap maps.[4] Stakeholders can use the interface to create their own custom evidence maps by filtering the evidence base according to region, study design, and economic sector.

Although a product in its own right, the evidence map was primarily used to support stakeholder engagement with the evidence base and subsequent decision making on the most effective review approach and scope (see section 3.8.1). The bodies of evidence presented in the map aided the identification of the interventions and outcomes of most interest to review stakeholders. The evidence map also guided our initial discussions about the most effective way to synthesise the evidence in answer to the review questions. The evidence map was published online as an open-access output in February 2017.[5]

3.1.2 FULL SYSTEMATIC REVIEW

We conducted an effectiveness systematic review (Snilstveit 2012). We focused on: (i) the effectiveness of interventions supporting women's participation in the wage labour market in higher-growth and/or male-dominated sectors; and (ii) the design features of such interventions. The review therefore included studies that measured the effects of interventions and that reliably attributed observed effects to the applied interventions. Individual effects were first synthesised into an overall estimate of treatment effects, and then disaggregated according to the identified design features of interventions.

[3] https://africacentreforevidence.org/outputs-2/
[4] http://gapmaps.3ieimpact.org/evidence-maps/land-use-change-and-forestry
[5] https://africacentreforevidence.org/outputs-2/

3.2 CRITERIA FOR INCLUSION AND EXCLUSION OF STUDIES IN THE REVIEW

Studies had to meet the following selection criteria to be included in the systematic review. A summary of the inclusion criteria for both the evidence map and the full systematic review is provided in Appendix 2.

3.2.1 POPULATION

Eligible studies had to meet all of the following population criteria to be included in the systematic review.

Women: The study sample had to include women aged 15 years or older. The study sample had to either be majority[6] female or the study results had to be disaggregated by gender. Where the study sample included women below the age threshold, the majority of the study sample had to either meet our age threshold or the study results had to be disaggregated by age.

Geographical location: Low- or middle-income countries (as classified by the World Bank at the time of data collection for the study, see Appendix 3). Middle-income countries included both lower-middle and upper-middle income countries.

3.2.2 EMPLOYMENT SETTING

The study had to evaluate the effects of relevant interventions (as listed in Table 1.1) applied in economic sectors with high or growing productivity *and/or* which were male-dominated.

Productivity here is used as a proxy for a range of indicators, such as growth in revenue, profits, income, employment. Sectors that had already achieved high productivity in LMICs (e.g. transportation) were included, as well as sectors that were likely to experience productivity growth in LMICs (e.g. finance, ICT).

A list of sectors was presented in section 1.3 of this report. During the conduct of the systematic review, we then investigated the relevance of each economic sector individually in the context of the included studies.

We included economic sectors (see section 1.3) with either horizontal or vertical male-dominance as defined in section 1.2. Employment status or experience was not a criterion for exclusion or inclusion.

[6] Majority refers to a portion of 51% or higher of the sample.

3.2.3 INTERVENTION

We included *any* intervention that aimed to overcome the barriers to women's wage labour market participation in LMICs. It was therefore not possible to exhaustively pre-specify a list of relevant interventions. Section 1.5 outlined the main categories of interventions that we expected to encounter in this review and provided examples of interventions for each category. The term intervention in this context referred to a policy, programme, strategy, technology, device or any other type of deliberate action.

We included multi-component (also known as bundled or combined) interventions only if: (i) all intervention components aimed to overcome a relevant barrier to women's wage labour participation and/or (ii) studies could attribute the observed outcome to an intervention component that aimed to overcome a relevant barrier.

3.2.4 OUTCOMES

To be included in the systematic review, studies had to evaluate the impact of interventions on one of these two final outcomes:

(1) Participation in formal or informal wage labour employment (in higher-growth and/or male-dominated sectors)
(2) Economic empowerment (following outcome 1: participation in higher-growth and/or male-dominated sectors)

Relevant indicators for each final outcome are specified in Table 3.1.

Table 3.1: Examples of relevant indicators for each outcome

Outcome	Relevant indicators
(1) Participation in formal or informal employment	• Employment status[7] • Under-employment • Nature of the employment (e.g. security of contracts, working conditions, salary and wage levels) • Progression and career prospect (e.g. promotion) • Changing employment from traditional to non-traditional sectors for women's employment
(2) Economic empowerment	• Household income and any other poverty-related measure • Control over household income/spending • Access to economic assets • Investment in economic assets/productivity • Individual savings • Women's well-being • Women's empowerment

In cases where the majority of intervention participants were not female, studies were eligible for inclusion only if the impacts on women were assessed separately from those on men (i.e. in sub-group analysis) or in comparison to men.

INTERMEDIATE OUTCOMES

Studies were only included in the systematic review if they assessed one or both of the two final outcomes listed in Table 3.1. Studies investigating intermediate outcomes only were excluded. The following intermediate outcomes were eligible for inclusion if they were reported as part of a study assessing final outcomes that was included in the review:

- employability/business skills (e.g. technical knowledge, adoption of technology, change in business practices)
- access to employment opportunities (e.g. awareness about job openings, changes in employment policies)
- employment-enhancing behaviours (e.g. attending job interviews)
- social capital (e.g. networks, self-esteem)
- policy change (e.g. change in labour laws, anti-discrimination policies).

[7] Excluding self-employment.

We also recorded information on unintended outcomes, for example an increase in violence against females and effect capture, as well as information on intervention costs or cost-effectiveness where reported.

3.2.5 STUDY DESIGN

We included studies using either of the following quantitative experimental or quasi-experimental study designs:

(a) Designs using a random or quasi-random method of group assignment in which one of the following was true:

- Units (individuals or clusters of individuals) were randomly assigned to treatment and control groups by the investigator using a fully random procedure, such as computerised random number generation;[8]
- A quasi-random procedure presumed to produce comparable groups had been used by the investigators. For example, allocation by date of birth or next person to walk in the door (i.e. the method of allocation falls short of full randomisation);
- Regression discontinuity designs were used in which participants were assigned by the investigator to intervention or control groups solely on the basis of a cut-off score on a pre-programme measure.

(b) Designs employing non-random methods of assignment, in which one of the following was true:

- The investigator controlled group exposure and assigned participants using a non-random procedure (e.g. alphabetically by surname);
- The investigator constructed the comparison group after the start of the intervention (e.g. by exploiting existing survey data);
- A natural experiment in which units exposed to the treatment and control conditions were determined by nature (e.g. change in policy or divergence in practice between regions) or by other factors outside the control of the investigators);[9]
- Assignment to conditions (treatment versus comparison) was by means of self-selection by participants or by administrator selection (e.g. by welfare officials).

Studies employing non-random methods of assignment had to use appropriate methods that take account of selection bias and confounding at the design and/or analysis stage in order to be included in this review. In the context of this review, these included statistical matching (e.g. propensity score

[8] With or without matching.

[9] Such studies depend on the premise that the 'assignment' of subjects to the treatment and control groups is equivalent to random assignment (though not in a controlled way). Natural experiments are often associated with IV, RDD, and difference-in-differences.

matching), difference-in-differences estimation, interrupted time series analysis, regression discontinuity analysis, instrumental variables (IV) regression, and certain forms of multivariate regression analysis such as the Heckman sample selection (two-step) model. If a study used matching and/or covariate adjustment it might be done individually or by groupings (clusters) of individuals, and it might be based on participant characteristics observed before or after the start of the intervention. If there was no matching or statistical adjustments, then pre-treatment information on equivalence had to be available and groups shown to be comparable.

Finally, control or comparison conditions in eligible studies referred to the population receiving no treatment, treatment as usual, an alternative treatment, or pipeline treatment.

3.3 SEARCH STRATEGY FOR IDENTIFICATION OF RELEVANT STUDIES

3.3.1 METHOD OF IDENTIFYING RELEVANT STUDIES

A comprehensive search strategy was used to search the international research literature for qualifying studies. The aim was to identify all available evidence relevant to the review question. Our objective was to be sensitive rather than specific. We deliberately formulated search strings and identified search sources that were over-inclusive. While this increased the number of citations to be screened, it reduced the risk of missing relevant studies. We searched different types of sources, including sources with a particular focus on LMICs (see Appendix 3 for the current World Bank classification of LMICs, grouped by region). We used a wide range of sources to capture both academic and 'Grey' literature and reduced the omission of relevant studies to ensure that our search was as unbiased as possible.

3.3.2 DATA SOURCES AND SEARCH STRATEGIES

Our search strategy rested on three pillars: (1) a formal search of academic databases using explicit search strings based on Boolean operators; (2) a formal search of Grey literature using mainly key word searches, but applying full search strings where organisational databases allowed the application of Boolean operators; and (3) an informal search using different snowballing techniques.

ACADEMIC DATABASES

A wide range of social science bibliographic databases was searched covering literature from international development, economics, sociology, psychology, education, and health care. These were:

- Web of Science
- Econlit (EBSCO)
- ERIC (EBSCO)
- Business Source Complete (EBSCO)
- PRISMA database

- Sociological Abstracts (CSA).

The search of academic sources was led by an information scientist, who developed a detailed search query (or string) based on the inclusion criteria, relying on the database's index terms where available and/or free-text terms. Synonyms and wildcards were applied as appropriate. Database thesauri were consulted to ensure that all appropriate synonyms had been included. The search query was tailored for each bibliographic database, and these were peer reviewed and piloted. There were no language restrictions to the search. A publication year filter to identify studies published since 1990 was used. This cut-off date was chosen as structural inequalities in women's economic empowerment only started to receive increasing attention in the 1990s (Kabeer 2012). A master search query for the Web of Science database is presented in Appendix 4.

The general key concepts that we used for the search query are presented below. These were directly informed by our inclusion criteria and the interventions and outcome framework presented in section 1.5. We combined search terms related to the four key concepts using the following Boolean combination: 1 AND 2 AND 3 AND 4. We deliberately opted not to search for outcome terms, as women's economic empowerment and labour market participation are not well defined. Thus we did not apply terms related to an outcome concept in our search string to avoid missing relevant citations.

Key concepts

1. Developing countries
2. Women
3. Type of study
4. Intervention.

Search 1 and 2 and 3 and 4 in Title, Abstract, Keyword, Subject Heading.

GREY LITERATURE SEARCH

The Grey literature search was informed by our content experts and advisory group members and covered a large variety of organisational repositories and websites. This selection included umbrella organisations with general knowledge repositories relevant to the review question, such as the World Bank's knowledge hub and the R4D database of the UK Department for International Development (DFID), as well as organisations specifically focused on a sub-area of knowledge relevant to the review question such as the International Centre for Research on Women (ICRW) and the ILO. A full list of sources is presented in Appendix 4.

USE OF SNOWBALLING TECHNIQUES

The snowballing search included hand-searching the content pages of key journals of particular relevance to the review question; backward citation searches (i.e. searching the reference lists of

included studies and seminal papers); forward citation searches (i.e. using Google Scholar to search for papers that cited included studies); requests to key authors and organisations to share studies with the review team; and a Twitter call for the wider community of practice to contribute relevant studies. Appendix 4 lists the sources for the snowballing search.

3.3.3 MANAGING AND DOCUMENTING THE SEARCH AND SELECTION PROCESS

Review management software (EPPI-Reviewer 4) was used to manage the entire review process (Thomas et al 2010). All potentially relevant items identified through the academic database search were exported to EPPI-Reviewer and then manually screened for eligibility, with EPPI-Reviewer used to keep track of decisions made about each citation. Search hits from organisational repositories and snowballing were stored in MS Word, and only the details of studies deemed relevant for the map, plus those over which there was any doubt, were transferred to EPPI-Reviewer. In such cases, it was necessary to check whether the item was already in EPPI-Reviewer before proceeding to manually enter details. Upon screening against the selection criteria, a record of all decisions taken (include/exclude/unsure) was kept in EPPI-Reviewer and MS Word, as appropriate.

All information retrieval and selection activities in the review are documented and described in sufficient detail in this final report so that the processes can be replicated by other researchers. Summary flowcharts are used to convey information, where relevant. Based on the Cochrane PRISMA checklist for reporting results of searching and screening, the following information was recorded: databases, database platforms, search strategy for at least one database, dates of search and timeframe.

3.3.4 STUDY INCLUSION

Selection of primary studies was based on the pre-developed selection criteria described in section 3.2. These criteria were piloted by two researchers, who screened a sample of search hits independently and then compared and discussed their assessments. Discrepancies were resolved by further examination of the respective titles and abstracts. If a final decision could not be reached, a third reviewer was asked to reconcile differences. This process was repeated until consistency in application of the selection criteria was achieved.

Following pilot testing, the remaining literature was screened for eligibility by individual reviewers (i.e. single screening), who worked through the selection criteria hierarchically. The study selection process was undertaken in two phases – screening for the map and screening for the in-depth review – and commenced once all the hits from the academic databases had been exported into the EPPI-Reviewer 4 database and all potentially relevant items from the Grey literature and snowballing searches had been saved in MS Word.

PHASE ONE: TITLE AND ABSTRACT SCREENING FOR THE MAPPING EXERCISE

We screened studies for inclusion in the evidence map on title and abstract only. During this phase, we screened two sets of studies: those that had been imported to the reviewing software and those that had been saved in MS Word. To this end, we started by manually examining the titles and abstracts of all records entered into the reviewing software after removing duplicates. The relevance of each item was assessed and decisions regarding each study recorded in EPPI-Reviewer. Items were included in the map if they appeared to meet the criteria outlined in section 3.2 on the basis of the information in the title and abstract, and excluded if they were clearly ineligible. Where we were in any doubt as to a study's eligibility (e.g. because no abstract was available, or it did not provide enough information), we classified it as 'unsure'. We gave studies the benefit of the doubt if there was reference within the wording of the title to (women of working age) AND (a relevant outcome OR a term suggesting that the study was an evaluation).

For sources that did not allow us to export bibliographic information, we screened the items that had been saved in MS Word, transferring the details of studies deemed eligible for the map, and any of which we were unsure, to EPPI-Reviewer 4.

In cases where the title and/or abstract were not in English, the translation service offered by Google (http://translate.google.com/) was used to translate the information into English; screening against the selection criteria then proceeded as normal.

PHASE TWO: FULL-TEXT SCREENING TO IDENTIFY STUDIES FOR IN-DEPTH REVIEW

We screened studies for inclusion in the in-depth review using full texts. These were obtained for studies included in the map and for the items marked as 'unsure', and detailed manual examination of the full reports was undertaken independently by pairs of reviewers. Disagreements between the reviewers' decisions were resolved by identification of the source of the disagreement, re-reading of the text and discussion. If a final decision could not be reached, a third reviewer was asked to reconcile the differences.

In the event that we could not determine the eligibility of a study (e.g. because the full text was unavailable, insufficient information was provided, or the only version we had was in a language other than English), the study authors were contacted to request additional information about the study or access to English-language versions. In cases where no English-language version or additional information could be obtained, the study was excluded from the review.

3.3.5 CRITERIA FOR DETERMINATION OF INDEPENDENT FINDINGS

Sometimes one evaluation leads to several study reports (e.g. working papers and journal articles). Efforts were made to identify all affiliations between study reports before coding commenced, using

information on study sample sizes, intervention details, grant numbers and so on. In cases where multiple reports were found to relate to a single study, the reviewers chose one as the main report (e.g. the publication containing the most complete data set). When extracting data, the full set of available reports was used to code each study.

Where a single report described more than one study (e.g. a single publication could describe a series of evaluations conducted in different countries, using different data sets), each study was coded individually as if they had come from separate reports.

Sometimes an intervention had been evaluated several times. Where we found that several evaluations were based on the same data, these were treated as part of the same study, even if the reports were written by different authors. If the intervention had been evaluated on multiple occasions using different data sets, then we treated the different reports as separate studies whilst noting their relationship. Where it was unclear whether multiple reports provide independent findings, the authors of primary studies were contacted for clarification.

We used meta-analysis to synthesise the results across primary studies. In a single meta-analysis, it is important to include only one effect size measure per study. Estimated treatment effects cannot be regarded as independent of each other if the underlying data are derived from the same sample populations (i.e. some participants contribute information to more than one effect estimate).

Individual effect size estimates may be correlated if, for example, the study analyses:

- different sub-groups of the treatment population (e.g. young women, highly educated women);
- outcomes at different times but on the same units (e.g. midline and endline findings);
- multiple outcome constructs for the same type of outcome (e.g. effects on hourly wages and earnings);
- multiple interventions with the same sample of participants;
- multiple treatment groups and the same control group;
- effects using different methods;
- several types of treatment effect estimates, e.g. average treatment effect (ATE), intention to treat analysis (ITT);
- multiple time-points for the same individual (e.g. repeated observations for several follow-up periods); or
- variations of the above.

In such cases, one of the following approaches was pursued, as appropriate. Where studies reported multiple effect sizes by sub-group, we reported the data in separate analyses. In the event that we had more than one effect size per outcome construct and study, we combined different estimates within each study into one effect size per sub-group. Estimating a single (within-study) composite effect size involved computing a sample-weighted average effect size for each study that accounted for differences in sample size, using appropriate formulae (Borenstein et al 2009). If a study included in the review had

used several different techniques to estimate treatment effects for the same outcome (e.g. both statistical matching and regression analysis), the estimate with the lowest risk of bias was used in the meta-analysis.[10] In the event that risk of bias assessments were similar, we chose between different estimation approaches using the approach outlined by Tripney and colleagues (2013) (e.g. comparing the efficiency of the estimator in studies that used both matching and covariate adjustment).

3.4 DATA EXTRACTION AND CRITICAL APPRAISAL

3.4.1 DATA EXTRACTION

We used a predefined data extraction tool in order to systematically and transparently extract data from the included primary studies. The tool, which is presented in Appendix 5, was translated into a coding set on EPPI-Reviewer to extract the information required for both the mapping exercise and the in-depth review. The data were entered directly into the EPPI-Reviewer database.

For the in-depth analysis and synthesis of study results, full-text reports were examined and studies coded on variables related to the study context, the characteristics of the study samples, details of the intervention design and its implementation, the study methods and the outcome variables and data.

Two members of the review team piloted the data extraction tool, working independently on a purposive sample of eligible studies selected to test the tool on the full range of included impact evaluation designs and methods. This process was repeated until a very high level of consistency in reviewers' application of the codes was achieved, at which point the tool was finalised. Thereafter, the remaining studies were coded by individual reviewers, with a sub-set of full-texts being coded by different combinations of two reviewers independently extracting information from each study report and then coming together to compare their decisions. Any uncertainties and discrepancies were resolved by discussion, further review of the respective study reports, and where necessary, consultation with a third reviewer.

3.4.2 CRITICAL APPRAISAL

A critical appraisal tool was applied to assess the impact of bias on the trustworthiness of primary study results, where trustworthiness refers to the confidence of the review team that the findings reported in the included studies used for the synthesis were rigorous and credible. In order to assess the risk of bias of the primary studies, we adapted the Cochrane risk of bias tool for non-randomised studies (Sterne et al 2016), which we had previously used and adapted in international development reviews (Stewart et al 2015). Sterne and colleagues used a domain-based risk of bias assessment covering the following six indications of trustworthiness: (i) selection bias; (ii) confounding bias; (iii) bias due to departures from

[10] A study may have used the same set of data but different estimation methods and published the results in a single report or in separate multiple reports (e.g. a different report on each of the estimation methods used).

applied interventions; (iv) bias due to missing data; (v) bias due to measurement of outcomes; and (vi) bias due to selection of the reported result. Each domain of bias received a low, moderate, high or critical risk of bias rating, allowing for a transparent calculation of an overall risk of bias score for each study. Studies with a critical risk of bias were included in the review but excluded from the synthesis.

The critical appraisal tool used to assess studies for the full review is presented in Appendix 6. It was piloted using a similar approach to that used for the data extraction tool. Two reviewers independently assessed each study and then came together to compare their decisions. Where these reviewers could not come to an agreement about the risk of bias rating for a particular study, a third reviewer was consulted.

3.5 EVIDENCE SYNTHESIS

This systematic review included an aggregative synthesis in order to: (i) distinguish between effective and non-effective interventions; and (ii) identify those configurations of intervention design features, participants and contextual characteristics that might be associated with a given outcome. The analysis focused particularly on identifying design features of interventions that were critical to their success. We used meta-analysis to identify overall effects and narrative synthesis and qualitative comparative analysis (QCA) for the disaggregated analysis according to the intervention design features (Gough et al 2012; Thomas et al 2014).

3.5.1 STANDARDISED EFFECT SIZES

In order to standardise outcomes, the calculation of effect sizes for each intervention outcome is required for both of the quantitative synthesis methods mentioned above. Extracted statistical information for effect size calculation was exported into MS Excel and reviewers documented the different computations and formulae used for the effect size estimates derived from each study. Other web-based resources (e.g. the Campbell Collaboration's effect size calculator) were utilised for the less common statistical representations.[11]

Standardised mean difference (SMD) or Response Ratio (RR) effect sizes present the most relevant metrics in the context of this review. SMD and RR were calculated for continuous outcome variables, while only RR were calculated for dichotomous outcome variables. By using SMD and RR calculated treatment effects, we indicated the ratio of, or difference between, treatment and control groups. Therefore, an increase in the outcome attributed to the intervention as compared to the control – an 'effective intervention' – was indicated by a SMD greater than 0 or an RR greater than 1. An SMD less than 0 or an RR between 0 and 1 indicated a reduction of outcomes for the population exposed to the intervention as compared to the control group – i.e. a harmful intervention. An SMD equal to, or not significantly different from 0, and an RR equal to 1 indicated that there were no observable changes in

[11] https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-Home.php

outcomes between the treatment and control groups. All effect sizes were coded such that positive effect sizes represented positive outcomes (e.g. the intervention lowered duration of unemployment, led to higher wages).

We corrected for potential sample bias in the effect sizes (due to small sample sizes) by converting them into Hedges *g* using the correction procedure developed by Hedges and Olkin (1985). The detailed approach to effect size calculations is provided in Appendix 7 including the conversion formula for effect sizes into percentage changes in outcomes.

3.5.2 UNIT OF ANALYSIS ISSUES

We identified and addressed any unit of analysis errors, which typically arise if the study conducts programme assignment and analysis at different levels and the analysis does not adequately account for this. This can be a problem in both randomised controlled trials and quasi-experimental studies where treatment allocation is clustered. Both the unit of assignment to treatment and comparison groups and the unit of analysis was coded for all studies. To correct for variation associated with cluster-level assignment, adjustments were made using the correction formula suggested in Hedges (2007). If sufficient information to apply this formula was not provided by the primary authors, we followed the approach described in Higgins and Green (2011).

3.5.3 DEALING WITH MISSING AND INCOMPLETE DATA

We extracted data on all variables of interest listed in the data extraction tool (see Appendix 5). Where studies lacked data considered essential for the review, we made thorough attempts to contact the original investigators. If the necessary data to compute effect sizes could not be retrieved from the authors, we considered imputing the missing data with replacement values. The proportions of missing participants were coded and included in the risk of bias assessment.

3.5.4 SYNTHESIS OF OVERALL EFFECTS

We used a statistical approach to synthesising results across the primary studies. Data extracted from the studies were pooled using meta-analysis. However, pooling of extracted data depended on such factors as the heterogeneity of the studies and study populations. Key features of the participants, interventions and outcomes were described in summary tables, along with effect size estimates and methodological quality characteristics. The analysis was conducted using the specialised built-in meta-analysis function within EPPI-Reviewer 4.

3.5.5 METHODS OF SYNTHESIS (1): META-ANALYSIS

Meta-analysis is the most rigorous method to synthesise quantitative evidence (Borenstein et al 2009; Lipsey and Wilson 2001). As a statistical approach it aggregates the numerical findings, such as effect

sizes, of primary research to report a pooled overall numerical value. This numerical value – the pooled effect size – expresses the overall finding derived from the combined primary research results. The pooled effect size reflects the direction and magnitude of the observed primary effect sizes, which are allocated different weights in the analysis depending on sample sizes and variance.

We report calculated effect sizes in tabular format as well as using forest plots. Where sufficient contextual homogeneity prevailed, effect sizes were averaged across studies by using an inverse variance weighting of the individual effect size. This weighting resulted in the individual effect sizes of studies with larger study samples being given more weight in the combined, pooled effect size. The meta-analyses were carried out using random effects statistical models.

The studies included in the review featured impact evaluation designs based on a random (or quasi-random) method of group assignment, and those that used non-randomised procedures. Some caution is needed when synthesising effect sizes from studies using different designs and estimation techniques. Where appropriate, we summarised across designs and made explicit our rationale for doing so. In addition, we calculated separate estimates of treatment effects for randomised and non-randomised studies.

ASSESSMENT OF HETEROGENEITY

In order to visibly examine variability in the effect size estimates, we used forest plots to display the estimated effect sizes from each study along with their 95% confidence intervals. Subsequently, and acknowledging the limitations of a quantification of heterogeneity and the different strengths of statistical approaches, the following test for heterogeneity were conducted: calculation of the Q statistic as a statistical test of heterogeneity (Hedges and Olkin 1985); and calculation of the i^2 and Tau^2 statistic to provide estimates of the magnitude of the variability across study findings caused by heterogeneity (Higgins and Thompson 2002; Higgins et al 2003).

SENSITIVITY ANALYSIS

To test the robustness of the results of the meta-analysis, a number of sensitivity analyses were conducted. Broadly, this involved collecting data on, and assessing sensitivity of findings to (i) the methods of the primary studies and (ii) the methods of the review.

The included studies varied methodologically. We conducted sensitivity analyses to examine the influence of these variations on the summary measures, in order to offer possible explanations for the differences between studies when interpreting the results. We examined whether the results were sensitive to study design, the risk of bias associated with the study, the degree of missing/incomplete data, and the way outcomes were measured and the timing at which they were measured.

As explained above, the main objective of the sensitivity analysis was to serve as a visual tool that allowed informal comparisons to determine whether the results of our meta-analyses were sensitive to

the methodological decisions of our review team. However, due to the controversy of pooling studies of random and non-random design, as well as of different risks of bias, we followed up the sensitivity analyses of these two variables with a one-way random effects ANOVA model calculated in EPPI-reviewer. That is, the mean effect size and standard error for each group of studies was calculated to test whether these means were statistically significant from one another (Lipsey and Wilson 2001).

MODERATOR ANALYSIS

We conducted moderator analyses to try to explain variations in effect sizes. Moderator analyses were reported in tabular format below each meta-analysis. Analyses were calculated using the same one-way random effects ANOVA model as applied to the sensitivity analyses.

Overall, we conducted moderator analyses to assess the following variables:

(1) socio-economic contexts;
(2) intervention characteristics;
(3) economic sector; and
(4) population characteristics.

Socio-economic contexts

Variables related to the socio-economic contexts in which the interventions are implemented are likely to moderate the effects of labour market interventions. For example, interventions in more affluent contexts might stand a better chance of yielding positive results. We therefore used the geographical region, including the country classification employed in the United Nation's *Roadmap for Promoting Women's Economic Empowerment* (UN 2013) and the World Bank classification of economies as relevant moderator variables.

UN classification of countries' socio-economic contexts:

- High fertility agrarian societies
- Declining fertility urbanising societies
- Declining fertility formalising economies
- Ageing societies.

World Bank classification of countries' socio-economic contexts:

- Low-income country
- Lower-middle-income country
- Upper-middle-income country.

Intervention characteristics

Variables related to the design of the labour market interventions are also likely to affect the outcome of the interventions. While we conducted a detailed QCA to unpack the relationship between different design features on intervention success, we investigated a small number of variables related to the intervention characteristics in the moderator analysis too. These referred to:

- the inclusion of a soft skills component in the intervention design;
- whether interventions were accessible to each gender or limited to women participants;
- whether interventions were provided by public or private providers.

These intervention design variables were assessed given their importance as flagged in related reviews on the impact of labour market interventions on female employment (Honorati and McArdle 2013; Kluve et al 2016). Lastly, we also used the dominant programmes evaluated in the studies included in our review as potential moderators. This allowed us to comment on the observed comparison between different programme designs in practice.

Economic sector

We also conducted moderator analyses using the economic sectors which were targeted for an increased participation of women in the included studies. These were:

- ICTs and electronics
- Manufacturing
- Construction
- Business administration
- Finance
- Services not otherwise considered.

Population characteristics

Lastly, we also attempted to analyse several population sub-groups of interest, including:

- the population categories used in the United Nation's *Roadmap for Promoting Women's Economic Empowerment* (UN 2013)
- the PROGRESS-plus categories.

The PROGRESS-plus framework was used in order to bring an equity lens to the review and synthesis process. Due to limited reporting, we were, however, only able to use the age of the participants as a moderating variable related to characteristics of the population.

3.5.6 METHODS OF SYNTHESIS (2): QUALITATIVE COMPARATIVE ANALYSIS

QCA investigates the configuration of different conditions and their association with intervention effectiveness (Thomas et al 2014). A condition refers to different design features in the context of our review, for example the use of behavioural techniques, gender sensitive design or the intensity of the intervention. Using QCA enables us to analyse whether configurations of different design features are present (or not) when the intervention has been effective (or not) in increasing women's wage labour participation. In comparing different types of configurations, QCA applies Boolean logic to establish necessary or sufficient conditions for intervention effectiveness. We refer to these conditions as the active ingredients of women's wage labour participation interventions in our review.

The advantage of this synthesis approach is the presentation and investigation of overlapping pathways to causality, which is to be expected in our review, as it is unlikely that any single intervention or design feature is the single cause of positive labour market participation effects. While meta-analysis is restricted to evaluating the effects of different variables individually, QCA allows us to evaluate the effects of different configurations of conditions. For example, conducting a moderator analysis as part of our meta-analysis, we could examine whether intervention intensity had a significant impact on intervention effectiveness *or* whether using behavioural approaches had a significant impact on intervention effectiveness, and so forth. In contrast, QCA allows us to compare different configurations of these design features (e.g. using a behavioural approach *and* a gender-sensitive design) and how these configurations – not the conditions themselves – are associated with intervention effectiveness. We assumed that providing different configurations of intervention design features could provide more policy-relevant synthesis results that could inform future policy and programme designs. This builds on a similar approach of using QCA to further interrogate the results of a meta-analysis developed by Brunton and colleagues (2014) in a systematic review on community engagement.

Our approach to QCA aimed to follow the six-step method for QCA proposed by Thomas and colleagues (2014) based on Rihoux and Ragin (2009). The authors propose the following six steps, but suggest that step five – consideration of the 'logical remainders' cases – might not be necessary when using QCA in the context of systematic reviews:

(1) building the data table
(2) constructing a 'truth table'
(3) resolving contradictory configurations
(4) Boolean minimisation
(5) consideration of the 'logical remainders' cases
(6) interpretation.

We used the studies included in the systematic review as our sole source to collect relevant conditions to construct the initial data table. In this process of identifying relevant conditions, we first consulted the studies reporting the most and least effective interventions and extracted their design features. This process of data extraction for the QCA used a separate extraction tool tailored to the identification of

intervention design features (Appendix 8) and the same consistency checks as the general data extraction process outlined above. The data extraction tool in Appendix 8 was used to identify a first list of conditions for constructing the data table. This list of conditions or design features was then shared with a range of stakeholders for feedback and further refined following a consultative workshop among the review team.

The data table then mapped the relevant conditions and outcomes for each individual study. We conducted a fuzzy set QCA in which interventions can be partially attributed to conditions. Following analysis of the data table, we developed the truth table, which displayed the conditions and configurations, and the number of studies with membership in each configuration set (Thomas et al 2014). The truth table itself then underwent careful analysis to investigate the logical coherence of the identified configurations. The results of the QCA were intended as an explanatory supplement to the meta-analysis findings. The QCA was designed to address objective (3) of our systematic review in case we were not able to conduct meaningful moderator analyses due to prevailing heterogeneity in the sample of included studies.

3.5.7 ASSESSMENT OF PUBLICATION BIAS

We aimed to assess publication bias using funnel plots followed by the 'trim and fill' method developed by Duval and Tweedie (2000). However, we identified an insufficient number of primary studies per intervention category (on average three studies per category) in order to conduct a meaningful assessment of publication bias.[12]

3.6 OVERALL ASSESSMENT OF THE QUALITY OF THE EVIDENCE

We applied the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) tool[13] to assess the overall quality of the evidence included in the review and the strengths of the review's findings. GRADE evaluates the quality of the primary evidence included in the review based on a range of factors, including: primary study limitations, inconsistency of the identified effect sizes, indirectness of the evidence (e.g. not evaluating final outcomes), and risk of bias ratings (Guyatt et al 2008). The quality of evidence rankings ranged from high to moderate, low and very low.

In addition to assessing the quality of the primary evidence, GRADE further assigns a strength of recommendation ranking to review findings. In this, GRADE combines the quality of evidence scores with additional variables such as uncertainty about the balance between desirable and undesirable effects, or uncertainty or variability in values and preferences. Strength of recommendation rankings are divided into 'strong' and 'weak'. However, recommendation rankings are more applicable for systematic reviews that aim to inform guideline development (Guyatt et al 2011) – an objective not shared by this

[12] The funnel plot for the only meta-analysis conducted, on interventions combining training with placement services, is provided in Appendix 14 for transparency.
[13] http://www.gradeworkinggroup.org/

review – and we consequently did not report strength of recommendation rankings in the discussion of findings in section 6.2. As a result, we did construct a summary table for each included intervention illustrating the review's findings regarding intervention effectiveness and GRADE's rating of the quality of evidence for each finding, but we did not provide a strength of the recommendation ranking (Appendix 9) (Guyatt et al 2011).

3.7 TREATMENT OF QUALITATIVE DATA

We did not include qualitative research studies in our review. Where qualitative research was reported as part of included impact evaluations, we extracted information about intervention implementation and context. This information was used to guide our grouping and analysis of the included studies.

3.8 STAKEHOLDER ENGAGEMENT IN DESIGNING AND PRODUCING THE REVIEW

In this review, we applied a range of mechanisms to support stakeholder engagement in designing and producing the review. These mechanisms included the creation of a review advisory group, regular joint meetings between the review team, the EPPI-Centre as the co-ordinating body, and DFID, and the production of an interactive evidence map. Each of these mechanisms aimed to ensure that the review's scope was of relevance to decision makers and that it addressed a relevant policy question. Two main features of this approach stand out for its implication for the design and production of the review: (i) using the interactive evidence map to jointly decide on the review scope; and (ii) identifying intervention design features associated with women's increased labour market participation.

3.8.1 INTERACTIVE EVIDENCE MAP

As indicated in section 3.1.1, this systematic review followed a two-stage approach commencing with a broad map of the available literature before deciding on relevant areas within the map to target the scope and synthesis of the review. The initial scope of the call for reviews was concerned with the broader area of women's economic empowerment. In order to both provide insights into this broad research area as well as to zoom into particular sub-sets of this literature for the systematic review, we conducted a full systematic search for all available impact evaluations on what works to support women's economic empowerment in LMICs. This search identified 501 impact evaluations, which we then mapped against an intervention-outcome framework to develop a matrix, as illustrated in Figure 3.1.

Figure 3.1 Example of the interactive evidence map

Mapping these 501 impact evaluations on what works to increase women's economic empowerment in LMICs allowed us to provide a visual guide to the patterns in the evidence base identifying evidence gaps as well as areas where different types of systematic reviews might be of value. The evidence map is further presented using an interactive interface built in HTML which allows users to tailor the map according to their preferences and to directly access the abstracts and full texts (where available) of the studies included in the evidence map. More information on this approach to evidence mapping and how its policy relevance can be enhanced can be found in the Departmental Guidance Note on policy-relevant evidence maps of South Africa's Department of Planning, Monitoring and Evaluation (DPME 2016).In this review, we presented the interactive evidence map to DFID and the EPPI-Centre in a joint meeting in February 2016. Based on the patterns in the evidence base, we then decided on the final scope of the systematic review, which narrowed the outcomes to wage labour participation only. The evidence map, however, remains a living document and has been updated throughout the review process with new impact evaluations published. The latest version of the map can be found on open access at: https://africacentreforevidence.org/outputs-2/.

3.8.2 IDENTIFYING DESIGN FEATURES

The second key contribution of these consultations to the review design and production was the identification of relevant design features for the QCA. A main emphasis of the call for research was identifying design features of interventions that were critical to intervention success. Such design

features, for example, can guide an understanding not just of what interventions work, but also on how and why these interventions work. However, there is a vast range of possible design features to investigate and we required an approach to ensure that the features tested in this review represented programme mechanisms and implementation aspects that were of policy interest. In this, we applied a range overlapping consultative processes. First, we developed an initial list of design features based on a review of empirical literature on labour market interventions for women in LMICs. This list was published in the protocol and is reproduced below:

- Gender-sensitive programme design and service provision:
    - Taking into consideration travel constraints (e.g. distance to homes)
    - Taking into consideration time constraints (e.g. when the programme is offered)
    - Taking into consideration caregiving constraints
    - Taking into consideration the gender of programme implementers
- Changes to social/professional norms (e.g. changing norms around women's appearance in public as an intervention mechanism)
- Dosage of the intervention (e.g. length of business training, size of loan)
- Quota/reservation approaches to ensure women's participation
- Demand-driven approaches (i.e. demands by markets)
- Financial input from employer
- Interventions working through subjective economic empowerment:
    - Increased economic self-reliance
    - Increased self-confidence
    - Increased risk taking
- Interventions addressing cognitive and social determinants of economic behaviour:
    - Protecting women from family and community pressures
    - Decision-making support
    - Increasing women's choice and autonomy
- Interventions addressing pervasive gender biases embedded in organisations and working environments
- Interventions applying explicit behavioural designs:
    - Commitment devices
    - Framing and identity cues
    - Designing for the risk preferences of the population (e.g. women being more risk averse)
    - Micro-incentives (e.g. performance-based rewards, social tokens)
    - Reminders

We then consulted widely on this list including our advisory group, DFID, the EPPI-Centre and labour market experts in South Africa to refine the identified design features. In addition, we also conducted a structured narrative synthesis (see Appendix 13) of the design aspects of all studies included in the systematic review. Thereafter, we then brought the results of the consultation, as well as the narrative

synthesis together in a one-day in-depth workshop for the extended review team in Johannesburg. This led to the formulation of an updated list of design features, which was then again shared with all review stakeholders. In total, our list of design features went through seven iterations, each of which was commented on and refined by different review stakeholders.

4 DESCRIPTIVE RESULTS

This section provides the descriptive results of the systematic review. It outlines how the included studies were identified and highlights their key characteristics. These characteristics refer to: study setting, population, intervention, outcomes, study design and risk of bias. The section aims to introduce the reader to the evidence base on which the review's findings are based. It thus contextualises the synthesis findings reported in the next sections. To recall, the focus of this systematic review is on providing a rigorous synthesis of impact evaluation evidence on the effects of interventions supporting women's participation in wage labour in higher-growth and/or male-dominated sectors in LMICs. The evidence described in this section therefore only refers to the characteristics of the identified evidence included in the systematic review.

4.1 RESULTS OF THE SYSTEMATIC SEARCH

The systematic search for impact evaluations on the effects of interventions supporting women's labour market participation in LMICs identified 16,091 citations (Figure 4.1). These were derived from 74 different sources compromising scientific databases of research studies as well as a range of Grey literature sources. Our exhaustive search for studies indexed in scientific databases was conducted between December 2016 and January 2017 and updated in April 2017. Our search of Grey literature sources was conducted in February 2017. All search hits were imported into EPPI-Reviewer (V.4.6.4.1) to facilitate data storage and management, except for the Grey literature searches; because they had to be entered manually, only those that were found to be relevant after initial screening (or of which the reviewers were doubtful) were entered into EPPI-Reviewer.

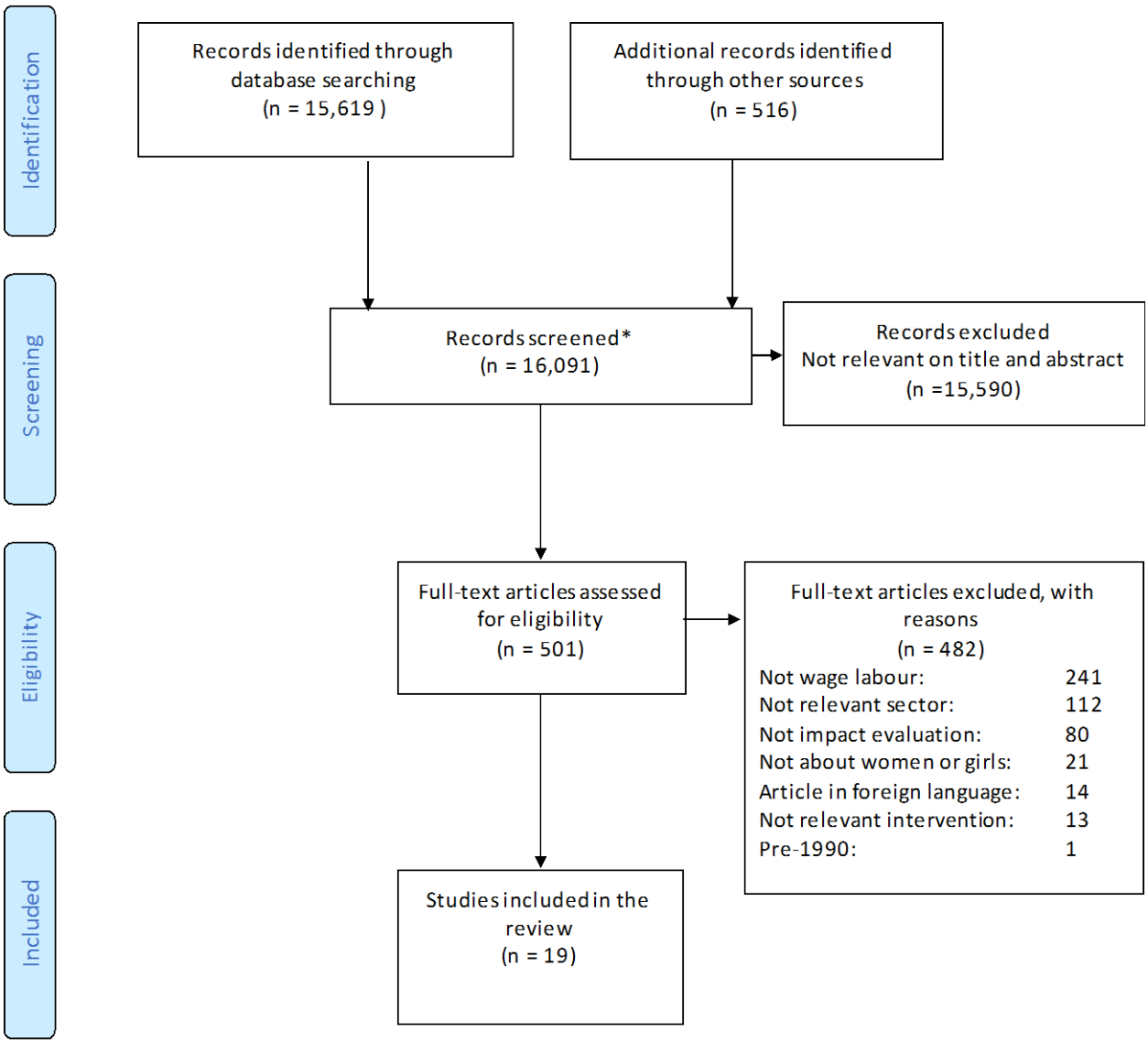
After screening on title and abstract, the large majority of citations were excluded as not relevant to the review question (n=15,590). We then retrieved the full texts of the remaining 501 studies and screened these against our inclusion criteria. This led to the exclusion of a further 482 studies. The main reason for exclusion on full text was that studies did not evaluate the effects of interventions on wage labour outcomes (n=241), for example examining small-business development outcomes. Of these, 40 studies only reported the number of employees by which the company had grown, which was regarded as an indicator of business development rather than wage labour.

We also excluded a large number of studies relating to irrelevant economic sectors (n=112). This referred to studies not examining wage labour market outcomes in higher-growth and/or male-dominated sectors, for example, nursing. However, in the majority of these studies, it was an absence of reporting on the economic sector that led to their exclusion. A total of 38 studies did not report economic sectors at all and a further 65 did not disaggregate labour market effects across sectors. Nine studies reported on sectors that could be identified as low-growth or female-dominated.

We excluded 80 studies that did not apply rigorous impact evaluation designs as outlined above. We further excluded 21 studies in which the evaluation results were not disaggregated by gender and 14

studies published in Spanish. Thirteen evaluations of minimum wage policies were also excluded, as was a single study in which the data was collected before our cut-off date of 1990. As a result, our final sample comprised 19 included studies.

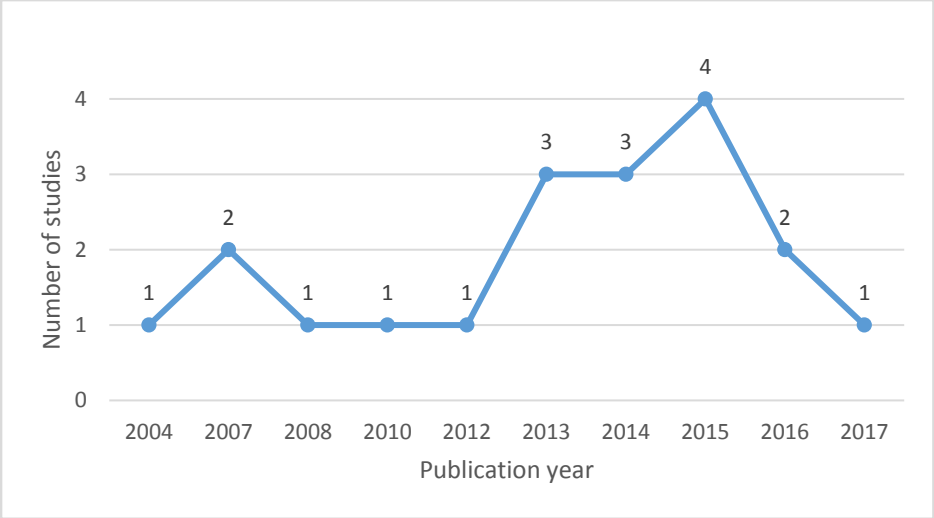
Figure 4.1 Flow chart diagram of search results and identification of studies



4.2 TYPES OF INCLUDED STUDIES

We included a total of 19 impact evaluations in our systematic review, which were reported in 23 papers (Appendix 10). Throughout the report, we use the term 'study' to refer to an independent data set regardless of how many papers reported on it.[14] Of the 19 studies, the majority were published as working papers (n=13) followed by journal articles (n=5) and a single evaluation report. The studies included in this review were published within the last two decades, with the earliest being Elías (2004) and most recent being Chen (2017). The majority of studies were published within three years between 2013 and 2015. Figure 4.2 below provides a cumulative number of the date of publication of studies.

Figure 4.2 Cumulative number of studies per year



The characteristics of all the 19 included studies are presented in Appendix 11.

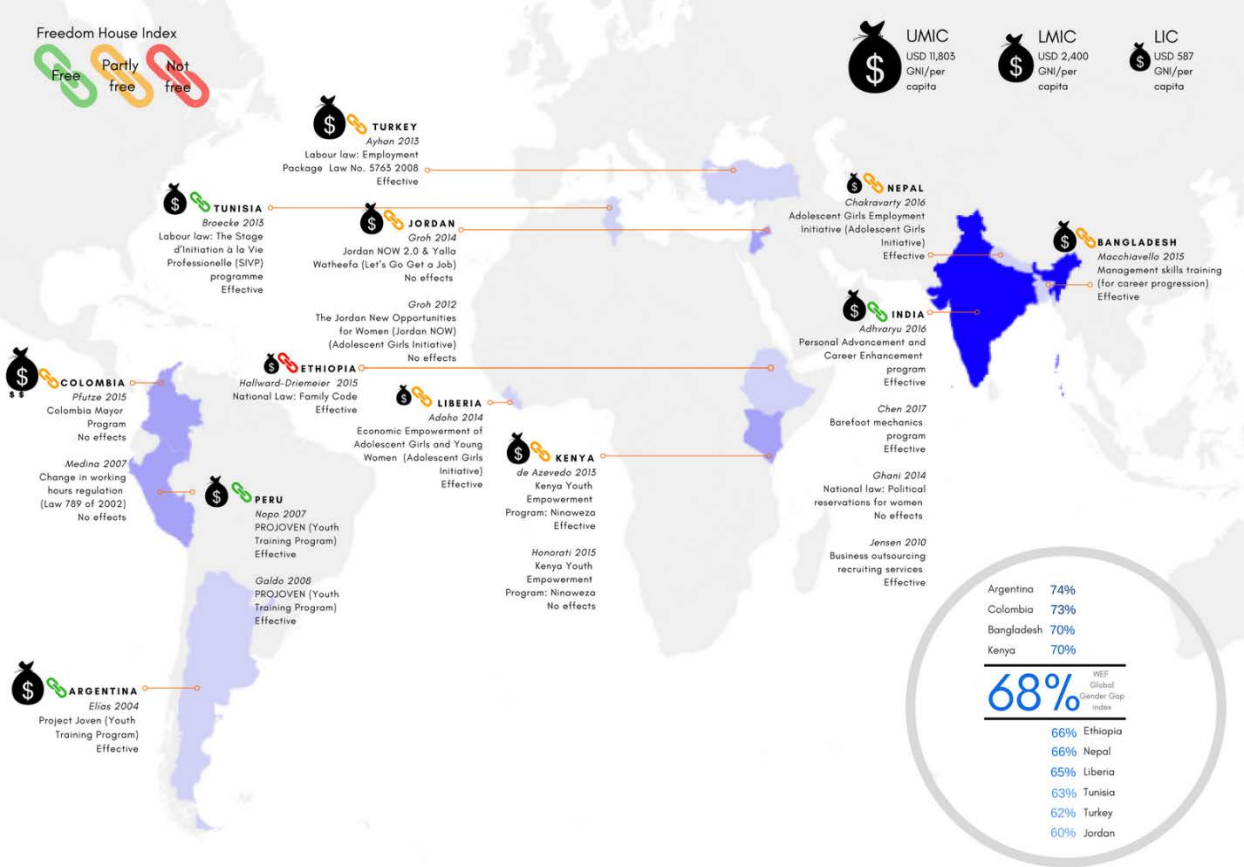
4.3 STUDY SETTING

There was large heterogeneity in the settings of the included studies. They varied in a range of socio-economic variables and represented a rich collection of contexts. First, in terms of geographical location, the studies were fairly evenly spread throughout continental regions. Nine studies were conducted in Asia, while five studies each were conducted in Latin America and Africa. In terms of individual country spread, India (n=4) was the most researched country followed by Colombia, Jordan, Kenya and Peru (all n=2). A single study was conducted in each of the following countries: Argentina, Bangladesh, Ethiopia, Liberia, Nepal, Tunisia and Turkey. Figure 4.3 presents a visual overview of the spread of countries in which the studies were conducted. Notable absences of countries with strong government labour

[14] Note that throughout this section and in related tables, studies are referred to by their first author and date in order to conserve space.

market interventions include: Brazil, China, Russia and South Africa. The cluster of research in India was linked to the country's manufacturing and business outsourcing economic sectors.

Figure 4.3 Overview of geographical regions of studies



In terms of socio-economic indicators, we investigated a number of dimensions. According to the World Bank classification of economies,[15] a majority of our included studies were classified as lower-middle-income countries (n=13), with an average GNI/per capita of USD 2,400. Only three countries were respectively classified as LICs (average GNI/per capita: USD 587) and UMICs (average GNI/per capita: 11,803). Two of the LICs were in Africa (Liberia and Ethiopia) and one was in Asia (Nepal). The three UMICs were Turkey, Argentina and Columbia. We also contrasted these GDP-driven classifications with UN country categories used in the UN's seminal roadmap to promoting women's economic empowerment report[16] (UN 2013). Here we found that 14 studies fell under 'Declining fertility, formalising economies'. Eight of these studies were from Asia, while five were from the Latin America and Caribbean region. Ethiopia was the only country from Africa classified under this category. Three studies were categorised as 'High fertility, agrarian societies', all of which were from Africa (Kenya n=2

[15] https://blogs.worldbank.org/opendata/new-country-classifications-2016
[16] http://www.womeneconroadmap.org/

and Liberia n=1). The last category was 'Ageing societies' (n=2), represented by one study each in Turkey and Tunisia.

Regarding urban and rural socio-economic categorisation, we found no clear trend, with the majority of studies spanning both rural and urban areas (n=10). Seven studies were conducted in urban areas tipping the evidence base slightly in favour of urban labour markets, as only two studies were conducted exclusively in rural settings. Of the latter, however, only a single study targeted employment in rural labour markets, with a specific focus on water infrastructure engineering (Chen 2017).

In terms of political indicators, we used the Freedom House Index[17] to indicate the type of political regime contextualising the interventions reported in the included studies. The Freedom House Index was chosen as it is the most widely cited index assessing the condition of political rights and civil liberties around the world; the index has been calculated and reported consistently for over 40 years. Only one study was conducted in a country rated as 'not free': Ethiopia. Eight studies were conducted in 'free' countries, while ten countries had some restrictions on political freedoms and were rated as 'partly free'. In order to zoom in on the particular situation of women in the countries in which the studies were conducted, we coded studies according to the WEF Global Gender Gap Index (World Economic Forum 2017), as well as its sub-dimension the Economic Participation and Opportunity Index. The Global Gender Gap index was chosen at it provides a multi-dimensional breakdown of women's economic empowerment and is a stand-alone index rather than a gendered breakdown of an existing index such as the Gender Development Index. Table 4.1 provides information on both these indexes and how they relate to the evidence included in this review.

Table 4.1 Overview of gender equality settings

WEF Global Gender Gap Index		WEF Gap Index: economic participation	
Above global average (n=6 studies)	Colombia (n=2) Index score: 0.727 Rank: 39	Above global average (n=6 studies)	Colombia (n=2) Index score: 0.749 Rank: 28
	Kenya (n=2) Index score: 0.702 Rank: 63		Kenya (n=2) Index score: 0.710 Rank: 48
	Argentina (n=1) Index score: 0.735 Rank: 33		Liberia (n=1) Index score: 0.612 Rank: 103
	Bangladesh (n=1) Index score: 0.698 Rank: 72		Argentina (n=1) Index score: 0.616 Rank: 101
Global average 0.68 (n= 6 studies)	India (n=4) Index score: 0.683	Global average 0.59 (n= 3 studies)	Ethiopia (n=1) Index score: 0.599

[17] https://freedomhouse.org/report/freedom-world/freedom-world-2017

	WEF Global Gender Gap Index		WEF Gap Index: economic participation
	Rank: 87		Rank: 106
	Peru (n=2) Index score: 0.687 Rank: 80		Peru (n=2) Index score: 0.594 Rank: 111
Below global average (n=7 studies)	Jordan (n=2) Index score: 0.603 Rank: 134	Below global average (n=10 studies)	Nepal (n=1) Index score: 0.578 Rank: 115
	Ethiopia (n=1) Index score: 0.662 Rank: 109		Turkey (n=1) Index score: 0.464 Rank: 129
	Liberia (n=1) Index score: 0.652 Rank: 114		Tunisia (n=1) Index score: 0.444 Rank: 131
	Nepal (n=1) Index score: 0.661 Rank: 110		Bangladesh (n=1) Index score: 0.410 Rank: 135
	Tunisia (n=1) Index score: 0.636 Rank: 126		India (n=4) Index score: 0.408 Rank: 136
	Turkey (n=1) Index score: 0.623 Rank: 130		Jordan (n=2) Index score: 0.381 Rank: 138

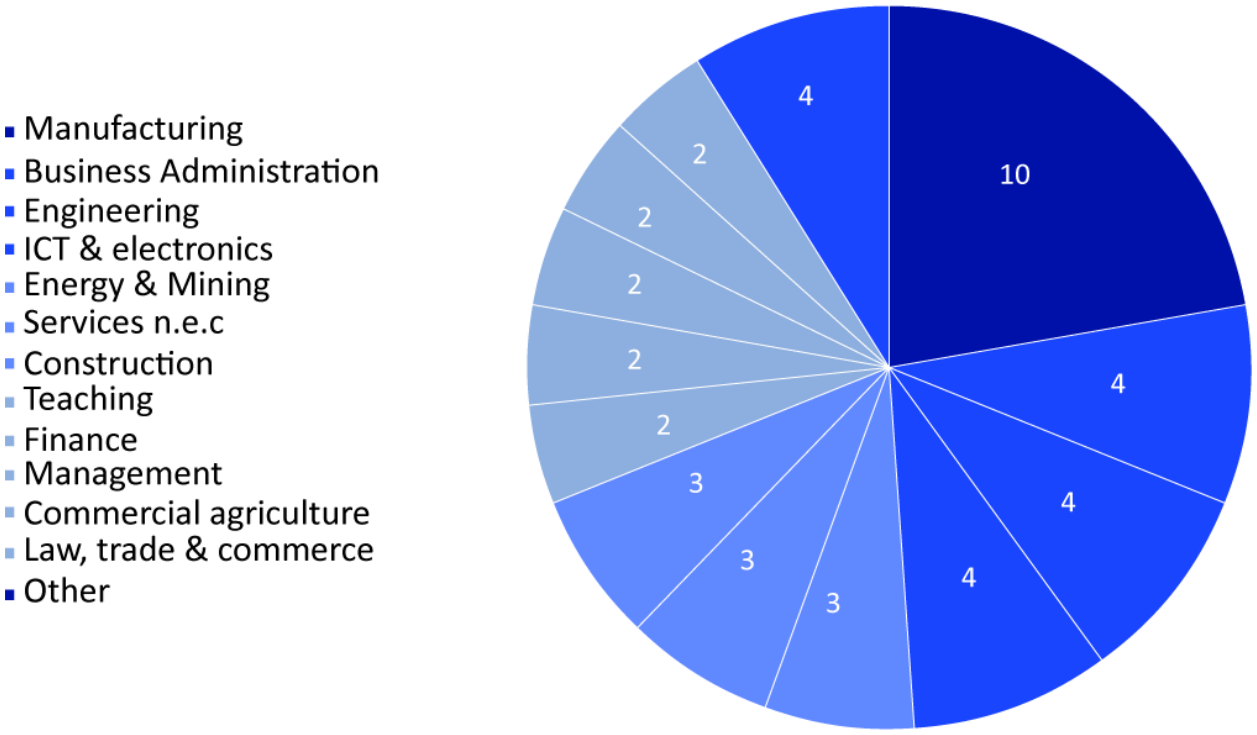
For the GAP Index, the global average as of 2016 is a value of 0.68, indicating that 68% of the global gender gap has been closed. The studies included in our systematic reflect this global average with an average index score of 0.672. This is driven by three groups of studies: one, studies that score significantly above the global average (Columbia, Argentina, Bangladesh, and Kenya); two, studies that score significantly below the global average (Turkey, Tunisia, Liberia, Nepal, Ethiopia, and Jordan); and, three, studies that reflect the global average (India and Peru). The evidence base is thus evenly split between studies with greater than average women's opportunities and participation (n=6), lower than average women's opportunities and participation (n=7), and on par with average women's opportunities and participation (n=6).

We can further contextualise this data by investigating the economic sub-component of the Gap Index, the Economic Participation and Opportunity Index (Column B, Table 4.1). This part of the Gap Index highlights the percentage of the economic gender gap that is closed in each country. Using this lens of investigation, we found that a majority (n=10) of our included studies were conducted in countries below the global average of 0.59. These ten studies had an average value of 0.39, indicating that only 39% of the economic gender gap was closed. In contrast, our review included only six studies that had higher than average economic opportunities for women (Columbia, Kenya, Liberia and Argentina). Three

studies from Ethiopia and Peru reflected the average global economic Gap Index. On average, our 19 included studies were conducted in countries that had closed only 51% of the economic gender gap, significantly below the global average.

In terms of labour market settings, we found that manufacturing was the most frequently cited higher-growth sector, with 10 studies targeting an increase of female employment in this sector (Fig 4.4). This was followed by the business administration sector (n=4), engineering (n=4), ICT and electronics (n=4), energy and mining (n=3); construction (n=3), and services not elsewhere classified (n=3).

Figure 4.4 Overview of the economic sectors in which women were employed



4.4 STUDY POPULATION

The included studies featured a diverse group of women. In most studies, the sample was mixed by gender (n=12) and only seven studies focused exclusively on women. Below, we provide a breakdown of the population characteristics of the women included in both types of studies. The unit of analysis in all studies was individual women. Other structures, such as household or self-help groups, were not used in the studies. The average age of women in the sample was 25.9 years. This reflects well the focus of a majority of the programmes on young women (15-35 years) (n=15). Only three studies included a sample of women with an average age above 30. In a single study the women's age was not specified.

We also investigated whether studies assessed women in particular transition periods, for example, returning to work after childbirth. Two transitions stand out from the sample of studies: (i) the transition from 'unemployment to employment' reported in 14 studies; and (ii) the transition from 'school and/or out-of-school to employment' reported in 12 studies. The transitions of 'within work promotion' and 'from informal employment to formal employment' were marginal, with two studies reporting on each. The same applies to the transition from rural to urban, which was reported in a single study. We can therefore add to the description of the average woman included in this review that she is out of work or has recently completed or dropped out of school.

In terms of level of schooling itself, in all but three studies, women had at least completed secondary education. In six studies, women had completed high school and in three further studies, had graduated from university. Thus half of the sample of women had at least high school education and the remaining half at least 8–9 years of formal education. The overall level of education in the included studies can therefore be described as reasonable, which might have been influenced by the interventions' objective of placing women in higher-growth sectors. The average woman included in this review does possess a decent level of education.

The information presented so far suggests that the women participating in the included interventions did present a special sub-set of the overall population. This situation was indeed actively created by the intervention designs themselves, as the vast majority of programmes included a deliberate targeting of the type of women who could gain access to the programmes. All but three interventions reported in the studies featured design criteria to target particular kinds of women. The most frequently used criteria were location, education, age and occupational status.

In terms of social indicators, only 11 studies provided information on women's marital status. Given the age of the majority of women and the interventions' targeting criteria, it is not surprising that a majority of women in these studies were not married. In 13 studies, the locations in which women resided or the women themselves were described as 'poor', 'very poor', 'disadvantaged', or 'vulnerable'. We also applied the PROGRESS-plus framework to assess multiple dimensions of potential vulnerabilities of the included samples of women. Only the dimension of disability was referenced in a single study. In summary, the average woman included in the studies featuring in this review can thus be described as: being unmarried, 25.9 years old, out of work or having recently completed or dropped out of school, having at least completed secondary school education,

Box 4.1: Key characteristics of the study population

The average women taking part in the reviewed labour market interventions:

- 25.9 years old (younger women)
- unmarried women
- out of work or having recently completed or dropped out of school
- secondary education qualification or higher (better-educated women)
- residing in disadvantaged settings (poorer women)

and residing in a disadvantaged setting.[18] Given this description, it can be stated that the women taking part in the labour market interventions included in our review did not necessarily constitute the most vulnerable women (i.e. they had a higher level of education) and rather represented younger, more mobile women who arguably had a better chance of finding employment. Box 4.1 presents a summary of the key characteristics of the women population featured in the studies included in the review.

4.5 INTERVENTION

Our systematic review covered 20 different interventions supporting women's labour market participation reported in the 19 included studies.[19] In four studies, two different interventions were applied and evaluated while a single study reported on three different interventions. Three interventions were evaluated across different settings. These referred to the World Bank's Adolescence Girls Initiative (AGI) which was implemented and evaluated in studies from Liberia, Nepal, and Jordan; three studies evaluating the impacts of Jovenes programmes in Latin America (Peru, n=2, Argentina n=1); and two studies assessing the effects of the Kenyan Youth Empow erment Program Ninaweza. We divide the included interventions into four overall groupings of programmes:

1. interventions that combine job skills training with job placement services (n=9)
2. interventions that only provide job skills training (n=2)
3. interventions that only provide job placement services (n=3)
4. macro-level interventions of different designs (n=6).

The most common labour market programme approach was the provision of job skills training paired with a subsequent intervention component to place programme participants in employment positions. Job training varied from technical skills (n=9) to soft or life skills (n=6) or a provision of both (n=4). Job placement support included job search assistance, mentoring and networking services, job vouchers, internships, and guaranteed temporary positions. This two-component programme design underwrote the interventions that were part of the Adolescence Girls Initiative programme portfolio, the Latin American Jovenes programmes and the Kenyan Ninaweza programme.

Two interventions only provided job skills training. Both these programmes focused on soft skills to address vertical labour market segregation. The programmes targeted female factory workers in the garment sector attempting to increase their representation among supervisor and management positions within the factories. Three studies applied a job placement support service as the only intervention component. This was an employee-employer searching, screening and matching service, a recruitment service for the business outsourcing sector, and a supply-side voucher to university graduates.

[18] Studies focusing on women that did not fit this average representation were Chen (2017), which focused on older, illiterate women in rural areas and Pfutze (2015), which focused on elderly women receiving social pensions.
[19] Six additional interventions are reported in the 19 studies, but did not target women's wage labour employment.

The six macro-level labour market interventions were heterogeneous. Two interventions could be broadly framed related to women's empowerment. These were a reservation of political seats in local municipalities in India and a change in property and marital laws in Ethiopia. Both interventions increased women's rights and political and economic participation by decree. Two further macro-level interventions provided national labour market subsidies to increase participation. In Tunisia, this was a national supply-side subsidy for university graduates, while in Turkey it was a demand-side subsidy reducing labour costs through the removal of social security contributions for employers. Lastly, two studies from Columbia investigated changes in working hour regulations and social pensions respectively. For all the included interventions, we extracted detailed information on design features, which is reported in section 5.6.1 as part of the QCA.

4.6 OUTCOMES

Our systematic review was concerned with two primary outcomes: (i) women's participation in formal or informal employment; and (ii) women's economic empowerment. Of the 19 included studies, all reported at least one of our measures for women's participation in formal or informal employment outcomes with eleven of those studies then also reporting on a measure of women's economic empowerment.

In relation to women's participation in formal or informal wage labour, the majority of the 19 studies (n=17) reported the outcome in a change in the formal employment status of participants. Two studies reported only on a change in career progression. Additional outcome measures under this outcome category were a change in the nature of employment (e.g. increased wages and social security) (n=11), changing employment or employers (e.g. rural to urban) (n=2) and under-employment (n=1).

In terms of the second primary outcome, women's economic empowerment, the eleven studies identified assessed heterogeneous indicators of empowerment. Seven studies reported on different indexes of women's empowerment; four reported on women's control over income and household spending; three reported on individual savings; while, lastly, women's well-being and access to economic assets was each reported just once. Of the seven studies that reported on indexes of women's empowerment, five focused on indexes for psychological empowerment, while the remaining two constructed their own survey instruments to measure empowerment numerically.

Our systematic review also investigated the intermediate outcomes reported in the included studies. Of the 19 included studies only eight reported on intermediate outcomes. These were almost exclusively measures of employability expressed as job-relevant skills in all instances (n=7). Four studies further investigated access to job-related information and networks as an outcome measure, while three investigated actual access to employment opportunities, such as attending job interviews.

We also investigated the extent to which cost data and cost-effectiveness calculations of the programmes were reported. Only nine studies reported some form of cost data. Four of these merely reported the costs of the intervention, whereas the remaining five conducted formal cost-effectiveness

calculations. Lastly, we assessed the reporting of unintended outcomes. Of the 19 studies, only four provided indications of unintended outcomes: two studies reported on adverse wage effects and two reported on adverse employment effects.

All included outcome measures were continuous outcome measures. We calculated SMD and RR for each outcome measure, with SMDs being the effect size used in the meta-analyses.

4.7 STUDY DESIGN

The 19 included studies applied four main types of impact evaluation designs: RCTs, natural experiments, prospective quasi-experimental designs and retrospective quasi-experimental designs. The vast majority were quasi-experimental (n=9) or RCT (n=8) designs. Only two studies reported results from natural experiments. Of the eight RCTs, no study applied a cluster randomised-controlled design. All interventions were randomly allocated to individuals. Within the quasi-experimental designs, we identified a wider range of study design. Five applied a prospective design based on baseline and endline data points. Of these, three studies used matching and regression techniques to construct experimental groups as comparable as possible. In contrast, two prospective quasi-experiments merely applied random sampling to construct experimental groups without applying formal matching or any other techniques to control for potential confounding variables between experimental groups. In addition, four quasi-experiments applied a retrospective design. These studies collected or accessed survey data sets and then used regression techniques to retrospectively assess how the applied intervention affected population groups differently. Finally, two studies exploited natural experiments in which national policies were phased in differently in the respective countries' federal structures.

The average sample size of the included evaluations was 2,339, with a range of 100 to 8,695 following the removal of outliers. The range of follow-up times in measuring outcomes varied from one month to three years, and on average was approximately 11 months. The median follow-up period was 12 months. In terms of data points, twelve studies conducted only a single follow-up, with five studies conducting two follow-ups. Two of the studies reported three follow-ups. In terms of experimental groups, the majority of studies assessed intervention effects against a clean control group (n=15). Four studies compared the intervention's effects against a second intervention as well as a control group. No studies compared intervention effects against 'business as usual' conditions, in which the control group received the standard practice or programme.

4.8 RISK OF BIAS OF THE INCLUDED STUDIES

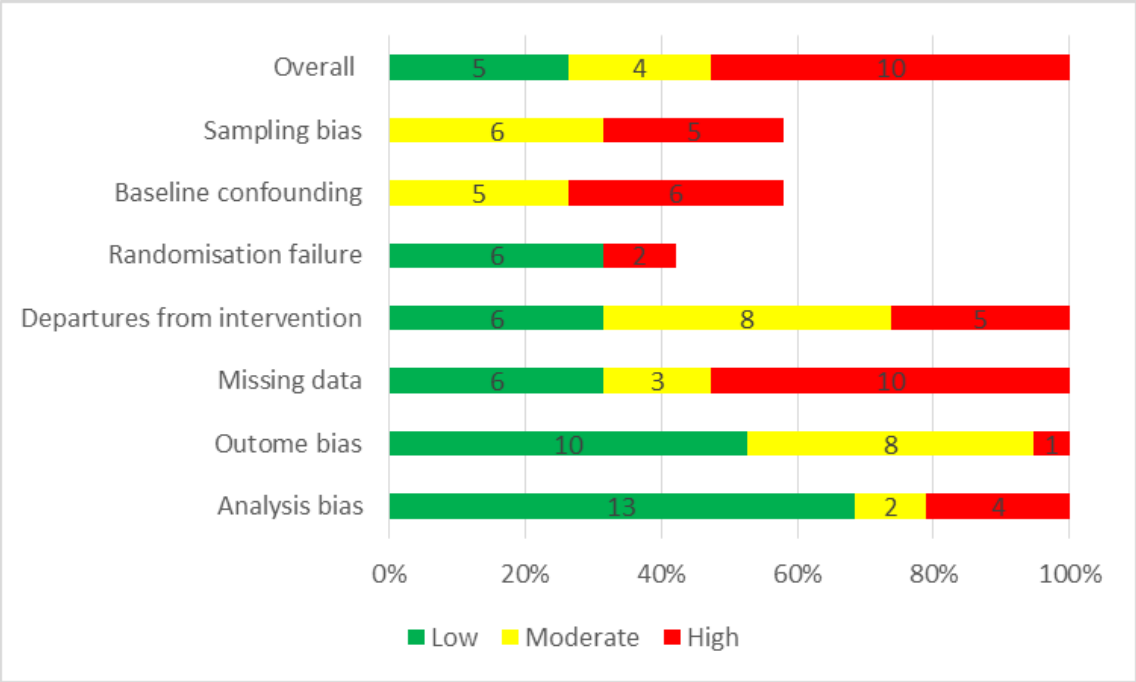
The 19 identified studies in our review were overall subject to a range of biases challenging the trustworthiness of the evidence base. Only five of the included studies had a low risk of bias. Four further studies were rated as having a moderate risk, leaving cautious concerns about whether the reported research results could be attributed to the intervention under investigation. In 10 out of the 19 studies, we identified a high risk of bias, indicating strong concerns about whether the reported

research results could be attributed to the intervention under investigation. Throughout the synthesis, we indicate the underlying risk of bias of the studies included in the different analyses.

All five studies found to have a low risk of bias applied RCT research designs. We consider these five impact evaluations to be the best available evidence on what works to increase women's labour market participation in higher-growth/male-dominated sectors in LMICs. The four studies with a moderate risk of bias comprised three prospective quasi-experimental research designs and one natural experiment. Overall, in these four studies the risk of bias assessment identified moderate concerns in the ability of the studies to adequately control for all potential confounding variables. The 10 studies with a high risk of bias constituted six quasi-experimental designs; four of these used retrospective regression designs based on a single data point and two merely used random sampling techniques to identify two experimental groups without matching techniques or control variables being applied. Three RCTs are also rated as having a high risk of bias. These ratings are mainly a consequence of non-compliance during the selection of participants, resulting in randomisation failure as well as a problem with the completeness and quality of the collected data. Lastly, a single natural experiment was also judged to have a high risk of bias. This rating resulted mainly from an absence of the methodological information required to transparently assess the study's risk of bias.

Figure 4.5 provides a summary of the full risk of bias ratings of the 19 included studies. This information details the high-level findings presented above. In total, missing data (attrition) is the strongest source of bias in the identified evidence with 10 studies being rated as of a high risk of bias in this domain. This is followed by high rates of baseline confounding biases (n=6), sampling bias (n=5) and bias due to departures from intended interventions (n=5). In terms of confounding and sampling biases, these mainly arose from quasi-experimental research designs attempting to create comparable experimental groups. In terms of missing data and departures from intended interventions, these biases largely arose due to challenges in implementing the intended interventions and controlling for the drop-out of participants. Randomisation failure, outcome reporting bias, and selective analysis presented minor sources of bias in the identified evidence. The detailed risk of bias ratings per study are provided in Appendix 12.

Figure 4.5 Overview of the risk of bias assessment



4.9 OVERALL QUALITY OF THE EVIDENCE

In addition to the risk of bias calculations, we also conducted a formal assessment of the overall quality of the evidence included in the review and how this influenced the review's findings. For this purpose, we applied the GRADE strengths of the evidence assessment tool, as explained in section 3.6. GRADE uses the above risk of bias ratings of the included primary studies combined with the consistency, directness and precision of the primary study results to assess the overall quality of the included evidence. The higher the quality of evidence, the more reliable the review results.

In our review, while we include 19 studies in total, we conducted five separate syntheses (see section 5). This was because of the heterogeneity of the included labour market participation interventions. Splitting the 19 included studies into homogeneous groupings of interventions leads to five intervention groupings, each of which is assessed in a different synthesis. The consequence of this is that the number of studies in each synthesis is reduced, which decreases the confidence that we can have in the evidence base informing each synthesis.

We then applied the GRADE framework to assess the quality of the evidence informing each of the five syntheses. Table 4.2 shows a summary version of the results of this assessment. The full version of the GRADE assessment is provided in Appendix 9. Applying the GRADE framework, we established that the overall quality of the evidence included in our five syntheses was low. Only the evidence included in the meta-analysis on combined training and placement interventions was of moderate quality. All the other

syntheses were based on either low-quality evidence (n=1) or very low-quality evidence (n=3). In summary, the small size and heterogeneous, low quality nature of the evidence base limits the findings of our systematic review. The findings from the syntheses based on intervention groupings with low- to very low-quality evidence have to be treated with particular caution.

Table 4.2 Summary GRADE Evidence Profile

Intervention category (outcomes)	No. of studies (design)	Quality
Combined training and placements		
Wage Labour	8 (5 RCTs)	⊕⊕⊕◯ Moderate
Income	9 (5 RCTs)	⊕⊕⊕◯ Moderate
Empowerment	5 (2 RCTs)	⊕⊕◯◯ Low
Soft skills training on promotion		
Career progression	2 (2 RCTs)	⊕◯◯◯ Very low
Empowerment	2 (2 RCTs)	⊕◯◯◯ Very low
Job placement services only		
Wage Labour	3 (3 RCTs)	⊕⊕◯◯ Low
Income	3 (3 RCTs)	⊕⊕◯◯ Low
Empowerment	3 (3 RCTs)	⊕⊕◯◯ Low
National labour subsidies		
Wage Labour	2 (0 RCTs)	⊕◯◯◯ Very low
Macro-level empowerment policies		
Wage Labour	2 (0 RCTs)	⊕◯◯◯ Very low
Empowerment	2 (0 RCTs)	⊕◯◯◯ Very low

5 SYNTHESIS RESULTS

In this section, we report the results of our synthesis of what works to increase women's participation in the wage labour market in LMICs. The synthesis is based on the results of the 19 impact evaluations assessing the impact of labour market interventions included in the review. A total of 20 interventions reported in the 19 included studies were assessed for their effectiveness on women's wage labour, income, and empowerment outcomes. We grouped these 20 interventions into five homogeneous groups of programmes in order to facilitate a meaningful synthesis. These five intervention groupings[20] are:

(i) Interventions combining training with job placement services (n=9)
(ii) Interventions providing soft skills training only to address vertical occupational segregation (n=2)
(iii) Interventions providing job placement services only (n=3)
(iv) Interventions providing national labour subsidies (n=2)
(v) Intervention altering changes to macro structures to support women's empowerment (n=2)

Table 5.1 provides an overview of all included interventions and intervention groupings in the synthesis. We were able to conduct a statistical meta-analysis for the effects of interventions combining training with job placement services only. In the remaining intervention groupings, the small number of studies paired with a lack of statistical information to calculate effect sizes prevented us from conducting a meta-analysis. We report the effects of these interventions using narrative synthesis based on structured summary of findings tables. In total, we therefore report one meta-analysis and four narrative syntheses for the above intervention groupings. From a methodological perspective, our meta-analysis on interventions combining training with job placement services presents the most rigorous synthesis of the evidence base as it draws on a significantly larger number of primary studies. Lastly, within the meta-analysis on interventions combining training with job placement services, we further conducted a QCA to investigate the design features of interventions associated with intervention effectiveness.

In this section, we first present the results of our meta-analysis on the effects of interventions combining training with job placement services. This is then followed by the presentation of each narrative synthesis on the remaining intervention grouping outlined in (ii)–(v) above. Finally, we provide the results of our QCA on the design features associated with the effectiveness of female labour market interventions in LMICs.

[20] These interventions categories differ marginally from the descriptive groupings in section 4.5 as we broke down the macro-economic interventions into two sub-categories for the purpose of narrative synthesis. Two studies could not be grouped in either intervention category.

Table 5.1 Intervention groupings for meta-analysis and narrative synthesis

Sub-category	Intervention	Outcome	Results (g; CI)	Synthesis applied
Training + job placement				
Adoho (2014) Low risk of bias	Adolescent Girls Initiative (World Bank)	Employment status Income	+ 0.302 (0.14, 0.42) + 0.087 (-0.07, 0.25)	Statistical meta-analysis
Chakravarty (2016) Moderate risk of bias	Adolescent Girls Initiative (World Bank)	Employment status Income	+ 0.166 (0.96, 0.24) + 0.214 (0.14, 0.28)	
Chen (2017) High risk of bias	Barefoot Mechanics Programme	Employment status Income	+ 0.33 (0.06, 0.60) + 0.21 (-0.06, 0.48)	
de Azevedo [21] (2013^{T1}) High risk of bias	Kenya Youth Empowerment Program: Ninaweza	Employment status Income	+ 0.287 (0.09, 0.48) + 0.291 (0.10. 0.49)	
de Azevedo (2013^{T2}) High risk of bias		Employment status Income	+ 0.178 (-0.02, 0.37) + 0.30 (0.11, 0.50)	
Elías (2004) Moderate risk of bias	ProJoven	Employment status Income	+ 0.146 (0.07, 0.22) + 0.016 (-0.06, 0.09)	
Galdo (2008) High risk of bias	ProJoven	Income	+ 0.185 (0.09, 0.28)	
Groh (2012) Low risk of bias	Adolescent Girls Initiative (World Bank)	Employment status Income	- 0.032 (-0.18, 0.12) - 0.037 (-0.18, 0.11)	
Honorati (2015) High risk of bias	Kenya Youth Empowerment Program: Ninaweza	Employment status Income	+ 0.053 (-0.10, 0.21) + 0.146 (-0.01, 0.30)	
Ñopo (2007) Moderate risk of bias	ProJoven	Employment status Income	+ 15% change in employment rates[22] +93% change in income[23]	

[21] The study by de Azevedo evaluated two different labour market interventions referred to as treatment 1 (T1) and treatment 2 (T2). For more detailed information please see section 5.1.
[22] Compared to +11% change in employment rates for men
[23] Compared to +11% change in income for men

Sub-category	Intervention	Outcome	Results (g; CI)	Synthesis applied
Soft skills training only				
Adhvaryu (2016) Low risk of bias	PACE Program[24]	Career progression Empowerment	+0.082 (0.01, 0.17) Mixed effects	Narrative synthesis
Macchiavello (2015) High risk of bias	Management skills training	Career progression Empowerment	More female managers Improved (self) perceptions of female managers	Narrative synthesis
Job placement only				
Groh (2012) Low risk of bias	Adolescent Girls Initiative (World Bank)	Employment status Income	+0.832 (0.68, 0.99) +0.842 (0.69, 1.00)	Narrative synthesis
Groh (2014) Low risk of bias	Search, screening and matching services	Employment status Income	No effects No effects	Narrative synthesis
Jensen (2010) Low risk of bias	Business outsourcing recruiting services	Employment status Income	2.4% increase in employment No effects	Narrative synthesis
National labour subsidies				
Ayhan (2013) High risk of bias	Employment law No. 5763	Employment status	3% higher probability of being employed	Narrative synthesis
Broecke (2013) High risk of bias	Labour law (SIVP)[25]	Employment status	+0.162 (0.08, 0.24)	Narrative synthesis
Macro-level empowerment				
Ghani (2014) High risk of bias	National law:[26] Political reservations for women	Employment status Empowerment	No effects Greater investment in public good accessed by women	Narrative synthesis
Hallward-Driemeier (2013)	National law: family code	Employment status Empowerment	15–24% increase in wage labour Delay in marriage age	Narrative synthesis

[24] Personal Advancement and Career Enhancement
[25] The Stage d'Initiation à la Vie Professionelle (SIVP) programme
[26] Formal name of the law: 73rd and 74th Indian Constitutional Amendment Acts

Sub-category	Intervention	Outcome	Results (g; CI)	Synthesis applied
Moderate risk of bias				
Macro-level (other)[27]				
Medina (2007) High risk of bias	Labour law: 789 of 2002	Employment status	Reduction of weekly work hours by 3.6%	Narrative synthesis
Pfutze (2015) High risk of bias	Colombia Mayor Pension Program	Employment status	No effects	

5.1 EFFECTS OF COMBINING TRAINING AND JOB PLACEMENT SERVICE INTERVENTIONS

We identified nine studies that evaluated the effects of combining training with job placement services as an intervention approach to enhance women's participation in wage labour markets in higher-growth/male-dominated sectors in LMICs. These interventions represented three types of labour market programmes. First, three studies evaluated the effects of the Latin America Youth Training model Jovenes (Elías 2004; Galdo 2008; Ñopo 2007). The Jovenes programmes combine technical and vocational skills training with subsequent internships and job placements for youth in a range of Latin American countries. The programme design thus deliberately tries to enhance youths' practical and employment-relevant skills as well as their work experience. The three studies included in this meta-analysis evaluated Jovenes programmes in Peru and Argentina.

Second, three studies evaluated the effects of programmes conducted under the World Bank's Adolescent Girl's Initiative (AGI).[28] These were the Economic Empowerment of Adolescent Girls and Young Women (EPAG) programme in Liberia (Adoho 2014), the Adolescent Girls Employment Initiative (AGEI) programme in Nepal (Chakravarty 2016), and The Jordan New Opportunities for Women (Jordan NOW) programme (Groh 2012). AGI-inspired programmes build on the initial approach of the Jovenes programmes – that is, combining practical skills training with work experience – but merge this programme design with livelihood- and empowerment-related programme components (Honorati and McArdle 2013). The programmes therefore included design features such as soft skills and life skills

[27] These interventions were too heterogeneous and did not fit in any intervention grouping. We therefore were not able to conduct a synthesis of these studies and only provide a narrative summary in Appendix 13.
[28] http://www.worldbank.org/en/programs/adolescent-girls-initiative

training as well as empowerment-related programme modules. In addition, the AGI model focuses exclusively on young women, while the Jovenes programmes target men and women.

Third, two studies evaluated the effects of Kenya's Youth Empowerment Program Ninaweza (de Azevedo 2013; Honorati 2015). The studies evaluated the impact of the programme in different sites and at different times, with the programme design varying between the evaluations. Ninaweza's design is closely linked to the AGI model of female labour market programmes, but is not exclusively focused on girls and places a special emphasis on ICT skills and employment. Lastly, Chen (2017) was an outlier, as the study evaluates the effects of the Barefoot Mechanics programme in India. Barefoot Mechanics is a programme designed by the Self-Employed Women's Association, an Indian NGO focused on the empowerment of women; it trains women to repair water pumps in rural areas and further guarantees them wage employment to conduct these repairs on behalf of the NGO. Barefoot Mechanics differs from the rest of the interventions as the programme is focused on elderly women, conducted exclusively in rural areas and in one sector of the economy, and is applied at a significantly smaller scale than the remaining interventions. Notwithstanding this outlier, all in all, the included interventions present a homogeneous programme approach and design.

Roughly half of the interventions were evaluated using quasi-experimental research designs (n=5), with the other half applying RCT evaluation designs. In terms of risk of bias, the nine studies included in the meta-analysis were subject to a range of biases. Only two studies were of low risk of bias (Adoho 2014; Groh 2012), with three being of moderate risk (Chakravarty 2016; Elías 2004; Ñopo 2007) and four being of high risk (Chen 2017; de Azevedo 2013; Galdo 2008; Honorati 2015). All but two studies assessed wage labour outcomes of women as a change in employment status as reported in survey data. Galdo (2008) relied on longitudinal survey and administrative data for measures of employment while Elías (2004) accessed administrative data sets to assess the wage labour outcomes of women. Finally, we were able to calculate effect sizes for all studies but Ñopo (2007), which only reported frequencies and percentage change of employment rates without any measure of variance or significance being stated. The Galdo (2008) study is not included in this meta-analysis on employment outcomes, as we were only able to calculate effect sizes for income outcomes given a lack of statistical information on the disaggregated employment outcome for females. For de Azevedo (2013), we calculate two effect sizes, as the study compared the effects of two different types of training and job placement services interventions: first, training that combined ICT and life/soft skills followed by job placement services (T1), and second, training that only provided ICT skills followed by job placement services, but no life/soft skills (T2).

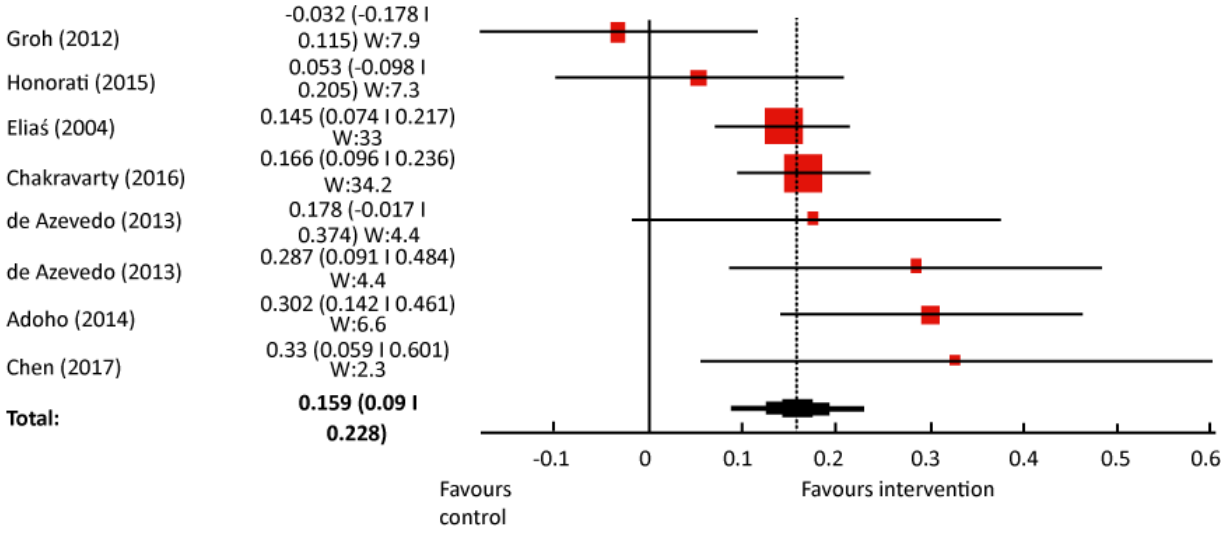
5.1.1 COMBINED EFFECTS OF TRAINING AND JOB PLACEMENTS ON LABOUR MARKET OUTCOMES

We conducted a meta-analysis to identify the overall effects of the described training and job placement services on women's wage labour market outcomes in higher-growth/male-dominated sectors in LMICs. We used a random effect meta-analysis model, as the true effect across studies was likely to differ due to various socio-economic backgrounds, intervention designs, etc. The results of the different meta-

analyses are graphically represented in forest plots. The results from sensitivity and moderator analyses are reported in tabular format. The effect sizes for wage labour and income outcomes are expressed in terms of the SMD of the respective outcome measures and display the change in outcomes in the group of women receiving the labour market intervention over the non-participants in the control group. The pooled effect size reported in the meta-analysis can be read as the number of standard deviation changes in the respective wage labour participation or income change of the experimental groups.

The results of our meta-analysis of combined training and placement interventions are presented in Figure 5.1. Applying the GRADE framework, the evidence on the effects of these intervention on women's wage employment is of **moderate quality** (Appendix 9). The meta-analysis includes eight different interventions and a total of 10,207 participants. Overall, we find that training and placement interventions lead to a positive combined effect size of 0.159 (0.09, 0.23). This identified effect can be described as small to moderate and represents a 7.8% difference in changes in employment rates in favour of the women participants. The meta-analysis results are subject to a moderate degree of heterogeneity (Q=14.8; p= 0.04; i^2=52.5%; tau^2=0.005), and the confidence intervals of all but one study overlap. In order to assess the robustness of the identified effect, we next report the results of our moderator and sensitivity analyses, which investigate whether the observed overall effect might be driven by variables other than the applied labour market interventions.

Figure 5.1: Meta-analysis of training and placement on wage labour outcomes



Heterogeneity: Q=14.8; p=0,0393; i^2=52,5%; tau^2= 0,0045; total participants: 10,207

SENSITIVITY ANALYSIS: TRAINING AND JOB PLACEMENTS ON LABOUR MARKET OUTCOMES

We investigated whether the variance in effect sizes might be caused by factors related to the applied evaluation design (i.e. study type, risk of bias, outcome measure and period of follow-up). For example, a more rigorous evaluation approach might yield systematically different effect sizes from those with a less robust evaluation design. We therefore investigated the sensitivity of our pooled effect estimate to the above design factors. Table 5.2 presents an overview of how the meta-analysis results vary if different groups of studies are combined according to the above design features. Differences in the pooled effect size for each variable could indicate that the overall results of the meta-analysis are sensitive to the design variable under investigation. It is, however, important to note that Table 5.2 presents merely an observational approach to uncover possible sensitivities that we then formally assessed statistically using a one-way random effects ANOVA model.[29]

Table 5.2 Sensitivity analysis of training and placement on wage labour outcomes

Variable	SMD	95% CI	Q	Tau^2	I^2	P-value	Sample	Sensitivity
Training + placement: all studies	0.159	0.090, 0.228	14.8	0.005	52.5%	0.0393	8	
Study type:								Results not sensitive
Randomised controlled trial	0.151	0.017, 0.285	12.8	0.016	68.6%	0.0159	5	
Quasi-experimental design	0.162	0.112, 0.211	1.69	0.00	0%	0.4290	3	
Risk of bias:								Results not sensitive
Low risk of bias	0.133	-0.193, 0.46	9.08	0.049	89.0%	0.0026	2	
Moderate risk of bias	0.156	0.106, 0.206	0.16	0.00	0%	0.69	2	
High risk of bias	0.189	0.062, 0.316	4.98	0.007	39.7%	0.174	4	
Period of follow-up:								Results not sensitive
1 year or less	0.172	0.047, 0.298	14.6	0.016	65.7%	0.0123	6	
2 year or less	0.156	0.106, 0.206	0.16	0.00	0%	0.69	2	
> 2 years	n/a	No observations					0	
Outcome measure								Results not sensitive
Employment status: survey data	0.166	0.075, 0.257	14.7	0.008	59.1%	0.023	7	
Employment status: admin data	0.146	0.074, 0.217	0	0.00	n/a	n/a	1	

[29] The same process applies to all sensitivity analyses reported in this review.

In our meta-analysis on combined training and placement interventions, we combined results from RCTs and quasi-experimental designs. However, the results of the meta-analysis were not sensitive to the applied study design (Q=5.42; p=0.88). The same findings hold true for the level of bias of the research. While we observed that a higher risk of bias led to larger effects, testing for the significance of this difference in effect sizes established that variances in the quality of studies did not influence the overall results of the meta-analysis (Q= 3.82; p=0.76; Q= 2.97; p=0.64). We further tested whether differences in the period of follow-up systematically affected the pooled effect size, but rejected this assumption too (Q=4.73; p=0.82). As only a single study did not use household survey data as an outcomes measure of employment, we could not run a formal sensitivity analysis for this variable. However, we observed that there did not seem to be a significant difference in results. In sum, we can therefore rule out that variances related to study design systematically influenced the results of our meta-analysis.

MODERATOR ANALYSIS: TRAINING AND JOB PLACEMENTS ON LABOUR MARKET OUTCOMES

In addition to assessing whether variables relating to study design influenced the robustness of the meta-analysis, we further conducted a range of moderator analyses. These investigated variables related to the intervention design, the characteristics of the population and the characteristics of the intervention setting that might systematically moderate intervention effects identified in the meta-analysis (Table 5.3). For example, training and placement programmes in UMICs might be more effective than in LICs, and so forth. We defined an a priori list of potential moderator variables in the review protocol (Langer et al 2017), which informed Table 5.3. Unfortunately, two moderator categories suffered from a lack of reporting within the included studies, and we were unable to assess the potential influence of these variables on the identified effects. These were PROGRESS-plus categories such as the age, religion and social capital of the women taking part in the intervention, as well as the UN categorisation of women as 'poor', 'extremely poor' and 'non-poor'. When reporting the moderator analysis, we used the same structure as in the sensitivity analysis based on an observational overview table followed by a one-way random effects ANOVA model. That is, Table 5.3 presents an overview of how the meta-analysis results vary if different groups of studies are combined according to the moderator variables under investigation. Differences in the pooled effect size for each variable could indicate that the overall results of the meta-analysis are sensitive to the moderator variable under investigation, which is then formally tested in the ANOVA model.

We first compared the effects of interventions in different socio-economic settings against each other Despite the observation that single programmes conducted in LICs have a higher pooled effect (*g*=0.302; 0.14, 0.46), the formal moderator analysis rejected the assumption that intervention effects in LICs and UMICs varied significantly from effects in low*er*-middle income countries (Q=6.69; p=0.48). The same finding emerged with regard to the UN classification of countries for which we had data on high-fertility agrarian societies and declining-fertility formalising economies (Q=7.34; p=0.38). In terms of intervention design, we further identified a range of non-significant moderator variables. The pooled effect of the meta-analysis was not moderated by differences in design factors related to the inclusion

of soft/life skills in the programme design (Q=6.73; p=0.69) or whether the intervention participants were of mixed or single gender (female only) (Q=6.87; p=0.76). In addition, we investigated whether the type of programme might systematically drive intervention effects, but when comparing the AGI-model against the Ninaweza and Jovenes programmes, we did not find any significant differences in the moderator analyses (Q=3.59; p=0.56). Lastly, we attempted to assess whether intervention effects varied significantly by the type of intervention provider (public vs private), but as all the included interventions were implemented by private providers, we were unable to investigate this moderator.

In terms of the economic sector that the intervention targeted for women's wage labour employment, we observed that the effects of programmes in the ICT and electronics sector and the service sector showed the largest effects (*g*=0.195 and *g*=0.207). The finance and business administration sector shows the smallest effects (*g*=0.009 and *g*=0.135). We could not formally test these results, as most interventions targeted employment in multiple higher-growth sectors simultaneously and the pooled effects per sector were therefore not based on independent effect sizes. Lastly, all but one intervention focused on the wage employment of young women (15–35 years), and we therefore could not assess the variance of effects from this moderator either. In sum, we did not identify a moderator variable that systematically influenced the results of our meta-analysis. Taken together with the results from the sensitivity analysis we are confident that our meta-analysis results are robust to a range of variables that could potentially systematically influence intervention effects. The studies included in the analysis, as well as the evaluated interventions and their reported effects present, by and large, a fairly homogeneous body of evidence, explaining the lack of significance in the sensitivity and moderator analyses.

Table 5.3 Moderator analysis of training and placements on wage labour

Variable	SMD	95% CI	Q	Tau²	I² (%)	P-value	Sample	Sensitivity
Training and placement: all studies	0.159	0.090. 0.228	14.8	0.005	52.5	0.0393	8	
Socio-economic contexts:								
LICs	0.302	0.142, 0.461	0	0.00	n/a	n/a	1	Results not sensitive
LMICs	0.142	0.048, 0.237	11.2	0.007	55.5	0.0468	6	
UMICs	0.146	0.074, 0.217	0	0.00	n/a	n/a	1	
High fertility agrarian societies	0.2	0.078, 0.323	5.92	0.008	49.4	0.115	4	
Declining fertility urbanising societies	n/a	No observations					0	
Declining fertility formalising economies	0.133	0.045, 0.221	7.72	0.004	61.1	0.0523	4	
Ageing societies	n/a	No observations					0	
Intervention characteristics:								Results not sensitive
Soft skills	0.152	0.058, 0.247	13.1	0.008	61.7	0.023	6	
No soft skills	0.19	0.036, 0.344	1.66	0.007	39.9	0.197	2	
Only women	0.173	0.034, 0.313	11.1	0.013	73.1	0.011	4	
Women & men	0.149	0.079, 0.219	3.52	0.0009	14.7	0.319	4	
Private provider	0.159	0.090. 0.228	14.8	0.005	52.5	0.039	8	
Public provider	n/a	No observations					0	
AGI-model	0.145	-0.012, 0.3	9.56	0.015	79.1	0.008	3	
Ninaweza	0.161	0.024, 0.299	3.51	0.006	43	0.173	3	
AGI + Ninaweza	0.152	0.058, 0.247	13.1	0.008	61.7	0.023	6	
Jovenes model	0.146	0.074, 0.217	0	0.00	n/a	n/a	1	
Economic sector:								
ICT & electronics	0.195	0.119, 0.271	7.35	0.0028	32	0.196	5	
Manufacturing	0.146	0.098, 0.193	1.47	0.00	0	0.418	3	
Construction	0.156	0.106, 0.206	0.159	0.00	0	0.69	2	
Business administration	0.135	-0.060. 0.331	7.19	0.0214	72.2	0.028	3	
Finance	0.009	-0.096, 0.115	0.624	0.00	0	0.429	2	
Services n.e.c.	0.207	0.0574, 0.356	3.05	0.008	67.2	0.081	2	
Population characteristics:								
Young women (15-35 years)	0.15	0.0798, 0.219	13.1	54.2%	0.0043	0.0416	7	
Women >35 years	0.33	0.059, 0.601	0	0.00	n/a	n/a	1	

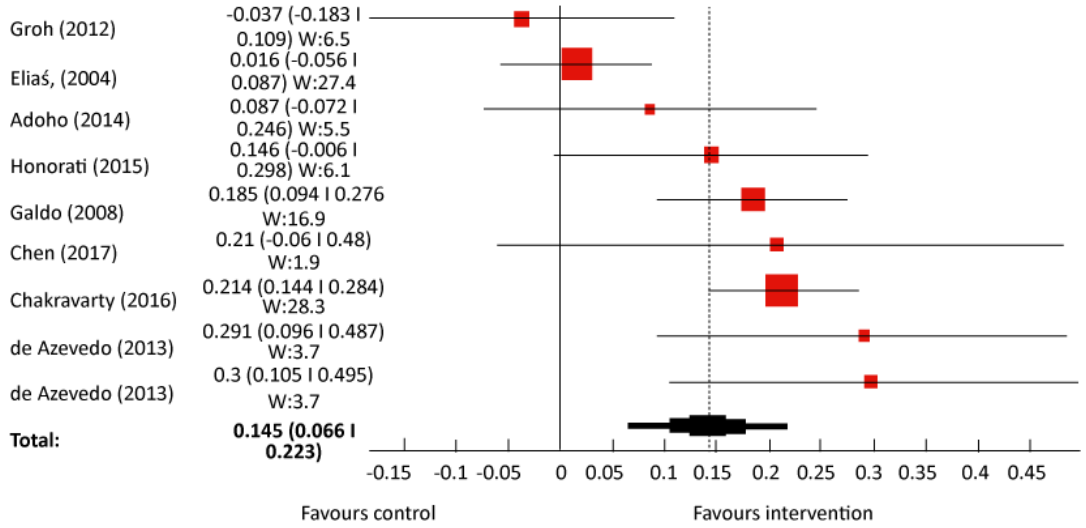
Variable	SMD	95% CI	Q	Tau^2	I^2 (%)	P-value	Sample	Sensitivity
PROGRESS-plus	Insufficient information							
UN categories of women	Insufficient information							

All in all, the results of our meta-analysis of eight combined training and job placement interventions suggest that there is a small to moderate effect on women's wage labour participation in higher-growth/male-dominated sectors. This effect is statistically significant with a narrow confidence interval and consistent across the included studies. Given the homogeneity in the intervention design and observed outcomes, no moderator variables were found to be significant and neither are the results of the meta-analysis sensitive to the applied study designs. Further, applying the GRADE framework, the evidence on the effects of combined training and placement's intervention on women's wage employment is of **moderate quality** (Appendix 9). We therefore present cautious evidence that combined training and placement interventions can improve women's wage labour employment outcomes. The small number of studies included in the meta-analysis as well as the risk of bias identified caution against strong claims to the interventions' combined positive effects.

5.1.2 COMBINED EFFECTS OF TRAINING AND JOB PLACEMENTS ON INCOME

Following the meta-analysis on the effects of training combined with job placements on women's wage labour employment, we next investigated whether these employment outcomes translated into an increased income for women. The results of this meta-analysis are presented in the forest plot in Figure 5.2. Applying the GRADE framework, the evidence on the effects of combined training and placement interventions on women's income is of **moderate quality** (Appendix 9). This meta-analysis includes nine different interventions, due to the inclusion of the Galdo (2008) study and features a total of 12,084 participants. The analysis indicated a positive effect of training combined with job placements on women's income of 0.145 (0.07, 0.22). This pooled effect size is slightly smaller than the effect identified in the meta-analysis on employment outcomes (c.f. *g*=0.159) and can similarly be described as a small to moderate effect size. Expressed in terms of percentage change, the pooled effect indicated a 7.2% difference in changes in income in favour of the women taking part in combined training and placement

Figure 5.2: Meta-analysis of training and placement on income



Heterogeneity: Q=27,9; p=0,00049;i^2=71,4%; tau^2=0.009; total participants: 12,084

interventions. The meta-analysis results on income are subject to a larger degree of heterogeneity (Q=27,9; p=0.0001; i^2=71,4%; tau^2=0.009), which is driven by the inclusion of the Galdo (2008) study. As per above, we assess the robustness of the identified combined effect through moderator and sensitivity analyses, which are presented in tabular format below.

SENSITIVITY ANALYSIS: TRAINING AND JOB PLACEMENTS ON INCOME

We assessed the sensitivity of our meta-analysis results against the same criteria related to study design as in the meta-analysis on employment outcomes. Table 5.4 presents the descriptive results of this analysis. Again, the meta-analysis results are not sensitive to the applied study designs, that is, RCTs and quasi-experimental designs (Q=6.18; p=0.99). A statistically significant difference exists in the effects of studies of a different risk of bias. Studies with a high risk of bias reported a significantly higher average effect (g=0.204) than studies with a low risk (g=0.021) (Q=3.43; p<0.05). There was no difference between studies with a low risk of bias and those with a moderate risk (Q=2; p=0.42) and between studies with moderate and high risks (Q=3,43; p=0.4). We examined whether the study results might differ according to the period of follow-up, but the meta-analysis results were not sensitive to this variable (Q=7.87; p=0.73). Lastly, we investigated the applied outcome measure as a potential moderating variable, but again did not find any significant differences between measures (Q=7.13; p=0.67). In sum, the results of the meta-analysis are thus robust to the applied study designs with the exception that high risk of bias studies generated significantly larger effect sizes than low risk of bias studies.

Table 5.4 Sensitivity analysis of training and placement on income

Variable	SMD	95% CI	Q	Tau^2	I^2	P-value	Sample	Sensitivity
Training and placement: all studies	0.145	0.07, 0.22	27.9	0.009	71.4%	0.0005	9	
Study type:								Results not sensitive
Randomised controlled trial	0.147	0.02, 0.27	10.9	0.0124	63.2%	0.0282	5	
Quasi-experimental design	0.146	0.03, 0.26	17.1	0.0103	82.4%	0.0007	4	
Risk of bias:								Results sensitive for low risk of bias vs high risk of bias
Low risk of bias	0.021	-0.1, 0.14	1.27	0.0017	21.3%	0.26	2	
Moderate risk of bias	0.115	-0.08, 0.31	15.1	0.0183	93.4%	0.0001	2	
High risk of bias	0.204	0.14, 0.27	2.43	0.000	0%	0.657	5	
Period of follow-up:								Results not sensitive
1 year or less	0.165	0.08, 0.25	13.5	0.007	55.7%	0.0352	7	
2 year or less	0.137	0.009, 0.27	16.7	0.011	88.1%	0.0002	2	
> 2 years	n/a	No observations					0	
Outcome measure:								Results not sensitive
Income: survey data	0.14	0.08, 0.16	26.4	0.11	73.4%	0.0004	7	
Income: admin data	0.016	-0.06, 0.09	0	0.00	n/a	n/a	1	
Income: longitudinal panel	0.185	0.09, 0.28	0	0.00	n/a	n/a	1	

MODERATOR ANALYSIS: TRAINING AND JOB PLACEMENTS ON INCOME

We next investigated whether the results of the meta-analysis on women's income were moderated by the variables related to the characteristics of the intervention setting, intervention design or population characteristics (Table 5.5). In terms of socio-economic settings, there were no significant differences between lower-middle income countries and UMICs (Q=6.12; p=0.42) or high-fertility agrarian societies and declining fertility formalising economies (Q=6.63; p=0.3). The results of the meta-analysis were thus not affected by the socio-economic setting in which the training and placement interventions were applied.

With regard to design characteristics, we found that the introduction of soft skills did not systematically alter the intervention effects compared to just technical and vocational skills training (Q=6.9; p=0.62). The same findings hold for the focus of interventions on single or mixed gender designs (Q=6.55; p=0.33). We next investigated whether there were systematic differences in the effects of the three different training and placement intervention models included in the synthesis. However, neither the difference in effects between the AGI and Jovenes models (Q=2.66; p=0.99) nor between Ninaweza and Jovenes (Q=2.99; p=0.2) were statistically significant. The same result applies to the difference between the AGI and Ninaweza programmes and the Jovenes-modelled interventions (Q=6.42; p=0.52). The difference in programmes designs thus does not translate into a systematic difference in wage labour effects.

We further examined the income effects of training and placement interventions in different economic sectors. Again, this was based on observation only, given that independence of effect sizes is undermined in these calculations. Bearing in mind this caveat, we found that the ICT and electronics, manufacturing, and engineering sectors display larger than average income effects, while the finance and services sectors displayed smaller than average income effects. Lastly, due to a lack of diversity in intervention designs, we could not assess whether the meta-analysis results were moderated by the type of implementing organisation (i.e. public vs private provider) and whether younger women experienced significantly different labour market effects from older women.

Table 5.5 Moderator analysis of training and placement on income

Variable	SMD	95% CI	Q	Tau²	I²	P-value	Sample	Sensitivity
Training and placement: all studies	0.145	0.07, 0.22	27.9	0.009	71.4%	0.0005	9	
Socio-economic contexts:								Results not
LICs	0.145	0.07, 0.22	0	0.00	n/a	n/a	1	sensitive
LMICs	0.178	0.08, 0.28	12.2	0.008	59.2%	0.0315	6	
UMICs	0.098	-0.09, 0.26	8.27	0.013	87.9%	0.0041	2	
High fertility agrarian societies	0.192	0.09, 0.29	4.18	0.003	28.2%	0.243	4	
Declining fertility urbanising societies	n/a	No observations					0	
Declining fertility formalising economies	0.113	0.00. 0.22	21.9	0.011	81.7%	0.0002	5	
Ageing societies	n/a	No observations					0	
Intervention characteristics:								Results not
Soft skills	0.161	0.07, 0.26	13.5	0.009	62.9%	0.0193	6	sensitive
No soft skills	0.117	-0.02, 0.26	9.11	0.011	78%	0.0105	3	
Only women	0.341	0.02, 0.67	61.5	0.102	95.1%	0	4	
Women and men	0.168	0.05, 0.28	16.5	0.001	14.7%	0.0025	5	
Private provider	0.145	0.07, 0.22	27.9	0.009	71.4%	0.0005	9	
Public provider	n/a	No observations					0	
AGI model	0.098	0.06, 0.26	10	0.016	80%	0.0068	3	Results not
Jovenes model	0.098	-0.07, 0.26	8.27	0.013	87.9%	0.0041	2	sensitive
Ninaweza	0.228	0.13, 0.33	2.06	0.000	2.74%	0.358	3	
AGI + Ninaweza	0.161	0.07, 0.26	13.5	0.009	62.9%	0.0193	6	

Variable	SMD	95% CI	Q	Tau^2	I^2	P-value	Sample	Sensitivity
Economic sector:								
ICT and electronics	0.202	0.14, 0.26	4.4	0.001	9.17%	0.354	5	
Manufacturing	0.209	0.16, 0.26	2.47	0	0%	0.65	5	
Construction	0.137	0.01, 0.27	16.7	0.011	88.1%	0.0002	3	
Engineering	0.213	0.14, 0.29	1.54	0	0%	0.463	2	
Business administration	0.177	-0.06, 0.41	10.5	0.035	81%	0.0052	3	
Finance	0.053	-0.13, 0.23	2.88	0.011	65.3%	0.0897	2	
Services n.e.c.	0.094	-0.02, 0.21	8.27	0.008	75.8%	0.016	3	
Population characteristics								
Young women (15-35 years)	0.127	0.04, 0.23	20.2	0.01	70.3%	0.0026	8	
Women >35 years	0.33	0.06, 0.60	0	0.00	n/a	n/a	1	

In summary, our meta-analysis on the effects of nine combined training and placement interventions finds a small to moderate effect on women's wage labour participation in higher-growth/male-dominated sectors. This effect is statistically significant with a narrow confidence interval and consistent across the included studies. The identified effect is robust to a range of moderator and sensitivity analyses, but the risk of bias of studies does influence the results, with an upward bias for low-quality evidence. Different factors relating to intervention design and context do not influence programme effects. Further, applying the GRADE framework, the evidence on the effects of combined training and placement interventions on women's income is of **moderate quality** (Appendix 9). We therefore present cautious evidence that combined training and placement interventions can improve women's income following positive wage labour employment outcomes. The positive gains of wage employment in higher-growth male-dominated sectors thus translate into positive income effects of a similar scale. However, the small number of studies included in the meta-analysis, as well as the risk of bias, indicate caution against strong claims in relation to the interventions' combined positive effects on income.

5.1.3 COMBINED EFFECTS OF TRAINING AND JOB PLACEMENTS ON WOMEN'S EMPOWERMENT

Following the meta-analysis of the effects of training combined with job placement services on wage employment and income outcomes, we also investigated whether the observed positive changes in employment and income translated into women's empowerment. However, women's empowerment as an outcome was only measured in five combined training and placement interventions, which are listed in Table 5.6.[30] These are to two AGI-modelled interventions, one Ninaweza programme, one Jovenes-modelled intervention, and the Barefoot Mechanics programme. Empowerment was measured at an individual level using measures of economic empowerment (n=5) such as control over household spending and measures of psychological empowerment (n=2) such as confidence levels and outlook on life. One study measured empowerment at a macro-economic level using the Gender Occupational Segregation Index. Overall, the five studies included in this synthesis are subject to a range of biases with two studies being of high and moderate risk of bias respectively and only a single study being rated as of a low risk of bias. Applying the GRADE framework, the evidence on the effects of combined training and placement's intervention on women's empowerment is of **low quality** (Appendix 9). Table 5.6 presents a structured breakdown of the empowerment related findings in the primary studies.

[30] Women's empowerment was measured in Adoho (2014) too but the effect cannot be disaggregated to the wage labour programme component.

Table 5.6 Narrative synthesis of training and placement on empowerment

Study	Programme	Context	Findings
Chakravarty (2016) **Moderate risk of bias**	The Employment Fund and the Adolescent Girls Employment Initiative (Adolescent Girls Initiative) TVET training + Life skills training + Job placement	Asia: Nepal UN: Declining fertility, formalising economies GAP: 0.661 (#110) GII: 0.497 (#115) Programme has positive effect on employment and income.	**Economic empowerment:** EF training participants reported having more money of their own, more control over household spending, and more access to mentors who can advise them on work-related matters. **Psychological empowerment:** Strong gains on psychological empowerment, including significant increases in self-confidence both in life and with regard to entrepreneurial activities.
Chen (2017) **High risk of bias**	Barefoot Mechanics programme TVET training + job reservation	Asia: India UN: Declining fertility, formalising economies GAP: 0.683 (#87) GII: 0.530 (#125) Programme has positive effect on employment and income.	**Economic empowerment:** Programme participants spent more on female-favoured consumption goods and had a greater say in household spending decisions.
de Azevedo (2013) **High risk of bias**	The Kenya Youth Empowerment Program: Ninaweza TVET training + Life skills training + Job placement	Africa: Kenya UN: High fertility agrarian societies GAP: 0.702 (#63) GII: 0.565 (135) Programme has positive effect on employment and income (not statistically significant).	**Subjective economic empowerment:** The RCT established that, among the young women who were not confident in their skill set, the Ninaweza programme was successful in bolstering their confidence. Treatment group participants saw gains in the kinds of life skills items that pertained to workplace behaviour and searching for a job.
Groh (2012) **Low risk of bias**	The Jordan New Opportunities for Women (Jordan NOW) (Adolescent Girls Initiative) Soft skills training + Wage subsidy voucher	Asia: Jordan UN category: declining fertility, formalising economies GAP index (#134): 0.603 GII index (#111): 0.478 Programme has no effect on employment and income.	**Economic empowerment:** There was no significant impact on attitudes towards the role of women. Part of the reason for the lack of effect might be that there was already strong agreement on attitudes related to work – 97% thought women should be allowed to work outside the home. **Psychological empowerment:** There was a significant and large negative interaction effect for receiving both treatments, so that graduates assigned to both treatments had no better current life evaluation than the control group.

Study	Programme	Context	Findings
Ñopo (2007) **Moderate risk of bias**	Youth labour training (ProJoven) TVET training + Job placement	Latin America and Caribbean: Peru UN category: declining fertility, formalising economies GAP index (#80): 0.687 GII index (#86): 0.385 Programme has positive effect on employment and income (not statistically significant).	**Economic empowerment:** Eighteen months after participation in the programme, gender occupational segregation reduces by 30%.

In terms of individual empowerment, three out of four studies identified positive effects on measures of economic empowerment, while a single study found no effects. Following the AGI intervention in Nepal as well as the Barefoot Mechanics programme in India, women had greater control over how financial resources were spent in the household (Chakravarty 2016; Chen 2017). In Kenya, the Ninaweza programme led to an increase in women's reported confidence and positive attitude towards looking for a job in the ICT sector (de Azevedo 2013). Only in Chakravarty (2016) were these observed improvements in economic empowerment accompanied by improvements in psychological empowerment. In all studies, empowerment outcomes followed the directions of intervention effects on employment and income outcomes. That is, in interventions that were found to be effective to increase wage labour participation and women's income, indicators of women's economic empowerment increased too. Ineffective interventions, on the other hand, had no effect on women's empowerment either. This is the case in Groh's (2012) RCT on the effects of the AGI-modelled Jordan NOW programme. The programme failed to support female employment and subsequent increases in income, which directly translated into non-significant effects on economic and psychological empowerment of women.

Ñopo's (2007) evaluation of the Peruvian Jovenes programme presents an outlier in this group of studies as it investigated empowerment on a macro-economic scale using the Gender Occupational Segregation Index. Using the Duncan Index to measure occupational segregation they found a positive change regarding the concentration of females in male-dominated sectors following the Peruvian Jovenes programme. Female participation increased by 30% in traditionally male-dominated sectors. This positive effect again illustrates the positive intervention effects on employment and income outcomes.

In summary, there is currently not sufficient evidence to support strong claims to positive effects of combined training and placement interventions on empowerment outcomes. The evidence base is too small, subject to a high risk of bias, and heterogeneous in the identified effects and applied outcome measures. Applying the GRADE framework, the evidence on the effects of combined training and placement's intervention on women's empowerment is of **low quality** (Appendix 9). There is a very cautious indication that empowerment outcomes might be observed if prior wage labour and income outcomes are positive. In all studies where this theory of change was observed, the intervention design

applied a specific focus on the individual empowerment of women, for example through incorporating soft/life skills components into the intervention design.

5.1.4 COST ANALYSIS OF TRAINING AND JOB PLACEMENTS ON LABOUR MARKET OUTCOMES

Lastly, we investigated the extent to which the included studies assessed the cost and potential cost-effectiveness of the applied training and placement interventions. Unfortunately, information on cost and cost-effectiveness was not reported consistently. Out of the nine studies included in the meta-analysis, only five reported data on intervention costs, and of those, only three conducted a formal cost-benefit analysis. We therefore cannot comment on the overall cost-effectiveness of interventions due to a lack of information reported in the included studies.[31]

Table 5.7 presents an overview of the cost data available in the included studies. In the five studies that did report cost data, the average programme cost per participants was USD 1,143 (range: USD 400–2,000) but programme lengths and intensity varied greatly. In the three studies that conducted a formal cost-benefit analysis, two found the combined training and placement interventions to be cost-effective (Elías 2004; Honorati 2015). These were the Peruvian Jovenes programme and the Ninaweza programme in Kenya. The evaluation of the AGI-modelled EPAG programme in Liberia, on the other hand, found that the programme was not cost-effective (Adoho 2014).

Table 5.7 Cost analysis of training and job placements on labour market outcomes

Study	Intervention cost	Cost-effectiveness
Adoho (2014)	USD 1,650 per participant for a 12-month programme.	**Not cost-effective** For the job skills training, it would take approximately 12 years to recoup the training costs, making it less cost-effective.
Elías (2004)	USD 2,000[32] per participant for an approximately 90-day programme.	**Cost-effective** The programme's rates of return varied from 7% over five years to 17.5% over a woman's lifetime.
Galdo (2008)	USD 514 per participant for a 3-month programme compromising 300 hours of classes and subsequent job placement.	**No cost-effectiveness calculations**
Groh (2012)	USD 400 per participants for a 9-day training course comprising 45 hours of classes and a job voucher of USD 210 per month for a	**No cost-effectiveness calculations**

[31] This finding applies to all intervention categories assessed in this review.
[32] The study dates from 2004 but it is unclear what year the price data is based on.

Study	Intervention cost	Cost-effectiveness
	maximum of six months (USD 1,260) resulting in a total programme cost of ≈ USD 1,660.	
Honorati (2015)	USD 1,150 per participant for a 6-month programme comprising a 2-week lifeskills training, 5-week business training, 5-week TVET training, and 12-week job placement.	**Cost-effective** These estimates show a KES 9,623 gain for females, which implies that it would take about 10 months to offset the costs of the programme.

5.1.5 SUMMARY FINDINGS OF TRAINING AND JOB PLACEMENTS INTERVENTIONS

Our synthesis of the effects of interventions providing training combined with job placements established cautious support that these interventions could be effective to support women's wage labour participation in higher-growth/male-dominated sectors, and their subsequent income. First, based on a meta-analysis of eight interventions, we identified a pooled positive effect size of 0.159 (0.09, 0.23), which translates into a 7.8% greater increase in employment for women taking part in the training and placement programmes as compared to a control group. This change in wage employment is robust to variables related to study design, intervention characteristics and settings, and characteristics of the women taking part. Applying the GRADE framework, this finding of combined training and placement's positive effect on women's wage employment in higher-growth/male-dominated sectors is based on a **moderate quality of evidence** (Appendix 9).

Second, based on a meta-analysis of nine interventions, we identified a positive pooled effect size of training and placement intervention on women's income of 0.145 (0.07, 0.22). This effect size expressed a 7.2% greater increase of income for women taking part in the interventions as compared to a control group. The effect was equally robust to a range of sensitivity and moderator analyses. Studies with a high risk of bias systematically yielded larger effects. Applying the GRADE framework, this finding of combined training and placement's positive effect on women's income is based on a **moderate quality of evidence** (Appendix 9).

In terms of women's empowerment, there is currently insufficient evidence to support strong claims of positive effects of combined training and placement interventions. Applying the GRADE framework, the quality of evidence for this outcome can be described as of **low quality**. As a result, our narrative synthesis merely provides a very cautious indication that empowerment outcomes might be observed if prior wage labour and income outcomes are positive. That is, there is anecdotal evidence that combined training and placement interventions can support women's economic empowerment if the intervention is effective in increasing wage employment and income outcomes too. There is insufficient evidence on the cost-effectiveness of combined training and placement interventions.

5.2 EFFECTS OF SOFT SKILLS TRAINING TO ADDRESS VERTICAL OCCUPATIONAL SEGREGATION

We identified two interventions that provided soft skills training as the only intervention component to address vertical occupational segregation (Table 5.8) (Adhvaryu 2016; Macchiavello 2015).

Table 5.8 Narrative synthesis of soft skills training on addressing vertical segregation

Study	Programme	Context	Findings
Adhvaryu (2016) **Low risk of bias**	Personal Advancement and Career Enhancement (P.A.C.E.) programme Soft skills training (for career progression)	Asia: India UN: Declining fertility, formalising economies GAP: 0.683 (#87) GII: 0.530 (#125 Majority of the work force in factories are females but a majority of line managers and supervisors are male.	**Career progression:** Treated workers were less likely to leave during the programme, and exhibited substantially higher productivity up to nine months after programme completion (SDM: +0.082 (0.01, 0.17). This led to being assigned to more complex tasks and a greater likelihood of promotion. **Empowerment:** First, treatment workers exhibited greater acquisition and use of information; second, treatment workers were more likely to be saving for children's education; third, increase in self-regard.
Macchiavello (2015) **High risk of bias**	Management skills training (for career progression) Soft skills training (for career progression)	Asia: Bangladesh UN category: declining fertility, formalising economies GAP index (#72): 0.698 GII index (#119): 0.520 80% of the work force in factories are females but 95% of line managers and supervisors are male.	**Career progression:** When the trainees were deployed in supervisory roles, production line workers initially judged females to be significantly less effective, and there was some evidence that the lines on which they worked underperformed. But after around four months of exposure, both the perceptions and performance of female supervisors caught up to those of males. **Empowerment:** Prior to the training, the study found that workers at all level of the factory believed that males were more effective supervisors than females. Evidence was presented that the exposure to female supervisors changed the expectations of male production workers with regard to promotion and expected tenure in the factory.

Applying the GRADE framework, the overall quality of evidence on soft skills training interventions on vertical occupational segregation is of **very low quality** (Appendix 9). The studies focused on female factory workers in the garment sectors in India and Bangladesh. Based on the observation that while a majority of workers in garment factories are females, the majority of supervisors in these factories are male, both studies designed a soft skills programme for women in order to enhance their chances of being promoted to managerial posts. Adhvaryu (2016) evaluated a training programme for female production line workers focused on a variety of life skills, including modules on communication, time

management, financial literacy, successful task execution, and problem-solving. Macchiavello (2015) reported on a programme for sewing machine operators that focused on training for skills necessary to become sewing line supervisors, such as production planning and technical knowledge, quality control, leadership and social compliance. Both programmes explicitly aimed to change females' self-perceptions regarding their ability to act as line managers, and, further, in the case of Macchiavello (2015) to changes males' perceptions of females as supervisors. Both studies applied RCT designs to evaluate the effects of the soft skill programmes, which were subject to a low risk of bias (Adhvaryu 2016) and to a high risk of bias (Macchiavello 2015). Table 5.8 presents a structured breakdown of both studies and their findings. We applied narrative synthesis to investigate the studies' findings due to a lack of statistical information reported to calculate effect sizes and the small number of studies identified.

There is a range of commonalities in the findings of both evaluations. First, both programmes identified a positive effect on women's promotion to managerial posts. In the PACE programme, this outcome was achieved through increased productivity and retention of female workers, which then led to their promotion. In the management skills programme in Bangladesh, programme participants were shortlisted already for potential promotion and subsequently took up these positions after the training programme. Second, both programmes aimed to support women in developing positive self-perceptions of themselves as managers, as well as attempting to address males' perceptions of females in managerial positions. The main mechanism to achieve this was exposure to women in management positions, which in both studies led to an increase in women's self-perceptions and confidence to act as line managers. However, in the Macchiavello study this exposure did lead to conflict, as male factory workers regarded female supervisors as competition and initially resist working with them. In the Adhvaryu (2016) study, women were also reported to develop an increased sense of self-efficacy, accessing a greater range of professional development opportunities and investing more in the education of their children. However, other measures of empowerment, for example mental health and risk behaviour, were not affected, so the overall results regarding women's individual empowerment were mixed. Third, both studies established that investment in women to assume managerial positions was beneficial to the factories themselves. Adhvaryu (2016) outlined how investments in women's access to promotions led to increased productivity and retention rates, which allowed the intervention to pay for itself several times over, implying that teaching soft skills in the workplace can be profitable for firms even in high turnover environments. In agreement, Macchiavello (2015) explained that while supporting women to get promoted could lead to conflict with male workers in the short term, the long-term effects of a more diverse managerial base were reported by factory owners to outweigh these short-term risks.

In summary, our systematic review identified two studies that reported positive effects of soft skills programmes to address vertical occupational segregation in garment factories. However, these results are based on a very small set of studies, which, in addition, are subject to moderate and high risks of bias. At this point, the available evidence base only allows us to indicate the use of soft skill programmes to address vertical occupational segregation as a relevant concept which requires further investigation. Applying the GRADE framework, the overall quality of evidence on soft skills training interventions on vertical occupational segregation can be described as **very low quality evidence** (Appendix 9).

5.3 EFFECTS OF JOB PLACEMENT SERVICES ONLY

Our systematic review identified three studies that evaluated the effects of job placement services only (Table 5.9) (Groh 2012; Groh 2014; Jensen 2010). Applying the GRADE framework, the overall quality of evidence on job placement service intervention is of **low quality** (Appendix 9). These interventions did not provide any form of skills training and only aimed to support women in accessing employment opportunities in higher-growth/male-dominated sectors. However, there was large heterogeneity in the applied programme designs to facilitate women's access to wage employment. As part of the Jordan New Opportunities for Women (Jordan NOW) labour market interventions, Groh (2012) provided women with job vouchers only that they could take to a firm while searching for jobs. The vouchers were valid for a period of six months and covered an amount equal to the mandatory minimum monthly wage of 150 JD (USD 210) per month. In a follow-up experiment under the Jordan Now programme, Groh (2014) also tested the effectiveness of a screening and matching service to pair employers with prospective employees. Here, prospective job applicants, that is graduate students, were tested and screened based on a range of educational and psychometric assessments and then matched with wage labour opportunities that fit their screening scores. Jensen (2010), lastly, provided recruiting services for employment opportunities in the business outsourcing industry in India. Women in rural communities were visited by recruitment agents over the course of three years and informed about employment opportunities and how to access them. In terms of the underlying programme mechanisms, Jensen (2010) and Groh (2014) both attempted to overcome a lack of information and matching between employers and prospective employees. Groh (2012), in contrast, attempted to overcome a barrier in the price and rigidity of women's labour.

All three programmes were evaluated using RCT designs and rated as of a low risk of bias. In terms of outcomes assessed, the three programmes each aimed to enhance women's wage labour employment, income and empowerment through the job placement services. In terms of study design and outcome measures, the three studies were thus homogeneous; but varied greatly in terms of the applied intervention design. As we were only able to calculate effect sizes for the Groh (2012) study, we synthesise the programmes' effects using narrative synthesis based on structured overviews of the three studies (Table 5.9).

Table 5.9 Narrative synthesis on job placement services only

Study	Programme	Context	Findings
Groh (2012) **Low risk of bias**	The Jordan New Opportunities for Women (Jordan NOW) (Adolescent Girls Initiative) Wage subsidy voucher	Asia: Jordan UN category: declining fertility, formalising economies GAP index (#134): 0.603 GII index (#111): 0.478	**Employment:** The analysis found that the job voucher led to a 40% increase in employment in the short run, but that most of this employment was not formal, and the average effect was much smaller and no longer statistically significant 4 months after the voucher period had ended. **Income:** The job voucher group earned 64 JD more per month than the control group at midline. However, by the endline, the difference had fallen to only 6 JD per month and was not statistically significant. **Empowerment:** There was no significant impact on attitudes towards the role of women (empowerment index: 0.084 change – non-significant). Further, there was a significant and large negative interaction effect for receiving both treatments, so that graduates assigned to both treatments had no better current life evaluation than the control group (change of -1.005 on Gallup survey – non-significant).
Groh (2014) **Low risk of bias**	The Jordan New Opportunities for Women (Jordan NOW) (Adolescent Girls Initiative) Screening and matching services	Asia: Jordan UN category: declining fertility, formalising economies GAP index (#134): 0.603 GII index (#111): 0.478	**Employment:** Although more than 1,000 matches were made, youth rejected the opportunity to even have an interview in 28% of cases, and when a job offer was received, they rejected this offer or quickly quit the job 83% of the time. **Income:** No effects as too few matches were made (DID: −9.360, not significant).
Jensen (2010) **Low risk of bias**	Business outsourcing recruiting services Job search assistance and placement support	Asia: India UN category: declining fertility, formalising economies GAP index (#87): 0.683 GII index (#125): 0.530	**Employment:** In villages that received the recruiting treatment, paid employment was 2.4 % higher for women aged 18–24. **Income:** No effects. **Empowerment:** Girls aged 5-15 in villages that received the recruiting services were 6 to 7 percentage points more likely to be in school and experienced an increase (z-score: 0.20–0.30) in body mass index, reflecting greater nutrition and/or medical care.

The identified effects of the three job placement programmes varied greatly. First, Groh's (2014) screening and matching service was found to be ineffective and did not influence any of the three assumed outcomes. Women were held back by a strong reservation prestige and rejected the vast

majority of matched employment opportunities. Jensen's (2010) evaluation of addressing informational and matching barriers through recruitment services, however, found significant positive effects on women's employment and empowerment. Three years after the intervention had commenced, women were 2.4% more likely to be employed in the business outsourcing industry, and this increase in wage employment further translated into increased investment in girl children in the rural communities. These changes in investment in girl children were large in comparison, and closed about 40% of the baseline boy-girl gap in school enrolment.

Second, Groh's (2012) wage vouchers were found to have large, short-term effects on women's employment in higher-growth sectors such as finance and business administration. Initial employment gains for women were up to 40% compared to the control group. However, these gains had disappeared four months after the voucher had expired, and long-term employment gains were not significant. It is also interesting to note that Groh's voucher experiment has two additional interventions arms: (i) a soft skills programme and (ii) vouchers combined with soft skills. The vouchers-only intervention design, surprisingly, not only outperformed soft skills training, but further also outperformed a combination of vouchers and soft skills. It is unclear why these results are observed and why there is no interaction effect between vouchers and soft skills programmes (Groh 2012).

In summary, we identified three studies that reported on the use of job placement services as the sole labour market intervention component. These interventions as well as their effects, were highly heterogeneous, and we cannot comment on the aggregate effectiveness of job placement programmes. There was contradicting evidence of effects within and across different outcomes. While the identified studies are of a low risk of bias, the body of evidence is too small and heterogeneous to allow for a strong synthesis of effects. Applying the GRADE framework, the evidence on job placement service intervention can be described overall as of **low quality** (Appendix 9).

5.4 EFFECTS OF NATIONAL LABOUR SUBSIDIES

In our systematic review, we identified only two interventions that evaluated the effects of national labour subsidies on women's wage labour participation in higher-growth/male-dominated sectors (Table 5.10) (Ayhan 2013; Broecke 2013). Applying the GRADE framework, the overall quality of evidence on national labour subsidies is of **very low quality** (Appendix 9). Both national labour subsidies were implemented in the context of macro-economic crises: the 2008/09 global financial crisis leading to a contraction of employment in Turkey (Ayhan 2013) and persistent high graduate unemployment in Tunisia (Broecke 2013). In response to these macro-economic challenges, the governments enacted changes to the macro-economic set-up. In Turkey this was a national labour stimulus referred to as the 'Employment Package Law', while in Tunisia, it was the labour law enacting the SVIP[33] programme. Turkey's employment package constituted a law that exempted employers from social security contributions (SSC) for employees who were women over the age of 18 years and men between 18 and 29 years. The intervention lasted for two years, during which time employers could benefit from the SSC

[33] French: The Stage d'Initiation à la Vie Professionelle, i.e. Initiation into the world of work.

exemption if they hired individuals within the target groups. In essence, the employment package can be described as a national-level demand-side labour subsidy.

Tunisia's SVIP programme, on the other hand, can best be described as a mix between a demand- and supply-side labour subsidy. While graduate students received a direct salary transfer from the government if they accessed employment opportunities under the scheme (i.e. supply-side subsidy), employers also received an SSC exemption if hiring under the graduate employment scheme. In order to be eligible, graduate students were required to have enrolled for science- and engineering-related degrees and had to register with the national employment services. In both labour subsidies in Turkey and Tunisia, the governments actively monitored the implementation and uptake of the subsidies and governance structures were in place to ensure the conditionality and targeting of the policies.

Given the nature of the intervention, both studies relied on national administrative labour market and panel data sets and were limited to a retrospective regression design in order to evaluate the effects of the national labour subsidies. The risk of bias of these designs was rated as high and the study authors acknowledged the limitations of the applied designs to control for all potential confounding variables. Table 5.10 presents a structured breakdown of both studies and their findings. We applied narrative synthesis to investigate the studies' findings as we were unable to conduct a meta-analysis to pool the studies' effects due to a lack of statistical information in the Ayhan (2013) study and the small number of studies identified overall.

Table 5.10 Narrative synthesis on national labour subsidies

Study	Programme	Context	Findings
Ayhan (2013) **High risk of bias**	Labour law: Employment Package Law No. 5763 2008. Demand-side macro-economic subsidy (reduction in employers' social security contributions)	Asia: Turkey UN: Ageing societies GAP: 0.623 (#130) GII: 0.328 (#69) Labour market intervention necessitated by financial crisis of 2008/09	**Employment:** The results suggested a positive effect of the reduction in non-wage costs on employment creation for the targeted group (women). The probability of being hired for a woman aged 30 to 34 increased by 3%.
Broecke (2013) **High risk of bias**	Labour law: The Stage d'Initiation à la Vie Professionelle (SIVP) programme Supply- and demand-side macro-economic subsidy (employees' salary paid by government and reduction in employers' social security contributions)	Africa: Tunisia UN: Ageing societies GAP: 0.636 (#126) GII: 0.289 (# 58) Labour market intervention necessitated by persistent high graduate unemployment.	**Employment:** The analysis showed clearly that SIVP beneficiaries had lower joblessness (DID: −0.069 – significant) and unemployment rates (DID: −0.090 – significant), and that they were much more likely to be hired in the private sector (DID: 0.288 – significant).

Following the introduction of the labour subsidies, both studies identified positive effects on women's wage labour participation in higher-growth/male-dominated sectors. Ayhan (2013) reported that, per quarter over a two-year period, women were 1.4–3% more likely to be hired in the industry and construction sectors than are men. This finding overlaps with Broecke (2013) who established that women's employment in professions related to the natural sciences and engineering increased by 11% following the introduction of the graduate subsidy. There is thus consistency between both studies' findings that national wage subsidies can positively affect women's wage labour employment in LMICs. Broecke (2013) further made a range of recommendations on the design of the subsidy programme, including a greater targeting of those graduates facing the highest risk of unemployment, as well as a combination of the subsidy with other labour market interventions such as training, counselling and job search assistance. These design-related recommendations are supported by Ayhan (2013), who also emphasised the importance or targeting and clear governance structures in the design of national labour subsidies.

In summary, our systematic review identified two studies that reported positive effects of national labour subsidies on women's wage employment in higher-growth/male-dominated sectors. However, these results are based on a very small set of studies, which, in addition, are subject to a high risk of bias. We are therefore unable to comment on the effectiveness of national labour subsidies in the context of women's wage labour employment beyond stating the need for further research based on the two positive effects presented above. Applying the GRADE framework, the overall quality of evidence on national labour subsidies can be described as **very low quality evidence** (Appendix 9).

5.5 EFFECTS OF MACRO-LEVEL POLICIES TARGETING WOMEN'S EMPOWERMENT

In our systematic review, we identified two interventions that aimed to alter macro structures in order support women's labour market participation and empowerment (Table 5.11) (Ghani 2014; Hallward-Driemeier 2013). Applying the GRADE framework, the evidence on macro-level empowerment interventions is of **very low quality** (Appendix 9).The two identified national policies were a change in state-level implementation of political reservations for women in India and the revision of the national family law in Ethiopia. Both policy reforms were designed as a direct response to existing gendered realities of economic and political participation, which disadvantaged women's opportunities. In India, the 73rd and 74th Constitutional Amendment Acts legislated a large-scale devolution and decentralisation of power to local government bodies. Amongst these were the reservations of a third of all seats at each local governance level for women, which was hoped to increase women's political participation and representation. In Ethiopia, the 2000 revision of the national family law legislated spouses' shared decision making regarding marital property, as well as removing restrictions on females working outside the home. These changes were intended to strengthen women's position within the household and to enhance their economic participation. Both interventions were evaluated using retrospective regression designs relying on longitudinal survey and labour market data. These designs were judged to have a moderate risk of bias (Hallward-Driemeier 2013) and of a high risk of bias (Ghani

2014). We applied narrative synthesis to investigate the combined studies' findings due to a lack of statistical information reported to calculate effect sizes and the small number of studies identified.

Table 5.11 Narrative synthesis of macro-level women's empowerment policies

Study	Programme	Context	Findings
Ghani (2014) **High risk of bias**	National law: Political reservations for women in India	Asia: India UN category: declining fertility, formalising economies GAP index (#87): 0.683 GII index (#125): 0.530	**Employment:** No effects on wage labour employment: While overall employment of women in manufacturing did not increase after the reforms (Log: 0.039; 0.154 – not significant), there was significant evidence that more women-owned establishments were created in the unorganised/informal sector (Log: 0.118; 0.024 – significant at 1% level). **Empowerment:** Reserved leadership positions can influence the allocation of local resources in line with a greater general provision of infrastructure and public goods towards women. A second candidate explanation is that the political reservations inspired women to start their own businesses by providing positive role models and nurturing aspiration.
Hallward-Driemeier (2013) **Moderate risk of bias**	National Law: Family Code Economic rights (making employment decisions without the consent of the husband) Increased property rights for women	Africa: Ethiopia UN category: declining fertility, formalising economies GAP index (#109): 0.662 GII index (#116): 0.499	**Employment:** A strong effect (15-24%) in increasing women's share in occupations that were not home-based, paid work, year-round employment, and those with average higher educational requirements. **Empowerment:** The revised family law increased the bargaining power of women and increased the age of marriage from 15 to 18.

Both macro-level empowerment interventions led to unexpected outcomes and pathways to change. First, while both studies identified positive long-term effects on women's participation in the economy, only the change in family laws led to an increase in women's employment in higher-skills sectors. The political quota led to an increase in women's ownership of businesses in higher-growth sectors, but not in female wage employment. Second, both studies reported long-term effects on women's empowerment following the introduction of the legal reforms. In India, the reservation of political seats in the long term enhanced the provision and allocation of public goods and local infrastructure towards women. Women in political leadership positions may have further served as role models to nurture the aspirations and career choices of other women. In Ethiopia, the revisions to the family law and women's enhanced economic status increased females' bargaining power within the household and led to a delay in the age of marriage from 15 to 18 years.

However, in both macro-level policy reforms the effects of the policy changes led to unintended consequences. The revision of the family law in Ethiopia was initially assumed to benefit married women most, but effectively empowered younger, unmarried women as, realising their new economic prospects, they delayed marriage to pursue these new economic opportunities instead. In India, changes in the political reservation of women were not targeted at nurturing economic empowerment of women and were seen primarily as a tool for political empowerment (Ghani 2014). However, changes in social norms and women's aspirations following the political reform then led to subsequent changes in women's pursuit of economic opportunities, primarily in terms of starting their own manufacturing companies.

In summary, these two macro-level empowerment interventions included in our systematic review provide insights into the complexities of altering economic and political structures to support women's wage labour market participation and empowerment. The small size and low quality of the evidence base does not allow us to comment on the overall effectiveness of these interventions. We observed that the interventions' effects on wage labour were mixed, but long-term increases in empowerment might manifest, though not necessarily through the channels assumed during policy design. Applying the GRADE framework, the evidence on macro-level empowerment interventions can be described as being of **very low quality** (Appendix 9).

5.6 DESIGN FEATURES OF INTERVENTIONS AIMING TO SUPPORT WOMEN'S WAGE LABOUR PARTICIPATION

In addition to assessing the aggregate effects of interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors, our systematic review also investigated what type of design features of the interventions were associated with intervention effects. In this investigation we used a two-stage process. First, we used narrative synthesis to derive a list of design features reported in the included studies. Second, we applied QCA in order to assess the configurations of these design features. That is, we tried to investigate the different combinations of design features and their interplay in reinforcing or mitigating intervention effects.

5.6.1 NARRATIVE SYNTHESIS TO IDENTIFY RELEVANT DESIGN FEATURES

We conducted a detailed narrative synthesis in order to generate a list of relevant design features of interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors. This detailed investigation of design features is limited to the nine interventions combining training and placement support services (section 5.1). The heterogeneity across all 26 interventions included in our systematic review did not allow for an investigation of common design features across all the individual programmes. For instance, the design of a national labour subsidy (e.g. Ayhan 2012) has few common design aspects with a soft skills programme to support individual women's promotion in the garment sector (e.g. Machiavelli 2015) or the combination of technical and vocational education and training (TVET) and internships to support youth employment (e.g. Ñopo 2007). As a result, we

focused our narrative synthesis on the most homogeneous and extensive intervention category identified in the review: the nine programmes combining training with job placement support services.[34]

As outlined in sections 3.5–3.6, our narrative synthesis was based on a structured extraction of study design features for each of the nine studies (Appendix 8). Following the extraction of features, we then engaged in a process of consultation and iteration to refine the list of design features. A total of seven iterations led to the identification of seven overall design features (Table 5.12). The seven design features constitute the aspects of training and job placement programmes that were associated with programme effectiveness in the included studies.

Table 5.12 List of intervention design features associated with programme effectiveness

Design features	Example of the design feature
1. Exposure to labour market participation enhancing social norms	• Conducive norms at work • Affirmative action programmes
2. Labour demand-led intervention design	• Consultation of private sector needs
3. Gender-sensitive intervention design	• Women-only programmes • Childcare facilities
4. Provision of soft/life skills and social empowerment training	• Confidence / aspiration building • Non-cognitive work skills
5. Participant profiling and targeting	• Narrow eligibility criteria • Gender-sensitive marketing
6. Clear governance structures for intervention providers	• Payment by results • Provider monitoring
7. Flexibility and responsiveness in intervention implementation and design	• Information management systems • Piloting and iteration

Below, we briefly describe each design feature before commenting on the results of the QCA on the configuration of features.

[34] Data extraction on the design features was conducted for all 26 interventions reported in the 19 included studies. However, following a presentation and joint workshop on the results of the design feature data extraction, we focused our narrative synthesis on the nine studies included in the meta-analysis for the reasons presented above. Data on the design features of the reaming interventions and their relation to the identified seven key features is available on request.

EXPOSURE TO LABOUR MARKET PARTICIPATION ENHANCING SOCIAL NORMS

Exposure to labour market participation enhancing social norms refers to aspects of the training and job placement interventions that support a shift in social norms regarding women's participation in higher-growth and male-dominated sectors. We identified a range of different programme features that were applied to this effect. In the design of the AGI-modelled EPAG programme in Liberia, Adoho (2014) reported an emphasis on the identification of female role models to change the perception of women's participation in higher-growth and male-dominated sectors such as ICT, professional driving and professional house and office painting. This was further enhanced by the formation of 'EPAG teams', groups of women who were provided with a coach or mentor to foster support networks and boost attendance and commitment through peer pressure and support.

A similar approach to building social norms through social organisation among women was applied by the Barefoot Mechanics programme in India (Chen 2017). Here, the programme design included small collectives of women who worked together as a team to fix broken water infrastructure in rural areas. These collectives aimed to provide women with support structures and a means to organise themselves in a male-dominated profession. The name of the programme 'Barefoot Mechanics' further aimed to build an identity of the women being trained and employed as water engineers.

A last example of an explicit intervention design to build labour market participation enhancing social norms refers to the AGI-modelled AGEI programme in Nepal (Chakravarty 2016). In the first iteration of the programme, it struggled to attract sufficient numbers of females interested in taking part in the 12-month training and job placement activities. As a result, the second iteration of the programme in 2011 launched an enhanced communication and outreach strategy to recruit more female trainees. This strategy included the use of newspaper and radio adverts targeting young women and communicating the benefits to women of being employed in non-traditional trades such as mobile phone repair, electronics or construction. These efforts were complemented by an in-depth partnership with local community leaders and organisations in order to encourage applications from women and marginalised groups for non-traditional professions. This partnership included a small monetary incentive in the form of a finder's fee equivalent to about USD 1.25 per person.

LABOUR DEMAND-LED INTERVENTION DESIGN

A second design criterion reported consistently throughout the nine included interventions was the need to tailor training and placement services to the labour demands of the private or public sector. Demand could relate to the type of profession targeted, for example software engineering (e.g. Honorati 2015) or business outsourcing (e.g. de Azevedo 2013) as much as to the particular skill required, for example soft skills (Groh 2012) or vocational skills (Ñopo 2007). All nine training and placement interventions were demand-led in terms of the skills and employment opportunities targeted. While in all interventions, the private sector had an opportunity to shape the labour demand, government too had a prominent role in this. For example, the Ninaweza programme in Kenya was designed around the

country's labour and skills needs identified in Vision 2030, Kenya's national development strategy. By and large, programmes sought both public and private sector inputs in terms of what skills and economic sectors to target. The most common ways to identify labour and skills demands referred to direct consultation with private and public sector decision makers, but national labour data sets were used to identify labour needs too (e.g. Galdo 2008).

A second way to ensure a labour demand-led intervention design related to ensuring employers' advance commitment to the programme. The EPAG and Barefoot Mechanics programmes, for example, established prior employment guarantees for the graduates of their training and placement interventions. The Ninaweza programmes likewise included an upfront financial contribution of private sector organisations to the programme. All in all, a demand-led approach to intervention design emerged as a key design feature in our narrative synthesis.

GENDER-SENSITIVE INTERVENTION DESIGN

Designing training and placement support programmes explicitly around the needs of female trainees in order to increase their attendance and effective participation presented a third design criterion in the included interventions. Such gender-sensitive design could take many forms and was explicitly mentioned as a key characteristic of programmes in all but one study. We identified three main approaches to implementing a gender-sensitive programme design. First, five interventions only offered the training and placement programmes to females. Such an access-focused approached was a popular mechanism, but not all programmes then tailored the interventions to women's needs. For example, while the AGI-modelled Jordan NOW programme was only accessible to young women, there were no further design aspects reported that tailored the programme to women's needs and preferences (Groh 2012).

Second, a number of programmes, mainly the Jovenes-modelled interventions, focused on mitigating against female drop-outs during the programme. In order to support females to access and complete the training and placement programme women received additional stipends if they had children under the age of five (e.g. Elías 2004) and training facilities included access to childcare (e.g. Galdo 2008). However, the overall programme model of the Jovenes labour market interventions was not tailored to women's needs.

Third, more recent training and placement interventions were explicitly designed to support women's labour market participation, particularly the World Bank's AGI and the Ninaweza programme in Kenya. In this programme design, women received additional stipends in order to access and participate in programme activities, but, what is more, the nature and approach of the activities themselves were tailored around women's needs. This included considering the location and timing of the programme activities to cater for demands on women's time and restriction around mobility; gender-sensitive messaging and marketing of the programme in order to attract female participants; being sensitive to the social opportunity cost of women when partaking in interventions focused on male-dominated sectors; and using behavioural techniques to increase the salience of the interventions for women (e.g.

Adoho 2014; Chakravarty 2016; de Azevedo 2015). In addition, these new types of female-targeted labour market interventions (Honorati 2012) also include explicit programme components related to access to sexual and reproductive health and life skills support (see next design feature).

PROVISION OF SOFT/LIFE SKILLS AND SOCIAL EMPOWERMENT TRAINING

A growing design feature referenced as a valuable contribution to increase the effectiveness of training and placement programmes is the incorporation of a life or soft skills programme component. This design approach was implemented by five interventions including the AGI-modelled programmes (Adoho 2014; Chakravarty 2016; de Azevedo 2013; Groh 2012; Honorati 2015). It is based on the assumption that female employees are not only held back by a lack of technical or vocational skills, but, in addition, by a lack of the non-cognitive skills required to access and retain employment opportunities. Such non-cognitive or soft skills can, for example, be setting personal goals, workplace ethics, communication, leadership, self-esteem, conflict resolution or problem solving. The lack of this type of skill can disadvantage women in job interviews as well as inhibit their performance when hired. Training programmes in the AGI-modelled interventions therefore combined technical and soft skills training within the intervention approach.

A second aspect of these programmes is the effort to support women's empowerment and livelihoods as part of the labour market programme. For example, if women are disadvantaged by societal norms to access employment in higher-growth sectors, it might not be sufficient to provide them with soft or technical skills. A range of programmes therefore applied livelihood- and empowerment-related design features such as supporting awareness of and access to sexual and reproductive health, forming women's collectives and self-help groups, and building women's subjective empowerment and self-confidence (e.g. Adoho 2014; Chen 2017). As indicated above, this focus on supporting women's soft or life skills as part of training and job placement interventions is a more recent development, but has been suggested as a lesson learned by the Jovenes programmes in Latin America too (Honorati and McArdle 2013).

INTERVENTION DESIGN INCLUDES PARTICIPANT PROFILING AND TARGETING

The fifth identified design feature concerns the manner in which women are identified and chosen for the labour market programmes. All but one training and placement intervention applied some form of participant profiling and targeting as a part of the programme design. That is, all but one programme restricted access to the intervention to a specific type of women, predefined by the programme design. The most common participant profiles were: participant age, with a focus on younger women; participants' location and residence; existing skills levels, with a focus on functional numeracy and

literacy skills; level of schooling, with a preference for a high level; and existing level of employment, with a preference for out-of-work and/or out-of-school youth and young adults.

This design criterion in practice often led to an exclusion of the most marginalised women from the applied training and placement interventions. Adoho (2014), for example, acknowledged the application of this detailed participant profiling and targeting: 'The requirement reflected a deliberate choice on the part of program designers in the face of a trade-off between serving the most vulnerable and serving those who could most readily make use of this relatively short training program' (Adoho 2014: 8). Studies justified such detailed screening and profiling, with the intervention objective of enhancing employment in higher-growth sectors, which required a more advanced level of education and employee mobility.

In most interventions (n=7), the funder of the programme, that is national governments or multilateral organisations, decided on the criteria for participant profiling and targeting. The criteria were then applied by the implementing agency, in all cases a for-profit entity. Such design reinforces the motivation to include women with the highest likelihood of finding employment in the combined training and placement interventions, as the implementing agencies were paid by the employment results of participants. To determine participant eligibility, five out of seven interventions conducted assessment tests before enrolling participants in the training and placement programmes. All in all, participant profiling and targeting were therefore a near-universal practice and were regarded as a crucial design feature in order to enhance the likelihood of generating positive labour market effects.

CLEAR GOVERNANCE STRUCTURES FOR INTERVENTION PROVIDERS

As indicated in the participant profiling design feature, most training and placement interventions were conducted by private entities. This characteristic gave rise to another important design feature of the included programmes: the establishment of clear governance structures for intervention providers. Eight of out nine combined training and placement interventions established clear governance structures for intervention providers. In practice, these structures were operationalised through a range of mechanisms. First, intervention funders relied on competitive bidding processes in order to choose intervention providers (e.g. Chakravarty 2016; Galdo 2008). Second, six out of nine programmes followed a payment by results approach (e.g. Adoho 2014; de Azevedo 2013). That is, intervention providers were paid conditionally on their success in placing women into employment positions. A linked mechanism to this was employed by the Jovenes programmes (e.g. Galdo 2008; Ñopo 2007). Instead of paying providers by results, prospective implementation agencies of combined training and placement interventions had to provide proof of their ability achieve this. This could, for example, be prior agreements with private companies to hire a certain number of programme graduates. In both versions, this mechanism provides a strong governance tool to ensure that the objective of the programme is more likely to be met.

Third, a common governance tool concerns the development of provider quality assessment tools and their subsequent monitoring. These served to ensure the credibility and quality of the intervention

providers. In the Jovenes and AGI intervention models, for example, the quality of the skills training and the qualifications of the instructors were monitored and assessed. Selected training and placement intervention also require intervention providers to be certified and to comply with standardised quality assessments (e.g. Ñopo 2007). Other indicators of provider quality reported were: range of courses offered, quality and security of facilities, and past track records.

FLEXIBILITY AND RESPONSIVENESS IN INTERVENTION IMPLEMENTATION AND DESIGN

The last design feature reported to support intervention effects was the need for programme design and implementation to remain flexible and responsive to changing contexts. This referred to the set-up of detailed monitoring and evaluation (M&E) systems in order to respond rapidly to changes in labour market conditions or to address implementation failures. The design of pilot programmes too was referenced under this design feature (e.g. Adoho 2014; Groh 2012). An example of the benefits of embedding strong M&E systems into the programme design was provided as described above in reference to the change in gender-sensitivity design aspects in the AGEI in Nepal (Chakravarty 2016). Both, the Jovenes and AGI models that combined training and job placement interventions emphasised strong monitoring systems as well as rigorous programme evaluations.

In summary, the narrative synthesis identified seven distinct intervention design features that supported interventions to address women's barriers to wage labour participation in higher-growth/male-dominated sectors in LMICs. We regard these seven features as crucial design aspects in their own right, and the studies included in our systematic review commented on their importance. These seven features represent the active ingredients of women's labour market participation interventions included in our review.

However, our narrative synthesis does not allow us to unpack the exact configurations of interventions and their interplay. For example, the EPAG programme evaluated by Adoho (2014) comprised all seven design features and led to a 0.30 standard deviation change in employment outcomes for women. At the same time, the Barefoot Mechanics programme investigated by Chen (2017) only comprised three of the seven features but led to a similar change on wage labour outcomes (g=0.33). We therefore conducted a QCA in order to systematically examine potential configurations of design features and their association with intervention effects.

5.6.2 QCA RESULTS: CONFIGURATION OF INTERVENTION DESIGN FEATURES

As indicated in section 3.6, our approach to QCA followed the six-step method for QCA proposed by Thomas et al (2014), based on Rihoux and Ragin (2009). We therefore commenced with the construction of a data table to investigate the configuration of design features in the included interventions. In this,

we only used the homogeneous set of ten[35] interventions included in the meta-analysis[36] and also used in the narrative synthesis to develop the seven design features. Table 5.13 presents the data table of our QCA, indicating the relevant conditions (i.e. design features) and outcomes (i.e. effect sizes) for each individual study. Studies were exclusively allocated binary codes for the presence (or absence) of intervention features. In order to develop the outcome metric, we relied on the effect sizes calculated per intervention (sample size corrected SMD: Hedge's *g*), which express a change in wage labour outcomes for women. All but one intervention were found to be effective and we therefore defined the outcome for our QCA analyses in terms of membership (or non-membership) in the set of highly effective interventions. We next calibrated our wage labour effect sizes into a fuzzy set outcome for the purpose of the QCA. We opted for a fuzzy set over a crisp set to allow for degrees of membership within the set of highly effective interventions. This decision was informed by the variance in effect estimates and the heterogeneity in studies' risk of bias ranking, which were found to moderate intervention effects to some extent. The criteria for membership in the fuzzy set were as follows:

1. Full membership in the set of 'highly effective interventions': if SMD >0.25.
2. More in than out of the set: 0.15 < SMD ≤0.25.
3. More out than in the set: 0 < SMD ≤0.15.
4. Fully out of the set: SMD ≤0.

Following analysis of the data table, we developed the truth table which displays the conditions, configurations and number of studies with membership in each configuration set (Table 5.14). We used the fs/QCA software package[37] in order to construct the truth table (Rihoux and Ragin 2009). Given the seven identified conditions and ten cases identified, the truth table could potentially feature up to 128 possible different configurations of these conditions. This considered, only six of these 128 configurations are represented by the ten cases tested.

[35] Ten cases based on nine studies as de Azevedo evaluates two different interventions.
[36] We were unable to calculate an effect size for Ñopo, which is not included in the meta-analysis of SMD. Here, we use an approximation of the percentage reported in Ñopo converted into *g* for the purpose of the QCA, which does not aim to aggregate and average effect sizes and is thus less sensitive to the inclusion of imprecise effects, in particular when membership in outcome sets is fuzzy.
[37] http://www.u.arizona.edu/~cragin/fsQCA/

Table 5.13 QCA Data table

Study	Condition							Outcome	
	Labour market enhancing norms	Demand-led	Gender-sensitive	Soft/life skills	Participant profiling	Provider governance	Flexibility & iteration	Effect size (SMD)	Highly effective fuzzy set
Adoho	1	1	1	1	1	1	1	0.30	1
Chakravarty	1	1	1	1	1	1	1	0.17	0.66
Chen	1	1	1	0	0	0	0	0.33	1
de Azevedo T2	0	1	1	0	1	1	0	0.18	0.66
de Azevedo T1	0	1	1	1	1	1	0	0.29	1
Elías	0	1	1	0	1	1	0	0.15	0.66
Galdo	0	1	1	0	1	1	0	0.19	0.66
Groh	0	1	1	1	1	1	1	-0.03	0
Honorati	0	1	0	1	1	1	0	0.05	0.33
Ñopo	0	1	1	0	1	1	0	0.05	0.33

Table 5.14 Truth table for QCA on the 7 design features

Labour market enhancing norms	Demand-led	Gender-sensitive	Soft/life skill	Participant profiling	Provider governance	Flexibility & iteration	Number of studies	Member highly effective intervention set	Raw consistency
0	1	1	0	1	1	0	4	0	0.5775
1	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**0.83**
1	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**1**	**1.000**
0	1	1	1	1	1	1	1	0	0
0	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1.000**
0	1	0	1	1	1	0	1	0	0.33

Configurations with raw consistency >.75 are shown in bold.

The truth table represented in Table 5.14 thus only represents these six configurations of design features of interventions. It presents the six different configurations of design features (rows 1 to 6) before indicating (column 8) how many studies reported on each configuration. For example, we can see in the first row that four studies reported on the configuration of demand-led AND gender-sensitive AND participants profiling AND provider governance. The ninth column indicates whether this particular configuration is associated with membership in the highly effective intervention set (which is not the case for the four design feature configuration in row 1). The last column then indicates the raw consistency of each configuration of design features and the effects of interventions included in the configuration. In QCA, consistency constitutes a metric to express how far the pattern of all the cases is consistent with sufficiency. That is, the extent to which a particular configuration of conditions is consistently necessary for the observed outcome. In our QCA, we follow Thomas and colleagues (2014) and Rihoux and Ragin (2009) in setting a threshold for consistency of 0.75 or above.

Setting the threshold of consistency at 0.75 leaves only three configurations of design features that have sufficient raw consistency (rows 2, 3 and 5 in bold). Of these, two configurations present single studies and only one is based on two studies (row 2). As recommended by Thomas and colleagues (2014), we next assessed the quality and consistency of our truth table and the identified configurations of design features. Unfortunately, this assessment established a contradiction and pattern in the design feature configuration that negated a conclusion of the QCA. To explain, the configuration of design features in row 2 based on two studies meets our criterion for consistency. This configuration indicates that if all seven design features are present, interventions are likely to be highly effective. At the same time, however, row 3 indicates, based on a single study, that interventions are likely to be highly effective if only three design features are present. There is therefore a contradiction between both configurations and, given the small number of cases in the QCA, we cannot resolve or provide an alternative explanation for this. We double checked the coding of all three studies in question in the two configurations (Adoho 2014; Chakravarty 2016; Chen 2017) and are confident that the design features of the applied interventions were captured accurately. The strong association of all seven design features with highly effective interventions in particular challenges the conclusion of this model of the QCA, as this configuration is based on the largest number of studies.

Thomas and colleagues (2014: 13) recommend that if the truth table reveals areas of concern that:

> reviewers return to the conceptual framework that their review is based upon and consider again the dimensions upon which included studies might differ. This, in turn, will prompt a re-examination of the conditions to be used in the synthesis and possibly lead to the incorporation of new, or different, conditions.

We considered this option carefully but decided against it and decided not to conclude the QCA. The identification of our seven design features followed a rigorous and extensive research and engagement process. We first conducted a narrative synthesis of the included studies in order to identify a preliminary list of features before engaging a range of stakeholders to refine this list of features. We are therefore reluctant to further iterate on this list of design features only to force the conclusion of the

QCA in order to potentially establish a configuration of design features. Given the data available in our review, we would rather position the seven identified design features of women's labour market interventions as a key finding in their own right than iterate further on this list of features in the pursuit of a meaningful QCA. That is, we prefer to highlight all seven identified design features as of equal importance to the effective implementation of women's labour market participation interventions instead of furthering our quest for a particular configuration or combination of these seven design features.

5.6.3 SUMMARY OF DESIGN FEATURES OF INTERVENTIONS AIMING TO SUPPORT WOMEN'S WAGE LABOUR PARTICIPATION

Our systematic review applied narrative synthesis and QCA in order to identify the design features of interventions aiming to support women's wage labour participation in higher-growth and/or male-dominated sectors in LMICs. Given the heterogeneity of the included interventions in our review, we limit our investigation of intervention design features to a homogeneous sub-set of studies: combined training and job placement interventions. Zooming in on this sub-set, the narrative synthesis led to the identification of seven design features reported in the included studies:

1. Exposure to labour market participation enhancing social norms
2. Labour demand-led intervention design
3. Gender-sensitive intervention design
4. Provision of soft/life skills and social empowerment training
5. Participant profiling and targeting
6. Clear governance structures for intervention providers
7. Flexibility and responsiveness in intervention implementation and design.

Following the identification of these seven intervention design features associated with programme effectiveness, we further attempted to unpack the specific configurations of these design features and their correlation with programme effects. The subsequently attempted QCA, however, was inconclusive and we therefore cannot comment on specific configurations of design features. All in all, this leaves us to conclude that the seven individual intervention design features constitute the active ingredients of combined training and placement programmes to support wage labour participation in higher-growth and/or male-dominated sectors in LMICs. Further research is required to assess the specific configurations and combination of individual design features and their resultant effects.

6 DISCUSSION OF REVIEW FINDINGS

We next provide a summary and discussion of our systematic review findings on the effects of interventions aiming to support women's participation in wage labour in higher-growth/male-dominated economic sectors in LMICs. We first indicate the limitations of our systematic review before providing a summary of the main findings, which is then briefly contextualised within existing debates on labour market interventions in LMICs.

6.1 LIMITATIONS OF THE SYSTEMATIC REVIEW

Our systematic review is subject to two groups of limitations. First, limitations that arise from the nature and size of the identified evidence base, and second, limitations that stem from the review process itself. Both sets of limitations are discussed below.

6.1.1 OVERALL SIZE AND NATURE OF THE EVIDENCE BASE

Despite an extensive search of both academic and Grey literature sources, our systematic review only identified a small and heterogeneous evidence base of women's wage labour interventions in higher-growth/male-dominated sectors.[38] In total, we only included 19 studies, and the largest number of studies within a single intervention category was nine. On average, each intervention category only featured three studies. As a result, it is challenging to conduct an extensive synthesis of the effectiveness of interventions, as the evidence base is spread thinly across different interventions. Our most rigorous synthesis comprised the effects of nine homogeneous labour market interventions.

In addition, the identified evidence base was subject to a range of biases. Only five of the 19 studies were judged to be of a low risk of bias. The majority of studies was judged to be of high risk. Missing data and baseline confounding presented the largest sources of bias in the identified evidence base. Further shortcomings in the applied impact evaluation designs were sampling bias and departures from intended interventions (n=5).

In order to formally assess the overall quality of the evidence base, we apply the GRADE framework, which combined the risk of bias rating of the included studies with an assessment of the consistency, precision and directness of the included evidence base. Applying the GRADE framework, we established that the overall quality of the evidence included in our five syntheses was low (Appendix 9). Only the quality of evidence included in the meta-analysis on combined training and placement interventions was of moderate quality. All other syntheses were based on either low-quality (n=1) or very-low quality evidence (n=3). In summary, this small size and the heterogeneous, low-quality nature of the evidence

[38] As for the search results themselves, the applied search strategy was deliberately designed to be over-inclusive and was not limited to any particular labour market intervention. The results of our search, as well as the subsequent evidence map, can thus be used as a starting point for future reviews that aim to target different labour market interventions and women's employment in LMICs.

base limits the findings of our systematic review. We are only able to reach cautious conclusions regarding the evidence on combined training and placement interventions; for all other intervention categories, the size and quality of the evidence base does not allow us to reach conclusions.

It is interesting to contextualise this finding against the results of our evidence map, which included a broader range of interventions and outcomes associated with women's labour market participation in LMICs. Here, we identified a large body of impact evaluations (n=501) that examined the impact of different types of interventions on women's labour outcomes. Zooming in on the body of evidence only concerned with supporting women's wage labour then excluded the vast majority of these studies. For example, a large number of studies (n=241) were excluded as they did not focus on wage labour outcomes but on women's self-employment. The body of literature identified in the evidence map was thus focused more strongly on interventions supporting women to start or expand their own businesses (e.g. microfinance programmes) and focused to a lesser degree on women's participation in wage labour. Lastly, the evidence map also illustrated a large challenge in the reporting of impact evaluation results. A total of 111 studies did not report or disaggregate outcome data according to the sector of employment, and, crucially, according to the gender of participants. The latter in particular does limit a gendered investigation of labour market effects.

6.1.2 LIMITATIONS AND POTENTIAL BIASES IN THE SYSTEMATIC REVIEW

A second set of limitations relate to the design of our systematic review and the process of conducting it. In terms of review design, our systematic review subscribed to an aggregative review approach aiming to use statistical meta-analysis to synthesise the results of a homogeneous body of interventions and outcomes. To achieve this objective, the scope of the systematic review was narrow in terms of included study designs and the question of interest in the synthesis (i.e. what interventions work and to what effect). This narrow scope, for example, excludes qualitative evidence on contextual factors that might mitigate intervention effects and causal pathways; our review was limited to rigorous quantitative impact evaluations that reported numerical data on interventions' effects on women's wage labour participation. In addition, our review scope also focused narrowly on higher-growth economic sectors, which challenges the generalisability of the review findings to the wider labour market.

In terms of conducting the systematic review, we applied a range of quality assurance mechanisms in order to limit potential biases to the research process introduced by the review team. First, this review followed a two-stage approach that first involved the creation and publication of an evidence map on the review topic. This aimed to mitigate against defining a review scope and question that did not meet stakeholders' priorities. Second, the review design and process was a priori outlined in a detailed and peer-reviewed review protocol to ensure the transparency and replicability of the review (Langer et al 2017). Third, our systematic review was based on an exhaustive search effort of both the academic and Grey literature. The applied search strategy was conducted and reviewed by two information scientists, who developed, tested and applied our search strategy. Fourth, we applied a structured coding and risk of bias tool in order to assess the trustworthiness of the included studies and to extract relevant information for the synthesis in a transparent and consistent manner. To ensure the uniform application

of these tools, two reviewers independently coded and appraised a sub-set of the included studies. In the case of disagreement between reviewers, a third reviewer acted as a moderator to reach a final decision. Fifth, we assessed the quality of our meta-analyses using moderator and sensitivity analyses. These analyses established that our meta-analysis was based on a homogeneous set of labour market interventions and outcomes, justifying the pooling of intervention effects. For the narrative synthesis on the design features of labour market interventions, we conducted an extensive process of consultation and iteration to formulate the final list of features. Lastly, throughout the review process we had guidance from a multi-disciplinary advisory group.

All in all, we are therefore confident that we have applied reasonable measures to reduce the potential bias in the design and conduct of this review as far as possible.

6.2 OVERVIEW OF MAIN FINDINGS

Keeping in mind the above limitations, we present an overview of the main results of our systematic review on the effectiveness of interventions aiming to support women's wage labour participation in higher-growth/male-dominated sectors in LMICs.

Our systematic review identified a total of 19 impact evaluations that investigated the effects of women's wage labour participation. These interventions were heterogeneous in nature and split across five groupings of interventions. In addition, this small and heterogeneous evidence base was characterised by a low quality of evidence in terms of risk and bias ratings and the GRADE framework.

The interventions included in our systematic review are mainly conducted in low*er*-middle-income countries, spread evenly across all continental regions, and in countries that featured a lower than average economic participation of women. The interventions, by and large, focused on young women who were out of work at the start of the programme but who had a higher level of education and skills. Interventions deliberately targeted younger and more mobile women in order to match programme participants with the labour demands in higher-growth sectors, particularly the manufacturing industries, the business administration sector, engineering and ICT and electronics.

Based on this body of evidence, we conducted a range of syntheses on the effects of five different interventions to support women's wage labour participation:

1. *Effects of combined training and job placement interventions*

   The statistical meta-analysis of eight combined training and job placement interventions established a pooled positive effect size of 0.159 (0.09, 0.23), which translates into a 7.8% greater increase in wage employment for women taking part in the training and placement programmes as compared to a control group. These interventions were then further effective in increasing women's income, and our meta-analysis on income outcomes identified a positive pooled effect size of training and placement interventions on women's income of 0.145 (0.07,

0.22) based on nine studies. This effect size expressed a 7.2% greater increase of income for women taking part in the interventions as compared to a control group. Both meta-analyses were robust to a range of moderator and sensitivity analyses, including variables related to study design, intervention characteristics and settings, and characteristics of the partaking women. This review finding is based on **a moderate quality of evidence** and can therefore be interpreted to provide cautious support for the effectiveness of combined training and job placement interventions.

We furthermore investigated whether a change in women's wage labour employment and income translated into a change in women's empowerment. There is insufficient evidence to investigate the aggregate effectiveness of combined training and placement interventions on women's economic empowerment. On observation, we identified anecdotal evidence that combined training and placement interventions can support women's economic empowerment if the intervention is effective in also increasing wage employment and income outcomes.

In summary, the meta-analysis of the effects of interventions providing training combined with job placements established cautious evidence that these interventions can be effective to increase women's wage labour participation in higher-growth/male-dominated sectors, and their subsequent income. This finding is based on a moderate quality of evidence. In terms of women's empowerment, there is currently insufficient evidence to rigorously assess and synthesise intervention effects. Lastly, there is insufficient evidence on the cost-effectiveness of combined training and placement interventions.

2. *Effects of soft skills training to address vertical occupational segregation*

There is insufficient evidence to investigate the aggregate effectiveness of soft skills programmes aiming to address vertical occupational segregation in LMICs. On observation, our systematic review identified two studies in the garment sector that found that two individual training programmes were effective in increasing women's promotion to managerial posts and in changing male and female perceptions of women acting as supervisors. However, these results were based on a very small set of studies, which, in addition, were subject to moderate and high risks of bias. This finding is therefore based on **very low-quality evidence**, and the available evidence base only allows us to indicate the use of soft skill programmes to address vertical occupational segregation as a relevant concept which requires further investigation.

3. *Effects of job placement services only*

There is insufficient evidence to investigate the aggregate effectiveness of job placement services as the sole labour market programme component. On observation, our systematic review identified three studies that reported that the effects of diverse job placement programmes were mixed. This included recruitment services, a screening and matching programme and a graduate wage voucher programme. The narrative synthesis of these three labour market programmes as the sole programme component is challenged by a large degree

of heterogeneity. The interventions as well as their effects are highly heterogeneous, and we therefore cannot comment on the aggregate effectiveness of job placement programmes. There is contradictory evidence of effects within and across different outcomes. This finding is based on **low-quality evidence**.

4. *Effects of national labour subsidies*

   There is insufficient evidence to investigate the aggregate effectiveness of national labour subsidies to support women's wage labour employment in higher-growth sectors in LMICs. On observation, our systematic review identified two studies that found that two national subsidy policies led to positive effects of women's wage labour in higher-growth sectors in the short term. However, these results were based on a very small set of studies, which, in addition, were subject to a high risk of bias. We are therefore unable to comment on the overall effectiveness of national labour subsidies in the context of women's wage labour employment. This finding is based on **very low-quality evidence**.

5. *Effects of macro-level policies targeting women's empowerment*

   There is insufficient evidence to investigate the aggregate effectiveness of macro-level women's empowerment interventions on women's wage labour outcomes. On observation, our systematic review identified two studies that found mixed wage labour outcomes from two macro-level policies, but positive long-term effects on women's empowerment, such as increased aspirations and economic participation. However, the small size and low quality of the evidence base does not allow us to comment on the overall effectiveness of these interventions. This finding is based on **very low-quality evidence**.

In addition, our systematic review also investigated the type of intervention design features that supported intervention effects on women's wage labour participation in higher-growth/male-dominated sectors. Given the heterogeneity of the included interventions in our review, we limited our investigation of intervention design features to a homogeneous sub-set of studies: combined training and job placement interventions. Zooming in on this sub-set, the narrative synthesis led to the identification of seven design features reported in the included studies:

1. Exposure to labour market participation enhancing social norms
2. Labour demand-led intervention design
3. Gender-sensitive intervention design
4. Provision of soft/life skills and social empowerment training
5. Participant profiling and targeting
6. Clear governance structures for intervention providers
7. Flexibility and responsiveness in intervention implementation and design.

Following the identification of these seven intervention design features associated with programme

effectiveness, we attempted to unpack their specific configurations and their correlation with programme effects. The subsequently attempted QCA was inconclusive, however, and we therefore cannot comment on specific configurations of design features.

All in all, this leaves us to conclude that the seven individual intervention design features constitute the active ingredients of combined training and placement programmes to support wage labour participation in higher-growth/male-dominated sectors in LMICs.

6.3 DISCUSSION OF MAIN FINDINGS

The findings of our review overlap with two existing reviews on labour market interventions in LMICs. Tripney and colleagues (2013) as well as Kluve and colleagues (2016) both conducted systematic reviews on TVET training in LMICs. Tripney's review and meta-analysis established a positive pooled effect size of TVET interventions on both employment and income outcomes for youth in LMICs. This pooled effect of 0.134 SMD (0.02, 0.24) is close to the effects identified in our review. The review concluded that TVET interventions had some promise as an approach to labour market interventions in LMICs but cautioned that the evaluation designs used to assess the interventions challenged the attribution of causality. Kluve's review came to similar conclusions. Conducting a global systematic review of training, entrepreneurship promotion, employment services and subsidised employment interventions which included a total of 113 counterfactual-based impact evaluations, Kluve and colleagues found the intervention effects to be particularly significant in LMICs. The review's meta-analysis identified effect sizes in the range of 0.06–0.18 SMD on employment and income outcomes, which overlap with the pooled effect sizes identified in our review. The review concluded that investment in youth labour market programmes had the potential to increase human capital and employment prospects in the long term. Lastly, Kluve's review also identified a range of design features of labour market interventions that overlap with the results of our review, including: participant profiling, participant engagement mechanisms and incentives for service providers.

There are also a number of reviews of broader economic interventions for women in LMICs: Buvinic and O'Donnell (2016), Peters and colleagues (2016) and ODI (2016). In *Revisiting What Works: Women, Economic Empowerment and Smart Design*, Buvinic and O'Donnell (2016) updated the results of the UN *Roadmap to Women's Economic Empowerment* report (UN 2013). With reference to wage labour interventions, they rated demand-driven job services as 'proven' for increasing the economic opportunities of young women, provided that the programmes levelled the playing field for women in training and working environments (Buvinic and O'Donnell 2016: 21). While our review agrees with the general direction of effects for demand-driven job services, in the context of higher-growth/male-dominated sectors, the effectiveness of the interventions cannot be regarded as proven yet. Buvinic and O'Donnell (2016) further suggested a range of design innovations in the implementation of economic programmes for women, many of which resonate with the findings of our synthesis on design features, such as: de-biasing service provision, nurturing subjective economic empowerment and designing for

cognitive and social determinants of economic behaviour as well as the pervasive gender biases embedded in organisations and working environments.

Our systematic review also overlaps with a body of literature more critical of wage labour interventions as an effective approach for social and economic development in LMICs. Based on the perceived low returns and subsequent cost-ineffectiveness of many wage labour intervention, skills training in particular, Blattman and Ralston (2015), McKenzie (2017) and Honorati and McArdle (2013), among other, argued that the focus of interventions should shift towards supporting entrepreneurship and capital-centric employment programmes such as start-up grants, cash infusions and in-kind capital transfers. Blattman and Ralston, for example, argued that:

> Skills training and microfinance have shown little impact on poverty or stability, especially relative to program cost. In contrast, injections of capital – cash, capital goods, or livestock – seem to stimulate self-employment and raise long term earning potential, often when partnered with low-cost complementary interventions (2015: ii).

While our review is not formally concerned with comparing the effects of wage labour interventions against self-employment interventions (which were excluded following stakeholder engagement on the evidence map), a number of identified impact evaluations did compare these different programme designs against each other. These are Blattman and Dercon (2016),[39] Adoho (2014), Chakravarty (2016) and Honorati (2015) and there thus seems to be an increasing body of evidence on this research question that could be used for further investigation.

A final body of research relevant to our systematic review relates to an emerging body of knowledge on 'cross-overs' of women into higher-growth sectors (e.g. Alibhai et al 2015; Campos et al 2015). This body of research investigated the characteristics of women who had already crossed over into higher-growth/male-dominated sectors rather than investigating the types of programmes that work best to facilitate such cross-overs. However, there were large synergies between the experimental impact evaluation evidence and the mixed-methods work on cross-overs. For example, Campos and colleagues (2015) identified psychological factors, information on cross-overs' rates of return and the influence of male role models and exposure to positive social norm as key drivers of women's search for employment in male-dominated sectors. Among these, there was a large overlap with the identified design features in our systematic review which were derived from the experimental impact evaluation evidence. It therefore seems sensible for further studies to explore the overlap between both bodies of literature in more detail.

The discussion above has attempted to outline a few of the potential synergies of our systematic review findings and the wider literature on women's labour market participation in higher-growth/male-

[39] Not included in the systematic review.

dominated economic sectors. In terms of the review findings themselves, the following implications for further research and practice were mentioned in the report and are summarised below.

6.3.1 IMPLICATIONS FOR DECISION MAKERS

- At this stage the overall evidence base on interventions supporting women's wage labour participation in higher-growth/male-dominated sectors is limited. The evidence base does not provide a clear picture of interventions that could be recommended for scale-up.
- Of all the reviewed interventions, training programmes combined with placement services show the most promise. The best evidence in this regard comes from two intervention models: the World Bank's AGI and the Jovenes approach. These two intervention models for combining training and placement services come closest to a design template for promising interventions.
- In terms of more granular intervention design implications, seven promising design attributes were identified, as listed in section 6.2.
- In general, the evidence map indicated that a focus on interventions supporting formal wage employment is lagging behind efforts to promote self-employment, e.g. microfinance, as a pathway to poverty reduction.

6.3.2 IMPLICATIONS FOR FUTURE RESEARCH

- There is a strong need to improve the reporting of labour market intervention impact evaluations in terms of disaggregation of outcome data by gender and sector of employment.
- Future research should also collect more detailed data on intervention costs and conduct formal cost-effectiveness analyses of the evaluated interventions.
- More research is required to compare the effects of wage labour interventions against self-employment interventions. At this stage, research on interventions supporting self-employment far outweighs research on interventions supporting wage employment.
- Existing research on cross-overs of women into higher-growth/male-dominated sectors could be used to inform future programme and impact evaluation design. This includes additional research on vertical rather horizontal occupational segregation.
- In terms of research on intervention design, an enhanced integration of research on social norms and behavioural insights within the more economic body of literature on labour market interventions would be of benefit.
- Given its promising findings in this systematic review, a more in-depth investigation into the overall effects and design lessons learned from the AGI portfolio of work on women's labour market participation would be a fruitful area for future research.

6.4 AUTHORS' CONCLUSIONS

This systematic review provided an evidence map and structured synthesis on the effects of interventions aiming to support women's participation in wage labour in higher-growth/male-

dominated sectors in LMICs. We identified a small evidence base that was heterogeneous in terms of the applied labour market interventions and of low quality in terms of the methodological trustworthiness of the studies and consistency of effects. Using statistical meta-analysis, we found that combined training and placement interventions increased women's wage labour participation by 7.8% and women's income by 7.2% compared to women not receiving these interventions. There is insufficient evidence to investigate the overall effect of combined training and placement interventions on women's economic empowerment.

There is currently a lack of evidence to comment on the effectiveness of the following interventions to support women's participation in wage labour in higher-growth/male-dominated sectors in LMICs: soft skills training to address vertical occupational segregation; job placement only interventions; national labour subsidies; and macro-level women's empowerment policies.

Our systematic review identified the following seven intervention design features of women's wage labour intervention in LMICs that can be described as the active ingredients of interventions supporting positive outcomes in women's wage employment: (1) exposure to labour market participation enhancing social norms; (2) labour demand-led intervention design; (3) gender-sensitive intervention design; (4) provision of soft/life skills and social empowerment training; (5) participant profiling and targeting; (6) clear governance structures for intervention providers; (7) flexibility and responsiveness in intervention implementation and design.

7 REFERENCES

Alibhai S, Buehren N, Papineni S (2015). *Female entrepreneurs who succeed in male-dominated sectors in Ethiopia*. Washington DC: World Bank. http://documents.worldbank.org/curated/en/2015/10/25157866/female-entrepreneurs-succeed-male-dominated-sectors-ethiopia

Blattman C, Dercon S (2016). *Occupational choice in early industrializing societies: Experimental evidence on the income and health effects of industrial and entrepreneurial work* (No. w22683). Cambridge, MA: National Bureau of Economic Research.

Blattman C, Ralston L (2015). *Generating employment in poor and fragile states: Evidence from labor market and entrepreneurship programs*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2622220

Bohnet I (2016). *What works: Gender equality by design.* Cambridge: Belknap Press.

Borenstein M, Hedges, LV, Higgins JPT, Rothstein HR (2009). *Introduction to meta-analysis*. Chichester: John Wiley and Sons.

Borrowman M, Klasen S (2017). *Drivers of gendered sectoral and occupational segregation in developing countries*. Courant Research Centre: Poverty, Equity and Growth Discussion Papers, No. 222. Göttingen: Georg-August-Universität Göttingen.

Brody C, De Hoop T, Vojtkova M, Warnock R, Dunbar M, Murthy P, Dworkin S (2015). Economic self-help group programs for improving women's empowerment: A systematic review. *Campbell Systematic Reviews*, 2015: 19.

Brunton, G., O'Mara-Eves, A. and Thomas, J., 2014. The 'active ingredients' for successful community engagement with disadvantaged expectant and new mothers: a qualitative comparative analysis. *Journal of advanced nursing*, *70*(12), pp.2847-2860.

Buvinic M, O'Donnell M (2016). *Revisiting what works: Women, economic empowerment, and smart design*. Washington: Center for Global Development. http://www.cgdev.org/sites/default/files/CGD-Roadmap-Update-2016.pdf

Campos F, Goldstein MP, McGorman L, Munoz Boudet AM, Pimhidzai O (2015). *Breaking the metal ceiling: Female entrepreneurs who succeed in male-dominated sectors*. Washington DC: World Bank. http://documents.worldbank.org/curated/en/753711467997247654/Breaking-the-metal-ceiling-female-entrepreneurs-who-succeed-in-male-dominated-sectors

Chinen M, de Hoop T, Balarin M, Alcázar L (2016). *Vocational and business training to increase women's participation in higher skilled occupations in low- and middle-income countries: protocol for a systematic review*. Campbell Collaboration. https://www.campbellcollaboration.org/library/women-training-participation-in-higher-valued-occupations.html

Cho Y, Honorati M (2014). Entrepreneurship programs in developing countries: A meta regression analysis. *Labour Economics*, 28: 110-130.

DPME (2016) *Policy-relevant evidence maps: A departmental guidance note*. Pretoria: Department of Planning, Monitoring and Evaluation.

Duval S, Tweedie R (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2): 455-463.

Gallup, ILO (2017). *Towards a better future for women and work: Voices of women and men*. Geneva: ILO. http://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_546256.pdf

Gibbs A, Willan S, Misselhorn A, Mangoma J (2012). Combined structural interventions for gender equality and livelihood security: A critical review of the evidence from southern and eastern Africa and the implications for young people. *Journal of the International AIDS Society*, 15(3, Suppl. 1): 1-10.

Gopalaswamy AK, Babu MS, Dash U (2016). *Systematic review of quantitative evidence on the impact of microfinance on the poor in South Asia*. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Gough, D., Oliver, S. and Thomas, J. eds., 2012. *An introduction to systematic reviews*. Sage.

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ: British Medical Journal*, 336(7650): 924.

Guyatt GH, Oxman A, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, deBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ (2011). GRADE guidelines: 1. Introduction to GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64: 383-394.

Hedges L (2007). Effect sizes in cluster-randomised designs. *Journal of Educational and Behavioural Statistics*, 32: 341-370.

Hedges L, Olkin I (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Higgins JPT, Thompson SG (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21: 1539-1558.

Higgins JPT, Green S (eds) (2011). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley-Blackwell.

Honorati M, McArdle TP (2013). *The nuts and bolts of designing and implementing training programs in developing countries*. Washington DC: World Bank. https://openknowledge.worldbank.org/handle/10986/16101

Ibanez M, Vollmer S, Khan S, Minasyan A, Sahoo S, Pasha A, Balasubramanian P, Le Thi Ngoc Tu, Straube S, Friede T (2016). *Title: Interventions to promote female economic empowerment in low- and middle-income countries: A systematic review of reforms to labor market, land, inheritance and participation of women in decision making*. Campbell Collaboration. https://www.campbellcollaboration.org/library/women-economic-empowerment-low-and-middle-income-countries.html

IGC (2016). *Emerging opportunities for women in Khyber Pakhtunkhwa: Industry (Manufacturing and Construction), Mining, Tourism, and Agriculture*. Oxford: International Growth Centre.

ILO (2016). *Women at work*. Trends 2016. Geneva: ILO.

Kabeer N (2012). *Women's economic empowerment and inclusive growth: Labour markets and enterprise development*. SIG Working Paper 2012/1. Ottawa: IDRC. https://www.idrc.ca/sites/default/files/sp/Documents%20EN/NK-WEE-Concept-Paper.pdf

Kahneman D (2011). *Thinking, fast and slow*. London: Macmillan.

Keef SP, Roberts LA (2004). The meta-analysis of partial effect sizes. *British Journal of Mathematical and Statistical Psychology*, 57(1): 97-129.

Kluve J, Puerto S, Robalino DA, Romero JM, Rother F, Stöterau J, Weidenkaff F, Witte M (2016). *Do youth employment programs improve labor market outcomes? A systematic review*. IZA Discussion Paper No. 10263. http://ftp.iza.org/dp10263.pdf

Langer L, Tripney J, Erasmus Y, Tannous N, Chisoro C, Opondo M, Zigana L, Obuku E, van Rooyen C, Stewart R (2017) *Women in wage labour: A systematic review of the effectiveness and design features of interventions supporting women's participation in wage labour in higher growth and/or male-dominated sectors in LMICs.* A systematic review protocol. London: EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London.

Lipsey MW, Wilson D (2001). *Practical meta-analysis*. London: Sage.

McKenzie D (2017). *How effective are active labor market policies in developing countries*. Policy Research Working Paper No. 8011. Washington, DC: World Bank.

McKenzie, D. and Woodruff, C., 2013. What are we learning from business training and entrepreneurship evaluations around the developing world? *The World Bank Research Observer*, 29(1): 48–82.

McKinsey Global Institute (2015). The power of parity: How advancing women's equality can add $12 trillion to global growth. New York: McKinsey. http://www.mckinsey.com/insights/growth/how_advancing_womens_equality_can_add_12_trillion_to_global_growth

ODI (2016). *Women's economic empowerment: Navigating enablers and constraints*. London: Overseas Development Institute.

Peters HE, Astone NM, Malik AA, Maret F, Heller C (2016). *Women's economic empowerment: A review of evidence on enablers and barriers*. Washington, DC: Urban Institute. http://www.urban.org/research/publication/womens-economic-empowerment-review-evidence-enablers-and-barriers

Petrosino, A. (2012). Interventions in Developing Nations for Improving Primary andSecondary School Enrollment of Children: A Systematic Review.https://doi.org/10.4073/csr.2012.19

Piza C, Cravo T, Taylor L, Gonzalez L, Musse I, Furtado I, Sierra AC, Abdelnour S (2016). *Business support for small and medium enterprises in low- and middle-income countries: A systematic review*. 3ie Systematic Review 25. London: International Initiative for Impact Evaluation (3ie).

Revenga A, Shetty S (2012). Empowering women is smart economics. *Finance & Development*, 49(1): 40-43.

Rihoux B, Ragin C (2009) *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. London: Sage.

Rosenthal, R. and Rubin, D.B., 1982. A simple, general purpose display of magnitude of experimental effect. *Journal of educational psychology*, *74*(2), p.166.

Sen A (1999). *Development as freedom*. Oxford, UK: Oxford University Press.

Snilstveit B (2012). Systematic reviews: From 'bare bones' reviews to policy relevance. *Journal of Development Effectiveness*, 4(3): 388-408.

Snilstveit B, Stevenson J, Menon R, Phillips D, Gallagher E, Geleen M, Jobse H, Schmidt T, Jimenez E (2016). *The impact of education programmes on learning and school participation in low-and middle-income countries*. 3ie Systematic Review Summary 7. London: International Initiative for Impact Evaluation (3ie).

Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355: 4919.

Stewart R, van Rooyen C, Korth M, Chereni A, Rebelo Da Silva N, de Wet T (2012). *Do micro-credit, micro-savings and micro-leasing serve as effective financial inclusion interventions enabling poor people, and especially women, to engage in meaningful economic opportunities in low- and middle-income countries: A systematic review of the evidence*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Stewart, R., Langer, L., Da Silva, N.R., Muchiri, E., Zaranyika, H., Erasmus, Y., Randall, N., Rafferty, S., Korth, M., Madinga, N. and de Wet, T., 2015. The Effects of Training, Innovation and New Technology on African Smallholder Farmers' Wealth and Food Security: A Systematic Review. *Campbell Systematic Reviews*, *11*(16).

Taylor G, Pereznieto P (2014) *Review of evaluation approaches and methods used by interventions on women and girls' economic empowerment*. London: ODI. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/8843.pdf

Thaler RH, Sunstein CR (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.

Thomas J, Brunton J, Graziosi S (2010). *EPPI-Reviewer 4.0: Software for research synthesis*. London: Social Science Research Unit, Institute of Education, University of London.

Thomas J, O'Mara-Eves A, Brunton G (2014). Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: A worked example. *Systematic Reviews*, 20(3): 67.

Tripney J, Hombrados J, Newman M, Hovish K, Brown C, Steinka-Fry K, Wilkey E (2013). Technical and vocational education and training (TVET) interventions to improve the employability and employment of young people in low- and middle-income countries: A systematic review. *Campbell Systematic Reviews*, 2013: 9.

Tripney J, Roulstone A, Vigurs C, Hogrebe N, Schmidt E, Stewart R (2015) Interventions to improve the labour market situation of adults with physical and/or sensory disabilities in low- and middle-income countries. *Campbell Systematic Reviews*, 2015: 20.

UN (2013). *A roadmap for promoting women's economic empowerment*. New York: United Nations. http://www.womeneconroadmap.org/sites/default/files/WEE_Roadmap_Report_Final.pdf

UN (2015). *Progress of the world's women 2015-2016: Transforming economies, realizing rights*. New York: United Nations Women. http://progress.unwomen.org/en/2015/pdf/UNW_progressreport.pdf

Vaessen J, Rivas A, Duvendack M, Palmer-Jones R, Leeuw FL, van Gils G, Lukach R, Holvoet N, Bastiaensen J, Hombrados JG, Waddington H (2014). The effect of microcredit on women's control over household spending in developing countries: A systematic review. *Campbell Systematic Reviews*, 10(8). https://www.campbellcollaboration.org/library/microcredit-effect-on-women-household-spending.html

World Bank (2012). *World development report 2012: Gender equality and development*. Washington DC: World Bank.

World Economic Forum (2017). *The global gender gap report 2017*. Geneva: World Economic Forum. http://www3.weforum.org/docs/WEF_GGGR_2017.pdf

Yoong J, Rabinovich L, Diepeveen S (2012). *The impact of economic resource transfers to women versus men: A systematic review*. Technical report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

APPENDICES

APPENDIX 1: AUTHORSHIP OF THIS REPORT

DETAILS OF ADVISORY GROUP MEMBERSHIP

Harsha Dayal	Research Director: South African National Department of Planning Monitoring and Evaluation
Thomas De Hoop	Senior Researcher: American Institute for Research
Sumana Hussain	Economic Advisor: UK Department for International Development
Marcel Korth	Senior lecturer: Witwatersrand University Senior consultant to the United Nations Office on Drugs and Crime on gender-based violence
Bernedette Muthien	Commissioner and Chair of Research & Policy Development Statutory Committee Constitutional Commission for Cultural, Religious & Language Rights South Africa
Josephilda Nhlapo-Hlope	Outcome Facilitator Social Cohesion: South African National Department of Planning Monitoring and Evaluation
Jenny Schreiner (interim)	Director-General: South African National Department of Women

DETAILS OF REVIEW GROUP MEMBERSHIP

Laurenz Langer	Africa Centre for Evidence, University of Johannesburg
Yvonne Erasmus	Africa Centre for Evidence, University of Johannesburg
Natalie Tannous	Africa Centre for Evidence, University of Johannesburg
Ekwaro Obuku	African Centre for Systematic Reviews and Knowledge Translation, Makerere University
Zafeer Ravat	Africa Centre for Evidence, University of Johannesburg
Charity Chisoro	Africa Centre for Evidence, University of Johannesburg
Mary Opondo	Africa Centre for Evidence, University of Johannesburg
Promise Nduku	Africa Centre for Evidence, University of Johannesburg

Janice Tripney	EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London
Carina van Rooyen	Africa Centre for Evidence, University of Johannesburg
Ruth Stewart	Africa Centre for Evidence, University of Johannesburg; EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London

ACKNOWLEDGEMENTS

We would like to thank our review Advisory Group – Harsha Dayal, Thomas De Hoop, Sumana Hussain, Marcel Korth, Bernedette Muthien, Josephilda Nhlapo-Hlope and Jenny Schreiner – for their insightful comments and inputs on the review. Sincere appreciation is also expressed to Alison Bullen for leading and conducting the scientific search for academic literature. We are also grateful for the helpful feedback and support received from members of the EPPI-Centre: Mukdarut Bangpan, Sandy Oliver, Claire Stansfield, Dylan Kneale, Jeff Brunton, Sergio Graziosi and James Thomas.

CONFLICTS OF INTEREST

There were no conflicts of interest in the writing of this report.

APPENDIX 2: SUMMARY OF INCLUSION AND EXCLUSION CRITERIA

CRITERIA USED FOR THE EVIDENCE MAP

Inclusion Criteria	Exclusion Criteria
Population	
Women: The study sample must include women aged 15 years and older. The study sample must either be majority female or the study results must be disaggregated by gender. *Geographical location:* Low- or middle-income country as classified by the World Bank at time of data collection for the study.	*Women:* The study sample is not majority female or the study results are not disaggregated by gender. *Geographical location:* High-income country as classified by the World Bank at time of data collection for the study.
Employment Setting	
The study must evaluate the effects of relevant interventions applied in economic sectors with high or growing productivity *and/or* which are male-dominated. This refers to: • Commercial agriculture • Energy (mining and quarrying, electricity, gas and water supply) • Trade • Transportation • Accommodation and food • Business administration services • Finance • Electronics and ICT • Maritime services • Wood pulp and forestry • Construction • Manufacturing • Higher education/Science and Technology. Studies assessing vertical labour market segregation are included regardless of the economic sector.	All other economic sectors are not relevant for inclusion unless the research relates to vertical market segregation.
Intervention	
We will include *any* intervention that aims to overcome the barriers to women's labour market participation in LMICs. Section 1.5 outlined the main categories of interventions that we expect to encounter in this review as well as provided examples of interventions for each category:	We will exclude macro-level interventions such as investment in basic to tertiary education, health care, citizenship, social welfare, and economic growth which are known to benefit labour market participation rates of the general population.

(1) Interventions overcoming discrimination by markets & work institutions (2) Interventions overcoming constraints in access to credit, finance and assets (3) Interventions overcoming constraints in employability or entrepreneurship (4) Interventions overcoming a lack of social capital & norms (5) Interventions overcoming behavioural factors	
Outcomes	
To be included in the evidence map, studies must evaluate the impact of interventions on one of these three final outcomes: (1) Participation in formal or informal employment (in higher growth and/or male-dominated sectors) (2) Entrepreneurial success (following outcome 1: participation in higher growth and/or male-dominated sectors) (3) Economic empowerment (following outcome 1: participation in higher growth and/or male-dominated sectors)	All other outcomes are not eligible for inclusion, including intermediate outcomes reported in studies that do not assess one of the final outcomes.
Methods	
We will include studies using either of the following quantitative experimental or quasi-experimental study designs: (a) Designs using a random or quasi-random method of group assignment in which one of the following is true: • Units (individuals or clusters of individuals) are randomly assigned to treatment and control groups by the investigator using a fully random procedure, such as computerised random number generation; • A quasi-random procedure presumed to produce comparable groups has been used, for example, allocation by date of birth or next person to walk in the door (i.e. the method of allocation falls short of full randomisation); • Regression discontinuity designs in which participants are assigned by the investigator to intervention or control groups solely on the basis of a cut-off score on a pre-programme measure. (b) Designs employing non-random methods of assignment, in which one of the following is true:	We will exclude all other types of study designs, for example, evaluations designs such as those without a control group or without multiple data points.

• The investigator controls group exposure and assigns participants using a non-random procedure (e.g. alphabetically by surname); • The investigator constructs the comparison group after the start of the intervention (e.g. by exploiting existing survey data); • A natural experiment in which units exposed to the treatment and control conditions are determined by nature (e.g. change in policy or divergence in practice between regions) or by other factors outside the control of the investigators); • Assignment to conditions (treatment versus comparison) is by means of self-selection by participants or by administrator selection (e.g. by welfare officials).	

CRITERIA USED FOR THE FULL SYSTEMATIC REVIEW:

Inclusion Criteria	Exclusion Criteria
Population	
Women: The study sample must include women aged 15 years or older. The study sample must either be majority female or the study results must be disaggregated by gender. *Geographical location:* Low- or middle-income country as classified by the World Bank at the time of data collection for the study.	*Women:* The study sample is not majority female or the study results are not disaggregated by gender. *Geographical location:* High-income country as classified by the World Bank at time of data collection for the study.
Employment Setting	
The study must evaluate the effects of relevant interventions applied in economic sectors with high or growing productivity *and/or* which are male-dominated. This refers to: • Commercial agriculture • Energy (mining and quarrying, electricity, gas and water supply) • Trade • Transportation • Accommodation and food • Business administration services • Finance • Electronics and ICT	All other economic sectors are not relevant for inclusion unless the study is investigating vertical labour market segregation.

• Maritime services • Wood pulp and forestry • Construction • Manufacturing • Higher education/Science and Technology Studies assessing vertical labour market segregation are included regardless of the economic sector.	
Intervention	
We will include *any* intervention that aims to overcome the barriers to women's labour market participation in LMICs. Section 1.5 outlined the main categories of interventions that we expect to encounter in this review as well as provided examples of interventions for each category: (1) Interventions overcoming discrimination by markets & work institutions (2) Interventions overcoming constraints in access to credit, finance, and assets (3) Interventions overcoming constraints in employability (4) Interventions overcoming a lack of social capital & norms (5) Interventions overcoming behavioural factors	We will exclude macro-level interventions such as investment in basic to tertiary education, health care, citizenship, social welfare, and economic growth which are known to benefit labour market participation rates of the general population.
Outcomes	
To be included in the full systematic review, studies must evaluate the impact of interventions on the final outcome: Participation in formal or informal employment (in higher growth and/or male-dominated sectors)	All other outcomes are not eligible for inclusion, including intermediate outcomes reported in studies that do not assess the final outcome.
Methods	
We will include studies using either of the following quantitative experimental or quasi-experimental study designs: (a) Designs using a random or quasi-random method of group assignment in which one of the following is true: • Units (individuals or clusters of individuals) are randomly assigned to treatment and control groups by the investigator using a fully random procedure, such as computerised random number generation;	We will exclude all other types of study designs, for example, evaluations designs such as those without a control group or without multiple data points.

• A quasi-random procedure presumed to produce comparable groups has been used by the investigators. For example, allocation by date of birth or next person to walk in the door (i.e. the method of allocation falls short of full randomisation); • Regression discontinuity designs in which participants are assigned by the investigator to intervention or control groups solely on the basis of a cut-off score on a pre-programme measure. (b) Designs employing non-random methods of assignment, in which one of the following is true: • The investigator controls group exposure and assigns participants using a non-random procedure (e.g. alphabetically by surname); • The investigator constructs the comparison group after the start of the intervention (e.g. by exploiting existing survey data); • A natural experiment in which units exposed to the treatment and control conditions are determined by nature (e.g. change in policy or divergence in practice between regions) or by other factors outside the control of the investigators); • Assignment to conditions (treatment versus comparison) is by means of self-selection by participants or by administrator selection (e.g. by welfare officials).	

APPENDIX 3: WORLD BANK CLASSIFICATION OF WORLD ECONOMIES

Region	Low-income economies	Lower-middle income economies	Upper-middle income economies
Europe and Central Asia		Armenia, Kosovo, Kyrgyz Republic, Moldova, Tajikistan, Ukraine, Uzbekistan	Albania, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Georgia, Kazakhstan, Macedonia FYR, Montenegro, Romania, Russian Federation, Serbia, Turkey, Turkmenistan
South Asia	Afghanistan, Nepal	Bangladesh, Bhutan, India, Pakistan, Sri Lanka	Maldives
Middle East and North Africa		Djibouti, Egypt, Morocco, Syrian Arab Republic, Tunisia, West Bank and Gaza, Yemen	Algeria, Iran, Iraq, Jordan, Lebanon, Libya
East Asia and Pacific	Democratic Republic of Korea	Cambodia, Indonesia, Kiribati, Lao PDR, Micronesia, Mongolia, Myanmar, Papua New Guinea, Philippines, Samoa, Solomon Islands, Timor-Leste, Tonga, Vanuatu, Vietnam	American Samoa, China, Fiji, Malaysia, Marshall Islands, Palau, Thailand, Tuvalu
Sub-Saharan Africa	Benin, Burkina Faso, Burundi, Central African Republic, Chad, Comoros, Democratic Republic of Congo, Eritrea, Ethiopia, The Gambia, Guinea, Guinea-Bissau, Liberia, Madagascar, Malawi, Mali, Mozambique, Niger, Rwanda, Senegal, Sierra Leone, Somalia, South Sudan, Tanzania, Togo, Uganda, Zimbabwe	Cameroon, Cape Verde, Republic of Congo, Côte d'Ivoire (Ivory Coast), Ghana, Kenya, Lesotho, Mauritania, Nigeria, São Tomé and Principe, Sudan, Swaziland, Zambia	Angola, Botswana, Equatorial Guinea, Gabon, Mauritius, Namibia, South Africa

Region	Low-income economies	Lower-middle income economies	Upper-middle income economies
Latin America and Caribbean	Haiti	Bolivia, El Salvador, Guatemala, Honduras, Nicaragua	Argentina, Belize, Brazil, Chile,[40] Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, Grenada, Guyana, Jamaica, Mexico, Panama, Paraguay, Peru, St. Lucia, St. Vincent and the Grenadines, Suriname, Uruguay,[41] Venezuela

As of 1 July 2016, https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

[40] Before 1 July 2013
[41] Before 1 July 2013

APPENDIX 4: SEARCH STRATEGY

ACADEMIC DATABASES TO BE SEARCHED USING THE MASTER SEARCH STRING:

- Web of Science
- Econlit (EBSCO)
- ERIC (EBSCO)
- Business Source Complete (EBSCO)
- PRISMA database
- Sociological Abstracts (CSA)

KEY CONCEPTS

1. Developing Countries
2. Women
3. Type of study
4. Intervention

Search 1 and 2 and 3 and 4 in Title, Abstract, Keyword, Subject Heading.

Example of search string applied and piloted in Web of Science:

1 AND 2 AND 3 AND 4 = 8,372 citations (search ran 11 January 2017)

1. SEARCH FOR 'DEVELOPING COUNTRIES' TERMS

Africa OR Asia OR Caribbean OR "West Indies" OR "South America" OR "Latin America" OR "Central America" or Afghanistan OR Albania OR Algeria OR Angola OR Antigua OR Barbuda OR Argentina OR Armenia OR Aruba OR Azerbaijan OR Bahrain OR Bangladesh OR Barbados OR Benin OR Byelarus OR Byelorussian OR Belarus OR Belorussian OR Belorussia OR Belize OR Bhutan OR Bolivia OR Bosnia OR Herzegovina OR Hercegovina OR Botswana OR Brasil OR Brazil OR Bulgaria OR "Burkina Faso" OR "Burkina Fasso" OR "Upper Volta" OR Burundi OR Urundi OR Cambodia OR "Khmer Republic" OR Kampuchea OR Cameroon OR Cameroons OR Cameron OR Camerons OR "Cape Verde" OR "Central African Republic" OR CAR OR Chad OR Chile OR China OR Colombia OR Comoros OR "Comoro Islands" OR Comores OR Mayotte OR Congo OR Zaire OR "Costa Rica" OR "Cote d'Ivoire" OR "Ivory Coast" OR Croatia OR Cuba OR Cyprus OR Czechoslovakia OR "Czech Republic" OR Slovakia OR "Slovak Republic" OR Djibouti OR "French Somaliland" OR Dominica OR "Dominican Republic" OR "East Timor" OR "East Timur" OR "Timor Leste" OR Ecuador OR Egypt OR "United Arab Republic" OR "El Salvador" OR Eritrea OR Estonia OR Ethiopia OR Fiji OR Gabon OR "Gabonese Republic" OR Gambia OR Gaza OR Georgia OR Ghana OR "Gold Coast" OR Greece OR Grenada OR Guatemala OR Guinea OR Guam OR Guiana OR Guyana OR Haiti OR Honduras OR Hungary OR India OR Maldives OR Indonesia OR Iran OR Iraq OR Jamaica OR Jordan OR Kazakhstan OR Kazakh OR Kenya OR Kiribati OR Korea OR Kosovo OR Kyrgyzstan

OR Kirghizia OR "Kyrgyz Republic" OR Kirghiz OR Kirgizstan OR "Lao PDR" OR Laos OR Latvia OR Lebanon OR Lesotho OR Basutoland OR Liberia OR Libya OR Lithuania OR Macedonia OR Madagascar OR "Malagasy Republic" OR Malaysia OR Malaya OR Malay OR Sabah OR Sarawak OR Malawi OR Nyasaland OR Mali OR Malta OR "Marshall Islands" OR Mauritania OR Mauritius OR "Agalega Islands" OR Mexico OR Micronesia OR "Middle East" OR Moldova OR Moldovia OR Mongolia OR Montenegro OR Morocco OR Mozambique OR Mocambique OR Myanmar OR Myanma OR Burma OR Namibia OR Nepal OR "Netherlands Antilles" OR "New Caledonia" OR Nicaragua OR Niger OR Nigeria OR "Northern Mariana Islands" OR Oman OR Muscat OR Pakistan OR Palau OR Palestine OR Panama OR Paraguay OR Peru OR Philippines OR Philipines OR Phillipines OR Phillippines OR Poland OR Portugal OR "Puerto Rico" OR Romania OR Rumania OR Roumania OR Russia OR Russian OR Rwanda OR Ruanda OR "Saint Kitts" OR "St Kitts" OR Nevis OR "Saint Lucia" OR "St Lucia" OR "Saint Vincent" OR "St Vincent" OR Grenadines OR Samoa OR "Samoan Islands" OR "Navigator Island" OR "Navigator Islands" OR "Sao Tome" OR "Saudi Arabia" OR Senegal OR Serbia OR Montenegro OR Seychelles OR "Sierra Leone" OR Slovenia OR "Sri Lanka" OR Ceylon OR "Solomon Islands" OR Somalia OR Sudan OR Suriname OR Surinam OR Swaziland OR Syria OR Tajikistan OR Tadzhikistan OR Tadjikistan OR Tadzhik OR Tanzania OR Thailand OR Togo OR "Togolese Republic" OR Tonga OR Trinidad OR Tobago OR Tunisia OR Turkey OR Turkmenistan OR Turkmen OR Uganda OR Ukraine OR Uruguay OR "USSR" OR "Soviet Union" OR "Union of Soviet Socialist Republics" OR Uzbekistan OR Uzbek OR Vanuatu OR "New Hebrides" OR Venezuela OR Vietnam OR "Viet Nam" OR "West Bank" OR Yemen OR Yugoslavia OR Zambia OR Zimbabwe OR "developing country" OR "developing countries" OR "developing nation" OR "developing nations" OR "developing world" OR "less-developed countr*" OR "less developed countr*" OR "less-developed world" OR "less-developed world" OR "lesser-developed countr*" OR "lesser developed countr*" OR "lesser-developed nation" OR "lesser developed nation*" OR "lesser developed world" OR "lesser-developed world" OR "under-developed countr*" OR "under developed countr*" OR "under-developed nation*" OR "under developed nation*" OR "under-developed world" OR "underdeveloped world" OR "under developed world" OR "underdeveloped countr*" OR "under-developed countr*" OR "Under developed countr*" OR "under developed nation*" OR "under-developed nation*" OR "underdeveloped nation*" OR "lower middle income countr*" OR "lower middle-income countr*" OR "lower middle income nation*" OR "lower middle-income nation*" OR "upper middle-income countr*" OR "upper middle income countr*" OR "upper middle-income nation*" OR "upper middle income nation*" OR "low-income countr*" OR "low income countr*" OR "low-income nation*" OR "low income nation*" OR "lower income countr*" OR "lower-income countr*" OR "lower income nation*" OR "lower-income nation*" OR "Low- and Middle- Income countr*" OR "Low and Middle Income Countr*" OR "underserved country" OR "underserved countries" OR "underserved nation" OR "underserved nations" OR "underserved world" OR "under served country" OR "under served countries" OR "under served nation" OR "under served nations" OR "under served world" OR "deprived country" OR "deprived countries" OR "deprived nation" OR "deprived nations" OR "deprived world" OR "poor country" OR "poor countries" OR "poor nation" OR "poor nations" OR "poor world" OR "poorer country" OR "poorer countries" OR "poorer nation" OR "poorer nations" OR "poorer world" OR "developing economy" OR "developing economies" OR "less developed economy" OR "less developed economies" OR "lesser developed economy" OR "lesser developed economies" OR "under developed economy" OR "under developed economies" OR "underdeveloped economy" OR "underdeveloped economies" OR "middle income economy" OR

"middle income economies" OR "low income economy" OR "low income economies" OR "lower income economy" OR "lower income economies" OR lmic OR lmics OR "third world" OR "lami country" OR "lami countries" OR "transitional country" OR "transitional countries" LMIC OR LMICs OR LIC OR LICs OR LMICs OR LMIC OR UMICs OR UMIC) OR ("khmer" AND "republic") OR ("cape" AND "verde") OR ("central" AND "african" AND "republic")

2. SEARCH FOR 'WOMEN' TERMS

Woman OR Women OR Women's OR Mothers OR young mother OR Female OR Females OR Wife OR Wives OR 'Young Women' OR 'older girls' OR Femini* OR Gender OR Maternal OR Maternity OR daughter OR Daughters

OR

((Student OR students OR adolescents OR adolescent OR youth OR adolescence OR "young adults" OR "young adult" OR teenagers OR teenage OR teenager OR teenaged OR NEETs OR NEET OR "factory worker*" OR "wage worker*" OR employee* OR "schoolleaver*" OR "school leaver*" OR "high school graduate" OR "not in education") AND (girls OR girl OR woman OR women OR female OR females))

3. SEARCH FOR 'TYPES OF STUDIES' TERMS

Systematic review*" OR "Observational stud*" OR "Longitudinal stud*" OR "impact stud*" OR "Impact evaluation*" OR "comparison stud*" OR "Longitudinal Analysis*" OR "impact analysis" OR "observational analysis" OR "non-comparison stud*" OR Imp-Act OR "random* control* trial*" OR "random* trial*" OR "comparison group*" OR "control group*" OR RCT OR experiment* OR "program* evaluation*" OR "experimental control*" OR "pilot scheme*" OR "Pilot stud*" OR "pilot program*" OR "time series" OR "comparative analysis" OR Quasi-experiment* OR post-test* OR posttest* OR "post test*" OR pre-test* OR pretest* OR "pre test*" OR "project apprais*" OR "cluster random* trial*" OR "propensity score matching" OR PSM Or "regression discontinuity design" OR RDD OR "difference in difference*" OR DID OR meta-analy* OR "meta analy*" OR "control* random* trial*" OR "case control*" OR matching OR "interrupted time series" OR "random* allocation*" OR "instrumental variable*" OR "research synthesis" OR "scoping review" OR "rapid evidence assessment*" OR "systematic literature review*" OR "Rapid Review*" OR QED OR "group design" OR "intervention group" OR "intervention groups" OR "controlled stud*" OR "comparative stud*" OR "between group" OR "between groups" OR "group difference*" OR "Quasi-experiment*" OR "experimental group*" OR "control community" OR "intervention community" OR "intervention communities" OR "control communities" OR "intervention condition*" OR "control condition*" OR "controlled condition*" OR "control participant*" OR "experimental condition*"

OR

(Design N3 quantitative) OR (Design N3 "comparison group") OR (Design N3 counterfactual) OR (Design N3 "counter factual") OR (Design N3 counter-factual) OR (Design N3 experiment*) OR (Study N3 quantitative) OR (Study N3 "comparison group") OR (Study N3 counterfactual) OR (Study N3 "counter factual") OR (Study N3 "counter-factual") OR (Study N3 experiment*) OR (Analysis N3 quantitative) OR (Analysis N3 "comparison group") OR (Analysis N3 counterfactual) OR (Analysis N3 "counter factual") OR (Analysis N3 counter-factual) OR (Analysis N3 experiment*) OR (random* N3 allocat*) OR (effectiveness N3 intervention*) OR (before N2 after study) OR (matched N3 group)

4. SEARCH FOR 'INTERVENTION' TERMS

"Job Market" OR "labour Power" OR "Labor Power" OR Employee* OR "Labor Market" OR "Labour Market" OR "Labour Force" OR "labor Force" OR "Labour Economy" OR "Human Capital" OR "Economic Capital" OR "Labour Demand" OR "Labour Supply" OR "Wage Labour" OR "Wage labor" OR "employment reservation" OR "labour reservation" OR "Job reserv*" OR "Wage Gap*" OR "Wage differential*" OR "Low wage*" OR "Work* Condition*" OR "labour incentive*" OR "monetary incentive*" OR "financial incentive*" OR "minimum wage" OR "minimum wages" OR "employment tax*" OR "labour tax*" OR "labor tax*" OR "job insecurity" OR "employment security" OR "labour market regulation*" OR "labor market regulation*" OR "labour law*" OR "labor law*" OR "labour regulation*" OR "labor regulation*" OR "labour polic*" OR "labor polic*" OR "Employment polic*" OR "employment law" OR "employment laws" OR "labour reform" OR "labor reform" OR "labour reforms" OR "labor reforms" OR "job security" OR "labor standard*" OR "labour standard*" OR "labor code*" OR "labour code*" OR "labor legislation" OR "labour legislation" OR "Labour Supply" OR "Labour market participation" OR "Labour Force Participation" OR "labor force participation" OR "Labour Market" OR "Labor Market" OR "labor market participation" OR "Employment Trend*" OR "Employment Pattern*" OR "Employment Potential" OR "Economic Conditions" OR "Economic Development" OR Infrastructure OR "Economic Opportunit*" OR "economic right*" OR "Public spending" OR "Public Finance" OR "finance law" OR "finance polic*" OR "Public Works" OR "Job creation" OR "employment guarantee*" OR "Trade polic*" OR "trade law*" OR "export polic*" OR "export law*" OR "Public Sector" OR "tariff* reduction*" OR "tariff* change*" OR "Trad* Openness" OR "Trade reform*" OR "Trade liberalisation" OR "Trade liberalization" OR "Preferential Trade Agreement*" OR "Trade PTA*" OR "Free Trade Agreement*" OR "Trade FTA*" OR "special economic zone*" OR "export processing zone*" OR "free trade zone*" OR "free zone*" OR "foreign trade zone*" OR "industrial park*" OR "industrial estate*" OR "urban enterprise zone*" OR FTZ* OR EPZ* OR "Monetary polic*" OR "Public Spending" OR "Fiscal Stimulation" OR SME OR SMME OR "small enterprise*" OR start-up OR "medium enterprise*" OR "medium siz* enterprise*" OR "High Skill*" OR "formal Enterprise*" OR "small and medium-sized enterprise*" OR "Labour Economy" OR "Human Capital" OR "Economic Capital" OR "Labour Demand" OR "Wage Labour" OR "Wage labor" OR "labour Power" OR "Labor Power" OR "labour quota*" OR "employment quota*" OR "employee quota*" OR "gender quota*" OR "affirmative Action" OR "Economic Empowerment" OR "Procurement norm*" OR "procurement Standard*" OR "Tax regulation*" OR "Tax certification" OR "tax simplification" OR "Income tax credit*" OR "demand-side" OR "business development" OR "Financial service*" OR "Personal finance" OR "Personal financial" OR "Domestic investment*" OR "Personal wealth" OR Microenterpris* OR "Micro-enterpris*" OR

"Productive resource*" OR "micro credit" OR Micro-lease OR Monetary OR Bank* OR Insurance OR "Economic participation" OR "Tax penalt*" OR "Second Earner*" OR "Tax incentive*" OR "Tax Regulation*" OR "Tax Certification" OR franchise OR lease OR "market access" OR "value chain*" OR "technical and vocational education and training" OR "Technical Education" OR "Vocational Training" OR "Vocational Education" OR "on the job training" OR "on-the-job training" OR "Technical and Vocational Voucher*" OR "Job Placement" OR "Retraining" OR "Re-training" OR Employability OR "Employment Support" OR "Training Support" OR "Capacity Building" OR "Appropriate Skills" OR "Role Models" OR "Business Network*" OR "labour network*" OR "enterprise network*" OR "industry network*" OR "Career option*" OR "Career Progression" OR "Business Leadership" OR "Job Counselling" OR "Career Guidance" OR "Career counselling" OR "job fair" OR "mentor*" OR "business advisory" OR "Business Skill*" OR "entrepreneur* training" OR internship* OR intern OR interns OR "job placement" OR "Soft Skills" OR Apprentic* OR "labour market information" OR "Job search*" OR "Job Seek" OR "Active Labour Market Polic*" OR ALMP OR "job matching" OR "employment support" OR "educational voucher*" OR "entrepreneur* training" OR "self-help group*" OR "empowerment group*" OR "peer support" OR "Participatory learning" OR Traineeships OR traineeship OR TVET OR "technical training" OR "tech* prep*" OR "technician education" OR "technical stud*" OR "technical cent*" OR "technical school*" OR "technical course*" OR "technical program*" OR "technical college*" OR "technical degree*" OR "technical diploma*" OR "technical qualification*" OR "vocational stud*" OR "vocational retraining" OR "vocational work experience" OR "vocational cent*" OR "vocational school*" OR "vocational course*" OR "vocational program*" OR "vocational college*" OR "vocational degree*" OR "vocational diploma*" OR "vocational qualification*" OR "vocational framework*" OR "industrial education" OR traineeship* OR "trade course*" OR "job training" OR "job-related training" OR "job-site training" OR "in-service training" OR "retraining" OR "training program*" OR "skill* training" OR "skill* development" OR "staff development" OR "work place learning" OR "work based learning" OR "work related learning" OR "work* education" or "work place education" OR "work based education" OR "work related education" OR "work* training" OR "work place training" OR "work based training" OR "work related training" OR "work* program*" OR "work place program*" OR "work based program*" OR "work related program*" OR "work experience program*" OR "workforce development" OR "labour market" or "labor market" OR "employment based education" OR "employment based training" OR "employ* training" OR "employ* education" OR "employ* development program*" OR "employ* program*" OR "employ* course*" OR "unemploy* training" OR "training for unemployed" OR "training for the unemployed" OR "occupation* education" OR "occupation* training" OR "occupation* program*" OR "occupation* course*" OR "business education" OR "office occupations education" OR "contract training" OR "school to career program*" OR "school to work program*" OR "career* education" OR "youth program*" OR "company training" OR "company-based learning" OR "investment in training" OR "Occupational Mobility" OR "Business Training" OR "Business Loan*" OR "Business Grant*" OR "Business ownership" OR "Business Credit" OR Asset* OR "land tenure" OR "Property" OR "title deeds" OR Livestock OR Capital OR micro-Saving* OR "Micro saving*" OR "flexible work" OR "flexible Hours" OR "Flexible employment" OR "Flexible Labour" OR "Flexible labor" OR "local productive system*" OR "collective action*" OR "economic cluster*" OR "industry cluster*" OR "enterprise cluster*" OR "matching grant*" OR formalization OR "business environment" OR "property registration" OR "regulatory framework*" OR "mandatory employee benefit*" OR "unemployment

benefit" OR "unemployment benefits" OR "payroll tax" OR "payroll taxes" OR "hour restrictions" OR "hiring rigidity" OR "hiring rigidities" OR "firing rigidity" OR "firing rigidities" OR "termination benefit" OR "termination benefits" OR "job insecurity" OR "employment security" OR "labour rigidity" OR "labor rigidity" OR "labor rigidities" OR "labour rigidities" OR "collective bargaining" OR "labour market regulation*" OR "labor market regulation*" OR "labour regulation*" OR "maternity protection" OR "employment search*" OR "work search*" OR "employment seek*" OR "work seek*" OR "Financ* literacy" OR "financ* inclusion" OR "financial awareness" OR "financial capabilit" OR "financial competence" OR "financial education" OR "financial knowledge" OR "financial literacy" OR "gender-responsive budget*" OR "female-owned enterprise*" OR "women-owned enterprise*" OR "female-owned business" OR "women-owned business" OR "female-owned industry*" OR "women-owned industr*" OR "female-owned factor*" OR "women-owned factor*" OR "business practice*" OR "labour practice*" OR "employment practice*" OR "procurement practice*" OR "business standard*" OR "labour standard*" OR "employment standard*" OR "procurement standard*" OR "business norm*" OR "labour norm*" OR "employment norm*" OR "procurement norm*" OR "Combined structural intervention*" OR "bundled service*" OR "livelihood program*" OR "poverty graduation program*" OR microfinance OR micro-finance OR "micro finance" OR microcredit OR micro-credit OR "micro loan" OR "micro-loan" OR microloan OR microloans OR "micro-loans" OR funding OR "Financial Support" OR Transport OR Sanitation OR Water OR telecommunications OR telecommunication OR "Business Technolog*" OR Electricity OR Electrification OR "clean stoves" OR "clean energy" OR "household fuel" OR power OR ICT OR "information and communication technolog*" OR "mobile technolog*" OR phone* OR telephone* OR credit OR Saving* OR Finance OR lending OR "fair trade" OR "ethical trade" OR "sustainability standard*" OR "Corporate social responsibility" OR CSR OR "corporate responsibility" OR "code of conduct*" OR "Gender Mainstream*" OR "gender awareness" OR "outreach program*" Or "advocacy Program*" OR "awareness campaign*" OR "awareness raising" OR Empowerment OR "Support group*" OR "advocacy group*" OR "Job Market" OR scholarship* OR subvention* OR stipend* OR donation OR bursary OR bursaries OR "tuition relief" OR "merit aid" OR "merit based aid" OR "merit-based aid" OR "merit award" OR advisory OR tutor* OR Loan* OR Grant* OR Credit

OR

(budget* OR "resource allocation" OR "public fund*") AND (gender)

OR

(land OR property) AND (right* OR conversion OR freehold OR titl* OR codification OR recognition OR customary OR certification)

OR

(career OR skill* OR work OR Performance) AND (Chang* or increas* or rise* or rising or rose or rais* or augment* or growth* or grow* or grew or improv* or gain* or motivat* or promot* or encourag* or enhanc* or boost* or achiev* or success* or succeed* or accomplish* or thrive* or thriving or achiev*

or attain OR enhance OR Upgrade)

OR

(Leave OR Care) AND (Matern* OR Patern* OR parent* OR Child OR Elder* OR disab*)

OR

(Work* OR employment OR Business OR Unemployment OR career OR employee OR job OR profession OR occupation) AND (Violence OR abuse OR exploitation OR harassment OR equity OR equality OR childcare OR "infant care" OR "child daycare" OR daycare OR "day care" OR nursery OR nurseries OR "nursery school" OR "nursery schools" OR pre-school OR pre-schools OR kindergarten OR "family caregiving," OR "informal caregiving," OR "unpaid caregiving" OR "Social Constraint*" OR Norm* OR fairness OR Inequality OR Discrimin* OR Dispar* OR "Self Esteem" OR self-esteem OR Self Confidence OR self-confidence OR motivation OR Equity OR Mainstreaming OR Exploit* OR equality OR information OR skill* OR training OR coaching OR Empowerment)

OR

(Pay* OR Remuneration OR Salar* OR Benefits OR Incentive* OR Financial OR Money OR Monetary OR Reward* OR Wage* OR Bonus OR Pension OR earning*) AND (change* OR Increase* OR Rise* OR Augmentation* OR Grow*)

OR

(subsid* OR subsidy OR subsidies OR subsidized OR subsidised) AND (Wage* OR Labour OR Employment)

OR

(Work* OR employment OR Business OR Unemployment OR career OR employee OR job) (Household OR residential OR domestic) AND (energy OR fuel) AND (choice OR switch OR switching)

OR

(industrialization OR industrialisation OR "industrial policy" OR "industrial policies" OR "investment policy" OR "investment policies" OR "domestic investment*" OR "foreign investment*" OR "foreign direct investment*" OR "fiscal policy" OR "fiscal policies" OR "monetary policy" OR "monetary policies")

GREY LITERATURE SEARCH STRATEGY

IMPACT EVALUATION AND SYSTEMATIC REVIEW REPOSITORIES

- UN Women Economic Empowerment: A road map – Database of Empirical Evaluations (http://www.womeneconroadmap.org/the_database)
- UN Gender Equality Evaluation Portal (http://genderevaluation.unwomen.org/en)
- 3ie Database of Systematic Reviews (http://www.3ieimpact.org/evidence/systematic-reviews/)
- enGENDER IMPACT A Gateway to Gender-Related Impact Evaluations (http://www.worldbank.org/en/topic/gender/publication/engender-impact-a-gateway-to-gender-related-impact-evaluations)
- 3ie Registry for International Development Impact Evaluations (RIDIE) (http://ridie.3ieimpact.org/)
- 3ie Register of Impact Evaluation Published Studies (RIEPS) (http://www.3ieimpact.org/evidence/impact-evaluations/)
- Campbell Collaboration Library: systematic reviews (http://www.campbellcollaboration.org)
- EPPI-Centre Systematic Reviews Library (https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=56)

SPECIALISED DATABASES

- Abdul Latif Jameel Poverty Action Lab (J-PAL) Evaluation and Publication Database (http://www.povertyactionlab.org/)
- ELDIS (http://www.eldis.org/)
- Innovations for Poverty Action (IPA) Database (http://www.poverty-action.org/project-evaluations/search)
- International Centre for Research on Women (http://www.icrw.org/)
- Labordoc (ILO) (http://labordoc.ilo.org/)
- RePEc (Research Papers in Economics)/IDEAS Economics and Finance Research: (http://ideas.repec.org/)
- Research for Development (http://r4d.dfid.gov.uk/)
- USAID Development Experience Clearinghouse (https://dec.usaid.gov/)
- World Bank Poverty Impact Evaluations Database (http://www1.worldbank.org/prem/poverty/ie/evaluationdb.htm)
- World Bank Independent Evaluation Group (IEG) (http://ieg.worldbankgroup.org)
- Youth Employment Inventory (YEI) (http://www.youth-employment-inventory.org/)

ORGANISATIONAL WEBSITES

- African Development Bank Evaluation Reports (http://www.afdb.org/en/documents/evaluation-reports)
- Asian Development Bank (ADB) Evaluation Resources (http://www.adb.org/site/evaluation/resources)
- Bill & Melinda Gates Foundation (http://www.gatesfoundation.org/)
- BRAC (http://www.brac.net/)
- Centre for Global Development (http://www.cgdev.org/)
- Ford Foundation (https://www.fordfound.org/)
- Harvard Women and Public Policy Program (http://wappp.hks.harvard.edu/about-wappp)
- Hewlett Foundation (http://www.hewlett.org/)
- Inter-American Development Bank Office of Evaluation and Oversight (https://www.iadb.org/en/evaluation)
- International Growth Centre (http://www.theigc.org/)
- Institute of Labour Economics (IZA) (http://www.iza.org)
- Institute of Development Studies (IDS) (http://www.ids.ac.uk/)
- Millennium Challenge Corporation (MCC)(https://data.mcc.gov/evaluations/index.php/catalog)
- National Bureau of Economic Research (NBER) (http://www.nber.org)
- Overseas Development Institute (ODI) (http://www.odi.org.uk/)
- Oxfam (https://www.oxfam.org/)
- Poverty and Economic Policy Research Network (PEP) (Project List https://www.pep-net.org/find-pep-project)
- Self Employed Women's Association (SEWA) (http://www.sewa.org/)
- UNDP International Policy Centre for Inclusive Growth (IPC-IG) (http://www.ipc-undp.org/)
- UNESCODoc (http://www.unesco.org/new/en/unesco/resources/online-materials/publications/unesdoc-database/)
- World Bank Labor Markets (http://www.worldbank.org/en/topic/labormarkets)

BACKWARD CITATION TRACKING

We screened the reference lists of all the included studies as well as existing systematic reviews for relevant studies. The following reviews and evidence maps were searched in this way:

- ILO What Works In Youth Employment Evidence Map (http://www.wwinye.org/wwinye/evidence-gap-map)
- 3ie Youth & Transferable Skills Evidence Gap Map (http://gapmaps.3ieimpact.org/evidence-maps/youth-transferable-skills-evidence-gap-map)
- Brody et al (2015)
- UN (2013)
- Buvinic and O'Donnell (2016)

- Kluve et al (2016)
- McKenzie and Woodruff, C (2013)
- Taylor and Pereznieto (2014)
- ODI (2016)
- Peters et al 2016)
- Piza et al (2016)
- Tripney et al (2013).

FORWARD CITATION TRACKING

We ran a citation search in Google Scholar to identify literature that had cited the studies included in our review. All citations for each included study were screened.

HAND-SEARCHING OF SPECIALISED JOURNALS

- *Economic Development and Cultural Change*
- *Feminist Economics*
- *Journal of Development Economics*
- *Journal of International Development*
- *World Development*

APPENDIX 5: DATA EXTRACTION TOOL

This tool will be translated into a coding set on EPPI-reviewer allowing for additional sub-levels of codes.

Questions	Answers
Section A: Administration	
Name of reviewer	Details (specify)
Linked reports	None / not known
	Linked (specify)
	Unclear (specify)
Language of report *If more than one report, choose the main report.*	English
	Other (specify)
Section B: Study characteristics	
Form of publication *If more than one report, choose the main report.*	Journal article
	Grey Literature (specify)
	Dissertation/thesis (specify)
	Other (specify)
Year of publication *If more than one report, choose the main report.*	1990-1994 (specify)
	1995-1999 (specify)
	2000-2004 (specify)
	2005-2009 (specify)
	2010-2014 (specify)
	2015-2017 (specify)
Broad aims of the study / research question	Explicitly stated (specify)
	Implicit (specify)
	Unclear/not stated (specify)

Questions	Answers
Study funding	Government or government-related agency (specify)
	Donor country agency
	Non-governmental organisation (NGO)/non-profit (specify)
	Multilateral agency (e.g. development bank or WHO) (specify)
	Academic/research institution (specify)
	Employer (specify)
	Other (specify)
	Unclear/not stated (specify)
When was the study conducted?	Initial year (specify)
	Final year (specify)
	Unclear/not stated (specify)
Section C: Subject characteristics	
Region *Details about the country/countries where programme is implemented are collected in Section D.*	Low-income country
	Lower-middle income country
	Upper-middle income country
	Africa
	Asia
	Europe
	Latin America & Caribbean
	Oceania
Region classification	High fertility agrarian societies
	Declining fertility urbanising societies
	Declining fertility formalising economies
	Ageing societies

Questions	Answers
Total number of study participants *This question refers to the number of subjects who were originally assigned to the treatment and control groups. Later in the tool there will be a question about the number of subjects who were actually observed/measured.* *Typically you can use baseline sample sizes for the assigned N, but the authors may have removed the subjects with incomplete data. In this event, indicate that the figure you provide is approximate.* *Report total N.*	Total 250 or less (specify)
	Total 251-499 (specify)
	Total 500 or more (specify)
	Unclear/not stated (specify)
Age *Select all that apply.*	Working aged adults aged 15-64 years (specify)
	Older adults aged 65+ (specify)
	Youth (15-35)
	Unclear/not stated (specify)
Is the study focussed specifically on young people?	Yes (specify)
	No
	Unclear (specify)
Is the study focussed specifically on transitions from school (incl. out-of-school) to work?	Yes (specify)
	No
	Unclear (specify)
Is the study focussed specifically on returning to work after childbirth?	Yes (specify)
	No
	Unclear (specify)
Is the study focussed specifically on migrating from rural to urban areas?	Yes (specify)
	No
	Unclear (specify)

Questions	Answers
Is the study focussed specifically on transitions from unemployment back into employment?	Yes (specify)
	No
	Unclear (specify)
Is the study focussed specifically on transitions from farm labour to non-farm labour	Yes (specify)
	No
	Unclear (specify)
Is the study focussed specifically on transitions from informality to formality?	Yes (specify)
	No
	Unclear (specify)
Is the study focussed specifically on transitions from conflict/trauma back into society?	Yes (specify)
	No
	Unclear (specify)
Sex	Females only
	Mixed (specify)
	Unclear (specify)
Descriptive characteristics of the sample	Years of schooling (specify)
	Income indicators (specify)
	Asset indicators (specify)
	Social capital indicators (specify)
	Employment indicators (specify)
Based on the above, identify the sample as per UN definition:	Very poor women
	Poor women
	Non-poor women

Questions	Answers
Other useful information about study participants.	Details (specify)
	None
Section D: Intervention characteristics	
Formal name (if any)	Not applicable (no formal name)
	Details (specify)
	Unclear (specify)
Scale (or availability) of the intervention *(specific to this particular study)*	International
	National
	Provincial/Regional (specify)
	Local (specify)
	Other (specify)
	Unclear/not stated (specify)
Which level of social organisation is targeted?	Individuals
	Households
	Community groups/collectives
	Entire communities/villages
	Others (specify)
Maturity of the intervention	Pilot (specify)
	Iteration following pilot/prior experience
	Scale up
Country or countries where the intervention is implemented *(specific to this particular study)*	Details (specify)
	Unclear (specify)
Area of country where the intervention is implemented (e.g. particular regions or cities) *(specific to this particular study)*	Details (specify)
	Unclear (specify)

Questions	Answers
Primary location *(specific to this particular study)*	Primarily urban
	Primarily rural
	Both
	Unclear/not stated (specify)
Type of intervention	Interventions to balance work & family responsibilities
	Increase women's financial returns
	Changing business culture/practice
	Macroeconomic changes
	Provision of infrastructure
	Microfinance
	Cash transfers
	Economic assets
	Changes to land titles, business ownership, inheritance
	Bundled services/combined structural interventions
	Interventions to provide education/skills for women
	Interventions to provide access to economic opportunities
	Interventions to provide work experience for women
	Interventions to provide support to female businesses and entrepreneurs
	Social organisation
	Changes in norms and attitudes
	Gender sensitive design
	Behavioural nudges
Barriers addressed by the intervention	Discrimination by markets and work institutions
	Constraints in access to credit, finance and assets

Questions	Answers
	Constraints in employability & entrepreneurship
	Lack of social capital and norms
	Behavioural (social & cognitive) barriers
Does the paper refer to one or more specific theories for how the intervention should work?	Yes (specify)
	No
	Unclear/not stated (specify)
Is the 'treatment' a single 'activity' or a combination of activities? *If multi-component (also known as bundled or combined interventions), provide details of the different components. Multi-component interventions will typically target a number of different barriers/constraints faced by recipients.*	Single
	Multi-component (specify)
	Unclear/not stated (specify)
Dates of operation *Include both start and ending dates of the intervention where these are available, and where relevant note if it is ongoing.*	Not applicable (specify)
	Details (specify)
	Unclear/not stated (specify)
Who funded the intervention?	Government or government-related agency (specify)
	Donor country agency (e.g. DIFD) (specify)
	Non-governmental organisation (NGO)/non-profit (specify)
	Multilateral agency (e.g. development bank or WHO) (specify)
	Individual donor (e.g. charitable foundations or private sector)
	Academic/research institution (specify)
	Employer of beneficiaries (specify)
	Beneficiaries
	Other (specify)
	Unclear/not stated (specify)

Questions	Answers
Who designed the intervention?	Government or government-related agency (specify)
	Donor country agency (e.g. DIFD) (specify)
	Non-governmental organisation (NGO)/non-profit (specify)
	Multilateral agency (e.g. development bank or WHO) (specify)
	Individual donor (e.g. charitable foundations or private sector)
	Academic/research institution (specify)
	Employer of beneficiaries (specify)
	Beneficiaries
	Other (specify)
	Unclear/not stated (specify)
Who implemented the intervention? *This question is not referring to the delivery of the intervention (see below for question about service provision).*	Government or government-related agency (specify)
	Donor country agency (specify)
	Non-governmental organisation (NGO)/non-profit (specify)
	Multilateral agency (e.g. development bank or WHO) (specify)
	Individual donor (e.g. charitable foundations or private sector)
	Academic/research institution (specify)
	Employer of beneficiaries (specify)
	Other (specify)
	Unclear (specify)
Political/economic/social/cultural context (at time of intervention)	Details (specify)
	Unclear/not stated (specify)
Target group of intervention	Not targeted specifically at any group

Questions	Answers
Select all that apply, and only if targeting of the intervention is specifically mentioned (i.e. the study may focus on a specific population such as the low paid, but this is not what this question relates to).	Women
	Women returning to work after childbirth
	People within a certain age range, e.g. under 30 years (specify)
	Young people making transition from school (out of school) to work
	People within a specific geographical location (specify)
	People with disabilities (specify)
	People within certain ethnic groups (specify)
	People with low education (primary or lower) (specify)
	People with secondary education (or equivalent) (specify)
	People with higher education (above secondary) (specify)
	People on low incomes/disadvantaged (specify)
	People in urban locations (specify)
	People in rural locations (specify)
	People migrating from rural to urban areas (specify)
	People unemployed at intervention start
	People who lost employment, looking to return to work
	People already employed/entrepreneur at intervention start (specify)
	First-time jobseekers only (specify)
	People of a particular religion
	People focussed on a particular occupation
	Groups stated as vulnerable
	Other (specify)
	Unclear/not stated (specify)

Questions	Answers
Descriptive characteristics of the sample	Age (specify)
	Years of schooling (specify)
	Income indicators (specify)
	Asset indicators (specify)
	Social capital indicators (specify)
Based on the above, identify the sample as per UN definition:	Very poor women
	Poor women
	Non-poor women
Does the intervention design include gender considerations? *The intervention may or may not have been targeted specifically at women.*	No
	Yes (specify)
	Taking into consideration travel constraints (e.g. distance to homes)
	Taking into consideration time constraints (e.g. when the programme is offered)
	Taking into consideration caregiving constraints
	Taking into consideration the gender of programme implementers
	Changes to social / professional norms
	Quota/reservation approaches to ensure women's participation
	Interventions working through subjective economic empowerment (e.g. self-reliance/esteem)
	Increased risk taking
	Interventions addressing cognitive and social determinants of economic behaviour (e.g. Decision-making support; protection from external pressures)
	Other features supporting a gender-sensitive approach/addressing gender biases embedded in organisations/working environment (specify)

Questions	Answers
	Unclear/not stated (specify)
Does the intervention apply other design features?	Interventions applying explicit behavioural designs (e.g. commitment devices, reminders)
	Financial input from employer
	Demand-driven approaches (i.e. demands by markets)
	Bundling of interventions
Does the intervention design focus specifically on improving women's employment in higher growth/male-dominated sectors, and if so which ones?	No
	Manufacturing
	Trade
	Construction
	Energy (mining & quarrying, electricity, gas & water supply)
	Wood pulp and forestry
	Transportation
	Accommodation & food
	Business administration services
	Electronics and ICT
	Finance
	Commercial agriculture
	Maritime services
	Higher education/Science & Technology
	Unclear
	Other (specify)
Does the intervention design include awareness training about the intervention to targeted participants?	Not applicable (no targeting of the intervention)
	Yes (awareness training available)
	No

Questions	Answers
	Unclear/not stated (specify)
What was the take-up of the intervention?	Details (specify)
	Unclear/not stated
Is there a specific selection process? *In other words, is eligibility conditional on (targeted) individuals meeting further requirements, such as passing an interview or test?*	No (any woman/any woman within targeted population is eligible)
	Yes (specify)
	Unclear/not stated (specify)
Setting or site (for delivery of service/s)	Not applicable
	Classroom (college or training centre) (specify)
	Online (distance learning) (specify)
	Workplace (specify)
	Community site (specify)
	Other (specify)
	Mixed/multiple sites (specify)
	Unclear/not stated (specify)
Service provider/s	Not applicable (specify)
	Public institution/contractor (specify)
	Private institution/contractor (specify)
	NGO/non-profit (specify)
	Other (specify)
	Unclear/not stated (specify)
Payment system (to service provider/s)	Not applicable (specify)
	Lump-sum budget (specify)
	Payment for services delivered (specify)
	Payment by outcomes (specify)
	Other (specify)

Questions	Answers
	Unclear/not stated (specify)
Duration of treatment: overall average duration a single cohort stays in the programme	Not applicable (specify)
	One day or less (specify)
	One day to 1 week (specify)
	One week (and 1 day) to 1 month (specify)
	One month (and 1 day) to 3 months (specify)
	Three months (and 1 day) to 6 months (specify)
	Six months (and 1 day) to 1 year (specify)
	One year (and 1 day) to 2 years (specify)
	Two years (and 1 day) to 3 years (specify)
	Three years (and 1 day) to 5 years (specify)
	More than 5 years (specify)
	Other (please specify)
	Unclear/not stated (specify)
Frequency of contact between beneficiaries and provider or treatment activity	Less than weekly
	Once a week
	3-4 times a week
	Continuous
	Other (specify)
	Unclear/not stated (specify)
Intervention dosage: hours per week	Not applicable (specify)
	1-10 hours (specify)
	11-19 hours (specify)
	20 hours and over (specify)
	Unclear/not stated (specify)

Questions	Answers
Other intervention dosage:	Size of loan/cash transfer/asset (specify)
	Other (specify)
Were incentives provided to intervention participants?	Yes, monetary benefits (e.g. stipend, transport allowance)
	Yes, non-monetary benefits (e.g. transport, childcare)
	No
	Unclear/not stated
Implementation of the intervention: was this monitored to assess whether this was delivered as intended?	Yes (monitored by programme staff)
	Yes (monitored by researcher)
	No
	Unclear/not stated (specify)
Were there any implementation problems? *For example, high drop outs, erratic attendance.*	Yes (specify)
	No
	Unclear/not stated (specify)
Role of study funder *Consider, for example, whether the funder was also involved in the design and delivery of the intervention, or was a member of the research team.*	Not applicable (details of study funding not reported)
	Not independent (specify)
	Independent
	Unclear/not stated (specify)
Role of evaluators *Consider, for example, whether the evaluator was also involved in the design and delivery of the intervention.*	Not stated
	Not independent (specify)
	Independent
	Unclear (specify)
Other relevant information about the programme (if any)	Details
	Not applicable (time-series design, so only one group)
	No treatment

Questions	Answers
Intervention (if any) received by the control/comparison group	Treatment as usual (specify)
	Alternative intervention (specify)
	Other (specify)
	Unclear/not stated (specify)
Section E: Key methodological characteristics	
Study design according to publication	Specify:
Unit of group assignment	Not applicable
	Individuals
	Groupings (clusters) of individuals (specify)
	Programme area, region, etc. (specify)
	Unclear/not stated (specify)
Method of group assignment *This question focuses on the initial method of assignment to groups. For non-random methods of assignment, please indicate whether:* • *the investigator controlled group exposure,* • *participants self-selected into the programme,* • *assignment was by means of administrator selection, or* • *study was a natural experiment (where exposure to the treatment and control conditions are determined by nature).*	Random (RCT design, with allocation done randomly after matching, blocking, stratification, etc.)
	Random (RCT design without matching, etc.)
	Quasi-random (e.g. investigator allocates by date of birth)
	Quasi-random (Regression Discontinuity Design)
	Non-random, but matched and/or statistically controlled (specify)
	Unclear (specify)
If study used matching, what variables are used? *Matching can be done on the propensity score or covariates. Check if study matched on (i)*	Not applicable
	Details (specify)
	Unclear/not stated (specify)

Questions	Answers
pre-test measures of some or all variables used later as outcome measures and (ii) other characteristics, such as demographic variables. Make a note if matching is done on endline characteristics only.	
If study used statistical controls, what variables are used? *Check if study controlled for (i) pre-test variables (i.e. measures of a dependent variable taken prior to treatment) and (ii) other characteristics, such as demographic variables.*	Not applicable
	Details (specify)
	Unclear/not stated (specify)
Proportion of sample with missing/incomplete data	No missing/incomplete data
	1-20% (specify)
	21-30% (specify)
	More than 30% (specify)
	Unclear/not stated (specify)
Section F: Outcome measurement	
Intermediate outcomes	Employability (specify)
	Access to employment opportunities (specify)
	Labour market participation enhancing behaviour(specify)
	Social capital (specify)
	Policy change (specify)
	Other (specify)
	Unclear (specify)
Final outcomes (labour market participation)	None
	Change in employment status (from unemployed to employed in high-growth/male-dominated sector), measured in terms of employment probability or participation rate

Questions	Answers
	Change in employment sector (from traditional sector for women's employment to high-growth/male-dominated sector)
	Change in employment status (from underemployment to full employment in high-growth/male-dominated sector)
	Nature of employment
	Progression/career prospects
	Firm size
	Other (specify)
	Unclear (specify)
Type of measure (i.e. data source) *Where relevant, state whether data pre-existing or not (e.g. the investigators may use pre-existing survey data, or collect original data).*	In-depth interviews
	Survey, questionnaire
	Administrative data
	Other (specify)
	Unclear/not stated (specify)
Number of post-test measurements	One
	Two
	Three
	Other (specify)
	Unclear (specify)
Timing of outcome measurements *Select all that apply.* *Indicate if this is the time that has lapsed since baseline measurement, start of intervention or end of the intervention.* *If different outcomes are measured at different time points, provide details.*	Not stated
	0-6 months (specify)
	7-12 months (specify)
	13-18 months (specify)
	Over 18 months (specify)
	Other (specify)

Questions	Answers
	Unclear (specify)
Evidence of unintended effects	Doing harm (specify)
	Effect capture (specify)
	Other (specify)
	Unclear (specify)
Evidence of cost data being collected	Cost data reported
	Cost-effectiveness calculations being reported
Section H: Effect size calculation and data entry	
Section G: Narrative findings	
Authors report interventions as:	Effective (specify)
	Ineffective (specify)
	Other (specify)
	Unclear (specify)
Copy & paste authors conclusions	

APPENDIX 6: CRITICAL APPRAISAL TOOL

Methodological appraisal criteria	Response		
	Yes	No	Comment
IF RANDOMISED CONTROL TRIAL, START AFTER CONFOUNDING BIAS. FOR ALL OTHER STUDY DESIGNS, START HERE. I. Bias in selection of participants into the study *Are participants selected in a way that minimizes selection bias?* [42] Appraisal indicators Consider whether:			
i. there is an adequate description of how and why sample was chosen (i.e. identified/selected/recruited)			
ii. there is adequate sample size to allow for representative and/or statistically significant conclusions			
iii. participants in the control[43] *group were sampled from the same population as that of the treatment*			
iv. group allocation process minimised potential risk of bias (e.g. using computer algorithms)			
v. the selection of participants into the study (or into the analysis) is based on participant characteristics observed after the start of the intervention			
Low risk of bias / Risk of bias / High risk of bias / Critical risk of bias	*Worth to continue: Y/N?*		
II. Bias due to confounding *Is confounding potentially controllable in the context of this study?* Appraisal indicators: Consider whether:			
i. there is potential for confounding of the effect of the intervention in this study. If yes, provide example of confounding domain in comment box.[44]			

[42] Selection bias can occur both in the way that individuals are accepted for participation in a study, and in the way that 'treatment' is assigned to individuals once they have been accepted into a study. This section deals with both these understandings of selection bias.

[43] The terms 'control' and 'comparison' group refer to any group with which the treatment of interest is compared that is presumed to represent conditions in the absence of that treatment, whether a true random control or not.

44 Confounding domains are those for which, in the context of this study, adjustment is expected to lead to an important change in the estimated effect of the intervention.

Methodological appraisal criteria				Response		
				Yes	No	Comment
ii. where matching was applied, it featured sufficient criteria[45]						
iii. where relevant, the authors conducted an appropriate analysis that controlled for all potential/remaining critical confounding domains after matching had been applied						
iv. *the authors avoided adjusting for variables identified after the intervention has been administered*						
v. the treatment and control group are comparable after matching/controls have been done. Select one of the following: ☐ No statistically significant differences ☐ Statistically significance differences ☐ Negligible descriptive differences ☐ Significant descriptive differences						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	*Worth to continue: Y/N?*		
IF RANDOMISED CONTROL TRIAL, SKIP I + II (ABOVE) AND START HERE! <u>Bias due to confounding (as a result of ineffective randomisation)</u> *Is allocation of treatment status truly random?* <u>Appraisal indicators</u> Consider whether:						
i. eligibility criteria for study entry are specified						
ii. there is a clear description of the randomisation process and methods are robust						
iii. the unit of randomisation and number of participants is clearly stated (pay special attention to treatment and control locations/ balance)						
iv. characteristics of both baseline and endline sample are provided[46]*and at endline the treatment and control group are comparable. Select one of the following:*						

[45] Matching can be done on the calculated propensity score or covariates. If the latter, it should ideally be done on pre-test measures and other characteristics, such as demographic. Answer 'no' if the study only matched on pre-test measures of some or all variables used later as outcome measures OR matched only on endline characteristics.

[46] Preferable condition: A RCT with appropriate randomisation procedure can be included without showing baseline data, as both experimental groups can be assumed to be equal at baseline by design.

Methodological appraisal criteria				Response		
				Yes	No	Comment
☐No statistically significant differences ☐Statistically significance differences ☐Negligible descriptive differences ☐Significant descriptive differences						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	*If critical risk of bias, treat as non-random study*		
III. Bias due to departures from intended interventions *Was the intervention implemented as laid out in the study protocol?* Appraisal indicators Consider whether:						
i. the critical co-interventions were balanced across intervention and control groups						
ii. treatment switches were low enough to not threaten the validity of the estimated effect of the intervention						
iii. implementation failure was minor and unlikely to threaten the validity of the estimated effect of the intervention						
iv. it is possible that the intervention was taken by the controls (contamination and possible crossing-over)[47]						
v. it is possible that knowledge of group allocation affects how the two study groups are treated during delivery and evaluation of the intervention[48]						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	*Worth to continue: Y/N?*		
IV. Bias due to missing/incomplete data (attrition) *Are the intervention and control groups free of critical differences in participants with missing/incomplete data?* Appraisal indicators Consider whether:						
i. outcome data are reasonably complete (80% or above)[49]						

[47] Whilst challenging in terms of estimating impact, spill-overs might be an important finding in itself.
[48] Consider only in extreme cases in which preferential treatment is clearly evident; blinding is generally not expected in social interventions.
[49] The assumption here that the level of attrition (or other forms of missing/incomplete data) is sufficiently low to not require adjustment.

Methodological appraisal criteria				Response		
				Yes	No	Comment
ii. If level of attrition (or other forms of missing/incomplete data) is more than 20%, are reasons for the missing data reported?						
iii. If level of attrition (or other forms of missing/incomplete data) is more than 20%, do the authors demonstrate similarity between remaining participants and those lost to attrition and are the proportion of participants with missing/incomplete data and reasons for missing/incomplete data similar across groups?						
iv. If level of attrition (or other forms of missing/incomplete data) is more than 20%, were appropriate statistical methods used to account for missing data? (e.g. sensitivity analysis)[50]						
v. If not possible to control for missing/incomplete data, are outcomes with missing/incomplete data excluded from analysis?						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	*Worth to continue: Y/N?*		
V. <u>Bias in measurement of outcomes</u> *Are measurements appropriate, e.g. clear origin, or validity known?* <u>Appraisal indicators</u> Consider whether:						
i. there was an adequate period for follow up[51]						
ii. the outcome measure (e.g. employment status, income) was clearly defined and objective[52]						
iii. outcomes were assessed using standardised instruments and indicators						
iv. outcome measurements reflect what the experiment set out to measure						
v. the methods of outcome assessment were comparable across groups						

[50] Select 'no' if the study addresses missing/incomplete data through simple estimates of missing data and observations.
[51] In many social science interventions, follow-up is not required to coincide with the start of the treatment; further, longer periods of follow up are often required to measure changes.
[52] Subjective measures (e.g. those based on self-report) are likely to have lower reliability and validity than objective measures.

Methodological appraisal criteria				Response		
				Yes	No	Comment
vi. were outcome assessors aware of the intervention received by study participants?[53]						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias	*Worth to continue: Y/N?*		
VI. Bias in selection of results reported *Are the reported outcomes consistent with the proposed outcomes at the protocol stage?* Appraisal indicators Consider whether:						
i. it is unlikely that the reported effect estimate has been selected for publication due to it being a particularly notable finding among numerous exploratory analyses						
ii. it is unlikely that the reported effect estimate is prone to selective reporting from among multiple outcome measurements within the outcome domain						
iii. it is unlikely that the reported effect estimate is prone to selective reporting from among multiple analyses of the outcome measurements, including sub-groups analysis						
iv. if sub-group/ancillary/adjusted analyses are presented, are these pre-specified or exploratory?						
v. the analysis includes an intention to treat analysis? (If so, was this appropriate and were appropriate methods used to account for missing data?)[54]						
vi. do the authors report on all variables they aimed to study (as specified in their protocol or study aims/research questions)?						
Low risk of bias	Risk of bias	High risk of bias	Critical risk of bias			
OVERALL RISK OF BIAS:						

[53] Consider only in extreme cases in which preferential treatment is clearly evident; blinding is generally not expected in social interventions.

[54] Usually in clinical RCTs, rare in social science: only rate if conducted.

APPENDIX 7: EFFECT SIZE CALCULATIONS

Corrected SMD and corrected Standard Errors (SE) will be estimated as follows:

$$SMD_{corrected} = SMD_{uncorrected} * \left[1 - \frac{3}{4 * (n_t + n_c - 2) - 1}\right]$$

$$SE(SMD)_{corrected} = SE(SMD)_{uncorrected} * \left[1 - \frac{3}{4 * (n_t + n_c - 2) - 1}\right]$$

For regression-based estimates, we will follow Keef and Roberts (2004: 100-101, 129, equation A9) to correct for potential sample bias in the effect sizes.

Many of the impact evaluation designs that we expect to see in this review are likely to use complex statistical analyses, and there is a lack of standard methods for computing effect sizes from these designs. In most cases, we expect our approach for computing effect sizes to be as follows.

CALCULATING STANDARDISED MEAN DIFFERENCES

For studies reporting matching-based estimates, the following formulae to compute g and its standard error will be used, where Yt and Yc are the post-intervention mean outcome in the treatment group and control group respectively:

$$SMD = \frac{Y_t - Y_c}{S_p}$$

To calculate S_p, the pooled standard deviation (the standard deviation of the outcome variable for both treated (S_t) and control (S_c) individuals), we will use the Hedges' approach described in Lipsey and Wilson (2001), where n_t and n_c are the sample sizes of the treatment and control groups:

$$S_p = \sqrt{\frac{(n_t - 1) * S_t^2 + (n_c - 1) * S_c^2}{n_t + n_c - 2}}$$

The standard error of *g* will be computed using the following formula:

$$SE\ (SMD) = \sqrt{\frac{n_t + n_c}{n_t * n_c} + \frac{SMD^2}{2 * (n_c + n_t)}}$$

Alternatively, in the event that the necessary information for calculating SE is not available, we will approximate it as follows, where *t* is the *t* statistic of the treatment effect:

$$SE\ (SMD) = \frac{SMD}{t}$$

For regression-based studies, we intend to use the following formula, described in Keef and Roberts (2004), where β is the coefficient of interest (i.e. yielding the impact of the intervention) and σ is the standard deviation of the error term in a regression:

$$SMD = \frac{\beta}{\sigma}$$

Where σ is not reported by the study authors (highly likely), we will use the following formula to compute an equivalent using the sample standard deviation of the dependent variable and the sample size for both treatment and control groups:

$$S_p = \sqrt{\frac{(SD_y^2 * (n_t + n_c - 1) - \frac{(\beta^2 * (n_t * n_c))}{n_t + n_c}}{n_t + n_c}}$$

Standard errors will be approximated using the following formula, where *t* is the *t* statistic for the regression coefficient:

$$SE\ (SMD) = \frac{SMD}{t}$$

Our second choice formula for effect size calculations of regression-based studies is based on Snilstveit et al. (2016).

For studies reporting regression coefficients:

$$d = \frac{2t}{\sqrt{n_t + n_c}}\ Var_d = \frac{2}{n_t + n_c} + \frac{d^2}{4(n_t + n_c)}$$

where *n* denotes the sample size of treatment group (*t*) and control (*c*). We will calculate the *t*-statistic by dividing the coefficient by the standard error. If the authors only report confidence intervals and no standard error, we will calculate the standard error from the confidence

intervals. If the study does not report the standard error, but report *t* we will extract and use this as reported by the authors.

CALCULATING RESPONSE RATIOS

For studies using matching, the following formulae to compute *g* and its standard error will be used:

$$RR = \frac{Y_t}{Y_c}$$

The standard error of *g* will be computed using the following formula, where *t* refers to either the *t* statistics/*p*-value of the regression coefficient or to the results of the *t* test for equality of means between groups after matching:

$$SE(RR) = \exp\frac{Ln(RR)}{t}$$

Alternatively, the standard error of *g* can also be computed as:

$$SE(RR) = S_p^2 * \left(\frac{1}{n_t * Y_t^2} + \frac{1}{n_c * Y_c^2}\right)$$

For regression-based studies, the following formula to compute *g* and its standard error will be used:

$$RR = \frac{Ys + \beta}{Yc}$$

The standard error of *g* will be computed using the same formula as presented for matched-based studies.

CONVERSION OF SMDS INTO PERCENTAGES

In general, an effect size of 0.10 reflects one-tenth standard deviation improvement for the treatment group (i.e. learners engaged in the mobile learning intervention) compared to the control group (Petrosino et al 2012). In an attempt to make this measurement scale less abstract, Rosenthal and Rubin (1982) convert the SDM to a percentage expression of the change between experimental groups. Based on this approach, a more intuitive translation of effect sizes is presented below. For example, a SDM of 0.10 represents and improvement of 5% in the intervention group.

Table A7.1 Translation of effect sizes

Effect size (cohen's *d*)	Percentage success (treatment)	Percentage success (control)	Difference in success rates (%)
0.00	50.0	50.0	0
0.10	52.5	47.5	5
0.20	55.0	45.0	10
0.30	57.4	42.6	14.8
0.40	59.8	40.2	19.6
0.50	62.1	37.9	24.2
0.60	64.4	35.6	28.8

Table adapted and updated from Petrosino et al (2010), based on the calculation formula by Rosenthal and Rubin (1982). Note calculation in the meta-analysis are based on *g,* a sample size correct version of *d*. The correction has marginal implications for the above translation.

APPENDIX 8: INTERVENTION DESIGN FEATURES EXTRACTION TOOL

<table>
<tr><td colspan="4">Study:
Intervention:</td></tr>
<tr><td colspan="4">Describe the intervention:
Please include intervention name, coding of intervention category on EPPI-Reviewer, and high-level description of the intervention.</td></tr>
<tr><td colspan="4">Intervention component:
Please include here the detailed breakdown of the intervention according to the different intervention components.</td></tr>
<tr><td>Component 1
E.g.: TVET course for six weeks</td><td>Component 2
E.g.: Employer voucher to cover 20% of the salary financed by government</td><td>Component 3</td><td>Component 4 (add more if needed)</td></tr>
<tr><td colspan="4">Mechanism(s) of change (per component if applicable):
Please include here how the designers of the intervention assumed it to lead to a positive change. For example, a training programme assumes that attendees will have skills that are needed by employers. The mechanism of change here is employment relevant knowledge and skills.</td></tr>
<tr><td>Mechanism 1
E.g. acquisition of labour-relevant skills</td><td>Mechanism 2
E.g. signalling of skills to employer</td><td>Mechanism 3</td><td>Mechanism 4 (add more if needed)</td></tr>
<tr><td colspan="4">Context(s) of change (per component if applicable):
Please include here how the immediate changes induced by the intervention can alter or be altered by contextual variables. These contexts need to be relevant to the intervention and its design itself. We are not interest in the wider context in which the study is conducted, e.g. urban areas or a period of economic growth. An example of a relevant context factor would be that women cannot use public transport and that the interventions provided a bus for the women to attend training sessions.</td></tr>
<tr><td>Context 1</td><td>Context 2</td><td>Context 3</td><td>Context 4 (add more if needed)</td></tr>
<tr><td>E.g. restricted mobility of women</td><td>E.g. stereotypes</td><td></td><td></td></tr>
<tr><td colspan="4">Possible areas for design features
Please consider the below areas of design features. Is any information provided on them that might be important? Do not feel you need to follow these classifications. These are just some areas to get you thinking and please add to them as you see fit.</td></tr>
</table>

How is the intervention delivered, where, when?	Who designed it, implemented it, financed it (e.g. public vs private sector; gender of the implementers; technical know-how of the implementers)?	How were people made aware of the intervention/selected into it? Was any form of incentives used?	What level of social organisation is targeted (e.g. individual, household, community? Is the intervention available to only women? Specify gender access. Who is the target group of the intervention?
Was there a role of M&E or other feedback mechanisms? Was there any evidence of implementation failures and lessons learned?	Were user fees charged or did the private sector contribute to the costs? Was there an emphasis on market-demand driving the intervention design?	How intensive was the 'dosage' of the intervention, including frequency of activities, interactions, size of provided resources, etc?	What is the intervention setting (e.g. classroom vs distance learning, class sizes, workplace, online etc)?
Any evidence of *gender-sensitive design*, e.g. quota for women, tailoring interventions around women's needs such as child care?	Any evidence of the use of *behavioural sciences (i.e. addressing cognitive or social determinants of economic behaviours)*, e.g. nudges in intervention design?	Any evidence of *participatory approaches* in the intervention design?	Any evidence of the intervention paying attention to *social or professional norms* around female labour market participation?
Any other notable design aspects to mention?	Any other notable design aspects to mention?	Any other notable design aspects to mention?	Any other notable design aspects to mention?
In conclusion, what do you think are the *distinguishing design features* of this intervention if any:			

APPENDIX 9: GRADE EVIDENCE PROFILE

	Quality assessment						GRADE Result
Intervention category (outcomes)	No. of studies (design)	Limitations	Inconsistency	Indirectness	Imprecision	Pooled effect	Quality
Combined training and placements							
Wage Labour	8 (5 RCTs)	Serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	0.159 (0.09, 0.23)	⊕⊕⊕◯ Moderate
Income	9 (5 RCTs)	Serious risk of bias	No serious inconsistency	No serious indirectness	No serious imprecision	0.145 (0.07, 0.22)	⊕⊕⊕◯ Moderate
Empowerment	5 (2 RCTs)	Serious risk of bias	No serious inconsistency	Serious indirectness	No serious imprecision	Narrative synthesis	⊕⊕◯◯ Low
Soft skills training on promotion							
Career progression	2 (2 RCTs)	Very serious risk of bias	No serious inconsistency	Serious indirectness	No serious imprecision	Narrative synthesis	⊕◯◯◯ Very low
Empowerment	2 (2 RCTs)	Very serious risk of bias	No serious inconsistency	Serious indirectness	No serious imprecision	Narrative synthesis	⊕◯◯◯ Very low
Job placement services only							
Wage Labour	3 (3 RCTs)	No serious limitation	Very serious inconsistency	No serious indirectness	No serious imprecision	Narrative synthesis	⊕⊕◯◯ Low
Income	3 (3 RCTs)	No serious limitation	Very serious inconsistency	No serious indirectness	No serious imprecision	Narrative synthesis	⊕⊕◯◯ Low
Empowerment	3 (3 RCTs)	No serious limitation	Very serious inconsistency	No serious indirectness	No serious imprecision	Narrative synthesis	⊕⊕◯◯ Low
National labour subsidies							
Wage Labour	2 (0 RCTs)	Very serious risk of bias	No serious inconsistency	No serious indirectness	Serious imprecision	Narrative synthesis	⊕◯◯◯ Very low
Macro-level empowerment policies							
Wage Labour	2 (0 RCTs)	Very serious risk of bias	Serious inconsistency	No serious indirectness	No serious imprecision	Narrative synthesis	⊕◯◯◯ Very low

	Quality assessment						GRADE Result
Intervention category (outcomes)	No. of studies (design)	Limitations	Inconsistency	Indirectness	Imprecision	Pooled effect	Quality
Empowerment	2 (0 RCTs)	Very serious risk of bias	Serious inconsistency	No serious indirectness	No serious imprecision	Narrative synthesis	⊕◯◯◯ Very low

APPENDIX 10: LIST OF INCLUDED STUDIES

Adhvaryu A, Kala N, Nyshadham A (2016). Soft skills to pay the bills: Evidence from female garment workers. Working Paper. London: Department for International Development. http://jobsanddevelopmentconference.org/wp-content/uploads/2016/10/ADHVARYU-Namrata-Kala-Anant-Nyshadham-PACE_aug2016.pdf.

Adoho F, Chakravarty S, Korkoyah DT, Lundberg M, Tasneem A (2014). The impact of an adolescent girls' employment program: The EPAG project in Liberia. Policy Research Working Paper Series: 6832. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/610391468299085610/The-impact-of-an-adolescent-girls-employment-program-the-EPAG-project-in-Liberia.

Ayhan SH (2013). Do non-wage cost rigidities slow down employment? Evidence from Turkey. IZA Journal of Labor Policy, 2(1): 1-23.

Broecke S (2013). Tackling graduate unemployment through employment subsidies: A look at the SIVP programme in Tunisia. IZA Journal of Labor Policy, 2(9): DOI: 10.1186/2193-9004-2-9.

Chakravarty S, Lundberg M, Nikolov P, Zenker J (2016). The role of training programs for youth employment in Nepal: Impact evaluation report on the employment fund. Policy Research Working Paper Series: 7656. Washington, DC: World Bank. http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-7656.

Chen YJ, Chindarkar N (2017). The value of skills: Raising the socioeconomic status of rural women in India. Development Policy Review, 35(2): 229–261.

de Azevedo TA, Davis J, Charleset M (2013). Testing 'what works in youth employment': Evaluating Kenya's Ninaweza program. The Global Partnership for Youth Employment, Volume 1: A Summative Report. https://www.iyfnet.org/sites/default/files/library/GPYE_KenyaImpactEval_V1.pdf.

de Azevedo TA, Davis J, Charleset M (2013). Testing 'what works in youth employment': Evaluating Kenya's Ninaweza program. The Global Partnership for Youth Employment, Volume 2: Appendices. (Linked study)

Elías VJ, Núñez FR, Cossa R, Bravo D (2004). An econometric cost-benefit analysis of Argentina's youth training program. Research Network Working Paper: #R-482. Washington, DC: Inter-American Development Bank. https://ideas.repec.org/p/idb/wpaper/3174.html.

Galdo J, Jaramillo M, Montalva V (2008). Household wealth and heterogeneous impacts of a market-based training program: The case of PROJOVEN in Peru. PMMA Working Paper: 2008-02. Nairobi: Poverty and Economic Policy (PEP) Research Network. https://ideas.repec.org/p/lvl/pmmacr/2008-02.html.

Ghani E, Kerr WR, O'Connell SD (2014). Political reservations and women's entrepreneurship in India. Journal of Development Economics, 108: 138-153.

Groh M, Krishnan N, McKenzie D (2012). Soft skills or hard cash? The impact of training and wage subsidy programs on female youth employment in Jordan. Policy Research Working Paper Series: 6141. Washington, DC: World Bank. http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-6141.

Groh, M., Krishnan, N., McKenzie, D. and Vishwanath, T., 2016. The impact of soft skills training on female youth employment: evidence from a randomized experiment in Jordan. IZA Journal of Labor & Development, 5(1), p.9. (Linked study)

Groh, M., Krishnan, N., McKenzie, D. and Vishwanath, T., 2016. Do wage subsidies provide a stepping-stone to employment for recent college graduates? Evidence from a randomized experiment in Jordan. Review of Economics and Statistics, 98(3), pp.488-502. (Linked study)

Groh M, McKenzie D, Shammout N, Vishwanath T (2014). Testing the importance of search frictions, matching, and reservation prestige through randomized experiments in Jordan. Policy Research Working Paper Series: 7030. Washington, DC: World Bank. http://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-7030.

Groh, M., McKenzie, D. and Vishwanath, T., 2015. Reducing information asymmetries in the youth labor market of Jordan with psychometrics and skill based tests. *The World Bank Economic Review*, *29*(suppl_1), pp.S106-S117. (Linked study)

Hallward-Driemeier MC, Gajigo O (2013). Strengthening economic rights and women's occupational choice: The impact of reforming Ethiopia's family law. Policy Research Working Paper: 6695. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/259861468021600567/Strengthening-economic-rights-and-womens-occupational-choice-the-impact-of-reforming-Ethiopias-family-law.

Honorati M (2015). The impact of private sector internship and training on urban youth in Kenya. Policy Research Working Paper Series: 7404. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/468501468185042536/The-impact-of-private-sector-internship-and-training-on-urban-youth-in-Kenya.

Jensen R (2010). Economic opportunities and gender differences in human capital: Experimental evidence for India. Working Paper: 16021. Cambridge, MA: National Bureau of Economic Research, http://www.nber.org/papers/w16021.

Macchiavello R, Menzel A, Rabbani A, Woodruff C (2015). Challenges of change: An experiment training women to manage in the Bangladeshi garment sector. Coventry: University of Warwick, Department of Economics.

Medina CD, Escobar JR (2007). The effects of changes in the legal work shift on wages and hours worked in Colombia. Coyuntura Social, 37: 63-87.

Ñopo H, Robles M, Saavedra J (2007). Occupational training to reduce gender segregation: The impacts of ProJoven. Research Department Working paper: 623. Washington, DC: Inter-American Development Bank https://publications.iadb.org/handle/11319/3331.

Pfutze T, Rodríguez-Castelán C (2015). Can a small social pension promote labor force participation? Evidence from the Colombia Mayor program. Policy Research Working Paper: 7516. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/577061468197093248/Can-a-small-social-pension-promote-labor-force-participation-evidence-from-the-Colombia-Mayor-program.

APPENDIX 11: SUMMARY TABLE OF THE INCLUDED STUDIES

Author	**Study design** *Study design* *Risk of bias rating*	**Intervention** *Programme name* *Categorization*	**Context** *Region & Countr:* *UN Category* *GAP Index (rank)* *GII Index (rank)*	**Population** *No. (of women)* *UN Category* *Transition*	**Industry sector**	**Outcomes** *Wage employment*	*Income & poverty measures*	*Women's empowerment*	*Cost data / calculations*
Adhvaryu (2016)	RCT Probit regression Low risk of bias	Personal Advancement and Career Enhancement (P.A.C.E.) programme Soft skills training (for career progression)	Asia: India UN: Declining fertility, formalising economies GAP: 0.683 (#87) GII: 0.530 (#125)	N: 2,706 UN: Unclear Transition: Within-profession promotion	Manufacturing	✔		✔	✔
Adoho (2014)	RCT DID OLS Regression Low risk of bias	Economic Empowerment of Adolescent Girls and Young Women (Adolescent Girls Initiative) TVET training + Life skills training + Job placement	Africa: Liberia UN: High fertility, agrarian societies GAP: 0.652 (#114) GII: 0.649 (# 150)	N: 2,106 UN: Unclear Transition: From school to work (including out-of-school) From unemployment to employment	Hospitality; Professional cleaning/Waste management; ICT; Professional house/ Office painting; Security guard services; Professional driving.	✔	✔	✔	✔
Ayhan (2013)	Quasi-experimental: retrospective DDD DDD regression High risk of bias	Labour law: Employment Package Law No. 5763 2008. Demand-side macro-economic subsidy (reduction in employers' social security contributions)	Asia: Turkey UN: Ageing societies GAP: 0.623 (#130) GII: 0.328 (#69)	N: 305,590 observations UN: Unclear Transition: From school to work (including out-of-school)	Services; Industry; Agriculture; Construction	✔			

Author	Study design	Intervention	Context	Population	Industry sector	Outcomes			
	Study design *Risk of bias rating*	*Programme name* *Categorization*	*Region & Countr:* *UN Category* *GAP Index (rank)* *GII Index (rank)*	*No. (of women)* *UN Category* *Transition*		*Wage employment*	*Income & poverty measures*	*Women's empowerment*	*Cost data / calculations*
				From unemployment to employment					
Broecke (2013)	Quasi-experimental: retrospective PSM OLS & PSM regression High risk of bias	Labour law: The Stage d'Initiation à la Vie Professionelle (SIVP) programme Supply- & demand-side macro-economic subsidy (employees' salaries paid by government & reduction in employers' social security contributions)	Africa: Tunisia UN: Ageing societies GAP: 0.636 (#126) GII: 0.289 (# 58)	N: 2,121 UN: Unclear Transition: From school to work (including out-of-school) From unemployment to employment	Law; Applied languages; Chemistry; Biology; Water treatment, agriculture & food industry; Physics; Engineering; Business; Health & social services; Mathematics	✔			✔
Chakravarty (2016)	Quasi-experimental: DID and PSM DDD & PSM regression (ITT) Moderate risk of bias	The Employment Fund and the Adolescent Girls Employment Initiative (Adolescent Girls Initiative) TVET training + Life skills training + Job placement	Asia: Nepal UN: Declining fertility, formalising economies GAP: 0.661 (#110) GII: 0.497 (#115)	N: 2,814 UN: Poor women Transition: From school to work (including out-of-school) From unemployment to employment	Farming; Poultry technician; Food preparation & hospitality; Electricians & electronics; Handicraft & incense stick making; Construction; Beautician/barber;	✔	✔	✔	

Author	Study design	Intervention	Context	Population	Industry sector	Outcomes			
	Study design *Risk of bias rating*	*Programme name* *Categorization*	*Region & Countr:* *UN Category* *GAP Index (rank)* *GII Index (rank)*	*No. (of women)* *UN Category* *Transition*		*Wage employment*	*Income & poverty measures*	*Women's empowerment*	*Cost data / calculations*
					Weaving, tailoring & garment making; Textile, carpentry, welding and masonry				
Chen (2017)	Quasi-experimental: random sampling IV regression High risk of bias	Barefoot Mechanics programme. TVET training + job reservation	Asia: India UN: Declining fertility, formalising economies GAP: 0.683 (#87) GII: 0.530 (#125)	N: 212 UN: Unclear Transition: From unemployment to employment From rural to urban areas	Water mechanics	✔	✔	✔	
de Azevedo (2013)	RCT: Different regression models (ITT) High risk of bias	The Kenya Youth Empowerment Program: Ninaweza TVET training + Life skills training + Job placement	Africa: Kenya UN: High fertility agrarian societies GAP: 0.702 (#63) GII: 0.565 (135)	N: 1230 UN: Unclear Transition: From school to work (including out-of-school) From unemployment to employment	ICT	✔	✔	✔	
Elías (2004)	Quasi-experimental: PSM DID; PSM regression	The Youth Training Program (PROJOVEN) TVET training + Job placement	Latin America and Caribbean: Argentina UN: Declining fertility, formalising economies	N: 8695 UN: Unclear Transition: From school to	Agriculture; Industry; Services; Construction.	✔			✔

Author	Study design	Intervention	Context	Population	Industry sector	Outcomes			
	Study design *Risk of bias rating*	*Programme name* *Categorization*	*Region & Countr:* *UN Category* *GAP Index (rank)* *GII Index (rank)*	*No. (of women)* *UN Category* *Transition*		*Wage employment*	*Income & poverty measures*	*Women's empowerment*	*Cost data / calculations*
	Moderate risk of bias		GAP index (#33): 0.735 GII index (#77): 0.362	work (including out-of-school) From unemployment to employment					
Galdo (2008)	Quasi-experimental: random sampling Quantile regression High risk of bias	Youth Training Program (PROJOVEN) TVET training + Job placement	Latin America and Caribbean: Peru UN: Declining fertility, formalising economies GAP index (#80): 0.687 GII index (#86): 0.385	N: 6548 UN: Unclear Transition: From school to work (including out-of-school) From unemployment to employment	Textiles and apparel; Services, mechanics and metalworking; Construction, carpentry and shoemaking.	✔	✔		✔
Ghani (2014)	Natural experiment OLS & IV regression High risk of bias	National law: Political reservations for women in India	Asia: India UN: Declining fertility, formalising economies GAP index (#87): 0.683 GII index (#125): 0.530	N: 3,606 UN: Unclear Transition: From unemployment to employment From informality to formality	Manufacturing sector.	✔		✔	

Author	Study design	Intervention	Context	Population	Industry sector	Outcomes			
	Study design *Risk of bias rating*	*Programme name* *Categorization*	*Region & Countr:* *UN Category* *GAP Index (rank)* *GII Index (rank)*	*No. (of women)* *UN Category* *Transition*		*Wage employment*	*Income & poverty measures*	*Women's empowerment*	*Cost data / calculations*
Groh (2012)	RCT: OLS Regression (ITT) Low risk of bias	The Jordan New Opportunities for Women (Jordan NOW) (Adolescent Girls Initiative) Soft skill training + Wage subsidy voucher	Asia: Jordan UN category: declining fertility, formalising economies GAP index (#134): 0.603 GII index (#111): 0.478	N: 1350 UN: Unclear Transition: From school to work (including out-of-school) From unemployment to employment	Teaching; Business administration (accounting and finance); Medical assistance (nursing & pharmacy).	✔	✔	✔	✔
Groh (2014)	RCT: OLS Regression (ITT) Low risk of bias	Experiment 1: Jordan NOW 2.0 Screening & matching services Experiment 2: Yalla Watheefa (Let's Go Get a Job) Screening & matching services	Asia: Jordan UN: Declining fertility, formalising economies GAP index (#134): 0.603 GII index (#111): 0.478	N: 1,538 UN: Unclear Transition: From school to work (including out-of-school) From unemployment to employment	IT; Accounting; Services (retail & marketing); Manufacturing; Education;	✔	✔		✔

Author	**Study design** *Study design* *Risk of bias rating*	**Intervention** *Programme name* *Categorization*	**Context** *Region & Countr:* *UN Category* *GAP Index (rank)* *GII Index (rank)*	**Population** *No. (of women)* *UN Category* *Transition*	**Industry sector**	**Outcomes** *Wage employment*	*Income & poverty measures*	*Women's empowerment*	*Cost data / calculations*
Hallward-Driemeier (2013)	Natural experiment (DID) DID regression Moderate risk of bias	National Law: Family Code Economic rights (making employment decisions without the consent of the husband) Increased property rights for women	Africa: Ethiopia UN: Declining fertility, formalising economies GAP index (#109): 0.662 GII index (#116): 0.499	N: 172 UN: Unclear Transition: From informality to formality	Human-capital-intensive occupations (index)	✔		✔	
Honorati (2015)	RCT OLS regression (ITT & ATT) High risk of bias	The Kenya Youth Empowerment Program: Ninaweza TVET training + Life skills training + Job placement	Africa: Kenya UN: High fertility, agrarian society GAP index (#63): 0.702 GII index (#135): 0.565	N: 558 UN: Unclear Transition: From school to work (including out-of-school) From unemployment to employment	Energy; Finance; ICT; Manufacturing; Tourism; *Juakali* (informal sector).	✔	✔		✔
Jensen (2010)	RCT OLS Regression Low risk of bias	Business outsourcing recruiting services Job search assistance and placement support	Asia: India UN: Declining fertility, formalising economies GAP index (#87): 0.683 GII index (#125): 0.530	N: 3200 households (individuals not specified) UN: Unclear Transition: From school to	Business process outsourcing industry (includes activities such as call centres, data entry and management, claims processing, secretarial services, transcription, online	✔	✔	✔	

Author	Study design	Intervention	Context	Population	Industry sector	Outcomes			
	Study design *Risk of bias rating*	*Programme name* *Categorization*	*Region & Countr:* *UN Category* *GAP Index (rank)* *GII Index (rank)*	*No. (of women)* *UN Category* *Transition*		*Wage employment*	*Income & poverty measures*	*Women's empowerment*	*Cost data / calculations*
				work (including out-of-school) From unemployment to employment	technical support, accounting and software development)				
Macchiavello (2015)	RCT ITT & OLS regression High risk of bias	Management skills training (for career progression)	Asia: Bangladesh UN: Declining fertility, formalising economies GAP index (#72): 0.698 GII index (#119): 0.520	N: 100 UN: Unclear Transition: Within-profession promotion	Garment industry.	✔		✔	
Medina (2007)	Quasi-experimental: retrospective cross-sectional DID regression High risk of bias	Labour law: Change in working hours regulation (Law 789 of 2002)	Latin America and Caribbean: Colombia UN: Declining fertility, formalising economies GAP index (#39): 0.727 GII index (#89): 0.393	N: 3,020,880 total observations; number of women not specified. UN: Unclear Transition: No information	Service; Trade; Industry; Financial institutions; Farming; Mining; Construction; Transport & telecommunications.	✔			

Author	**Study design** *Study design* *Risk of bias rating*	**Intervention** *Programme name* *Categorization*	**Context** *Region & Countr:* *UN Category* *GAP Index (rank)* *GII Index (rank)*	**Population** *No. (of women)* *UN Category* *Transition*	**Industry sector**	**Outcomes** *Wage employment*	*Income & poverty measures*	*Women's empowerment*	*Cost data / calculations*
Ñopo (2007)	Quasi-experimental: PSM Mean comparison Moderate risk of bias	Youth labour training (PROJOVEN) TVET training + Job placement	Latin America and Caribbean: Peru UN: Declining fertility, formalising economies GAP index (#80): 0.687 GII index (#86): 0.385	N: 1,383 UN: Unclear Transition: From school to work (including out-of-school) From unemployment to employment	Occupational segregation by gender based on Duncan Index and measured in three sectors: Sales personnel; Restaurant & food service; Domestic work.	✔	✔	✔	
Pfutze (2015)	Quasi-experimental: retrospective cross-sectional (IV) OLS regression (IV) High risk of bias	Colombia Mayor Program Changes to municipal social pensions	Latin America and Caribbean: Colombia UN: Declining fertility, formalising economies GAP index (#39): 0.727 GII index (#89): 0.393	N: 13,077 UN: Very poor women Transition: No information	Agriculture and related; Manufacturing; Trade & commerce; Services & tourism.	✔			

APPENDIX 12: RISK OF BIAS RESULTS PER STUDY

Title	Sampling bias	Bias due to confounding	Failed randomisation	Departures from intended intervention	Attrition bias	Outcome measurement bias	Selective reporting	Overall risk of bias
Adhvaryu (2016)			1	1	1	1	1	Low
Adoho (2014)			1	1	1	1	1	Low
Ayhan (2013)	2	3		2	3	2	1	High
Broecke (2013)	3	3		1	2	2	3	High
Chakravarty (2016)	2	2		2	2	1	1	Moderate
Chen (2017)	3	3		1	3	1	1	High
de Azevedo (2013)			1	3	3	1	1	High
Elías (2004)	2	2		2	3	1	1	Moderate
Galdo (2008)	3	2		3	1	2	1	High
Ghani (2014)	2	3		2	3	1	1	High
Groh (2012)			1	2	1	1	1	Low

Title	Sampling bias	Bias due to confounding	Failed randomisation	Departures from intended intervention	Attrition bias	Outcome measurement bias	Selective reporting	Overall risk of bias
Groh (2014)			1	1	1	1	1	Low
Hallward-Driemeier (2013)	2	2		1	1	2	2	Moderate
Honorati (2015)			3	2	3	2	3	High
Jensen (2010)			1	2	1	2	2	Low
Macchiavello (2015)			3	3	3	3	3	High
Medina (2007)	3	3		3	3	2	1	High
Ñopo (2007)	2	2		2	3	1	1	Moderate
Pfutze (2015)	3	3		3	3	2	3	High

1	low RoB
2	moderate RoB
3	high RoB

APPENDIX 13: NARRATIVE SUMMARIES OF INDIVIDUAL STUDIES NOT INCLUDED IN THE SYNTHESIS

Medina CD, Escobar JR (2007). The effects of changes in the legal work shift on wages and hours worked in Colombia. *Coyuntura Social*, 37: 63-87.

Context
This paper focuses on Colombia, an upper-middle-income country. In view of the fact that Colombia was still experiencing the remnants of a deep labour market crisis, the Colombian Congress passed a reform to the labour market regulation called Law 789 in 2002. This law focused, among other aspects, on the change in a number of daytime hours of work. It aimed to cover the whole country. Moreover, the law was subject to a two-year trial period, after which it would be altered or withdrawn if it did not produce positive results.

Population
The aim of the study was to sustain the active labour market policies and to ensure that the existing labour regulations became more flexible. The law aimed to alter the hourly wages and work hours of employees working in the formal sector. In an attempt to estimate the effects of this law, the authors had to introduce a treatment and control group. Selection criteria included age (under and over 25), gender (both male and female) and education levels (those who had finished their academic formation compared to those who had not). The paper mentioned that the workers in the formal sector represented 33% of the total labour force and that only 53% of the formal sector would be affected by the reform (treatment group), whereas 47% would represent the comparison group, which would not be affected by the reform. The eligibility criteria included people working between 6:00 p.m. and 10:00 pm, people working on Sundays or holidays, workers with a fixed hourly work shift and workers from the formal sector of the economy, as well as the vulnerable populations in society, for instance, the unemployed or the economically inactive. The control group comprised workers in either managerial or directive occupations as well as those who had a position of trust within the formal sector of an organisation.

Intervention
The active labour market policies included incentives for hiring hard-to-employ workers, a reduction of firing costs as well as the introduction of unemployment insurance and employment subsidies. Furthermore, the reform aimed to make labour regulation more flexible by increasing daytime working hours, reducing overtime pay for working on Sundays and national holidays and allowing for the possibility of making work shifts more flexible.

Outcomes and findings
The outcomes that were estimated in this study were the effects of legislative amendments on weekly working hours and hourly wages. After the law was changed, the night hours became standard hours and overtime night hours became regular overtime hours. A difference by gender was also recorded. The main findings (based on robust evidence) were that the change in law resulted in a decline in the hourly wage. The hourly wages for males (over 25 years old) working in the manufacturing sector in metropolitan areas reduced by over 11% because of the change in law. For females over 25, there was a

reduction of 3.6 hours of work per week for the treatment group in manufacturing. For females under 25, the reform had insignificant outcomes on hourly wages and hours per week.

There were additional findings are based on less robust evidence. Overall, the change in law seems to have had a positive outcome on all the affected employees with the exception of those employed in the manufacturing sector. The reform also led to an increase in hourly wages: the treatment hourly wages increased more than the comparison group, probably as a result of demand for labour. Male workers (above 25 years) worked 5.7 more hours per week because of the reform. There was an increase of up to 8% in the hourly wages of males over 25 who were not employed in the manufacturing sector. In the manufacturing sector, the hours worked increased by up to 3.2 hours per week.

Pfutze T, Rodríguez-Castelán C (2015*). Can a small social pension promote labor force participation: Evidence from the Colombia Mayor Program?* Policy Research Working Paper 7516: Washington, DC: World Bank.

Context
Colombia is an upper middle-income country according to the World Bank. The Colombia Mayor Program was established in 2003 with the main intention of reducing poverty in the country through cash transfers, which were anticipated to ease the liquidity in the economy. The purpose was to offer basic subsistence through non-contributory pension benefits to the senior citizens that did not have access to retirement funds as well as the elderly people who were living in conditions of extreme poverty.

Population
This intervention was available to rural and urban areas (mixed). There are three main programme eligibility criteria, namely age, household score and income. In terms of age, the beneficiary must not be within three years of reaching the official retirement age (which was 52 years for women and 57 years for men). The beneficiary's household must not score above a certain threshold in Colombia's system for the identification of potential social programme beneficiaries, and a beneficiary's income may not go above half the minimum wage. In order to study the effects of the programme the authors selected a sample that consisted of the following: the first age group was 52 or older for women and 57 or older for men, which was the minimum age of eligibility; the other age group was 60 to 65. The sample had more men than women, mostly amongst the relatively younger group. About half of the prospective recipients were married, whereas 24% were widowed, 14% were divorced and only 11% were single. In terms of education levels, around two-thirds had at least completed primary school, while only 9% had completed secondary education, with 2% having gone beyond secondary education; 13% were unfit and unable to work.

Intervention
Colombia Mayor provided non-contributory pension benefits as well as several indirect non-monetary benefits (the latter are unspecified in the paper). The amount of transfer was adjusted annually depending on the funds available. The money that was given out was not supposed to be more than half the value of the minimum wage. A specific number of places for new recipients were allocated to each municipality. However, if the number of eligible petitioners surpassed the number of offered slots, the intervention made use of a prioritisation strategy to determine who would receive the benefit.

Outcomes and findings

The outcome measured for the intervention was labour force participation. Overall, there was an increase in labour force participation amongst men and not women; the explanation suggested for this was that unconditional cash transfer programmes assist beneficiaries by making it easier for them to access funds, which will in turn boost economic activity. Men were now working less in manufacturing, whereas women were moving out of the primary sector and working less on agricultural land. The results generated by the intervention were much smaller in magnitude for the women, which indicates that it was statistically insignificant.

Overall, the results show that the programme had a positive outcome on other adults in the household, comparable to its outcome on the actual beneficiaries. The programme had the effect of increasing labour force participation for relatively young beneficiaries, i.e. men in their 50s and 60s. In addition, there were positive and statistically significant results for working alone, for working on agricultural land, rivers or the sea, and for working independently or working on one's own land. With regard to extent, it was found that such males were 13.8 percentage points more likely to work alone, 9.3 percentage points more likely to work on agricultural land, 14.7 percentage points more likely to work independently, and 6.5 percentage points more likely to work on their own land.

APPENDIX 14: PUBLICATION BIAS

As indicated in section 3.5.7, we identified an insufficient number of studies to assess the overall publication bias in the evidence base. In the review, we only identified a homogeneous grouping of interventions suitable for meta-analyses in reference to interventions combining training and placement services. This intervention grouping was based on nine studies yielding eight effect sizes. We provide the funnel plot for this meta-analysis below. On visual inspection, the plot suggests that there might be a modest prevalence of publication bias regarding studies with larger negative effect sizes and higher standard errors.