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

Janice Tripney, Jorge Hombrados, Mark Newman, Kimberly Hovish, Chris Brown, Katarzyna Steinka-Fry, Eric Wilkey

THE CAMPBELL COLLABORATION

| Corresponding author | Janice Tripney
 EPPI-Centre, Social Science Research Unit
 Institute of Education, University of London
 18 Woburn Square
 London WC1H 0NR
 United Kingdom
 E-mail: j.tripney@ioe.ac.uk |

Campbell Systematic Reviews

Table of contents

Executive summary

BACKGROUND

The increase of low-income, low-skilled youth in the labour market, particularly in developing countries, is a major concern internationally. In some regions of the world, young people are nearly three times as likely as adults to be unemployed. They are also more likely to work in the informal labour market than adults, in low quality jobs that offer limited socio-economic security, training opportunities, and working conditions. This enormous unlocked potential represents a substantial loss of opportunity for both individuals and society. With increasing emphasis being given to work- and skills-based solutions to economic competition and poverty in the developing world, comes a renewed focus on technical and vocational education and training (TVET) as a means to expand opportunities for marginalised youth. Although several reviews have attempted to summarise the existing research in this area, there are a number of limitations to these reviews. There is a need to systematically examine the evidence base to provide a picture of the types of TVET interventions being used to raise employment, to identify those that are effective and ineffective, and to identify areas in which more research needs to be conducted.

OBJECTIVES

The main objective of this systematic review was to summarise the available evidence on the effects of TVET interventions for young people in developing countries to inform policy, practice, and research.

The questions guiding this study were:

- What are the effects of different models of technical and vocational education and training (TVET) interventions on the employment and employability outcomes of young people, aged 15-24 years, in low- and middle-income countries?
- What do the findings suggest about moderating effects?

SEARCH STRATEGY

A systematic and comprehensive search was used to locate both published and unpublished studies. A wide range of major bibliographic databases were electronically searched, along with specialist and grey literature databases, and websites of relevant organisations. Reference lists of previous reviews and included studies were examined. In addition, we conducted forward citation checking exercises and attempted contact with authors and other relevant stakeholders.

SELECTION CRITERIA

Studies eligible for inclusion in the review were required to meet several eligibility criteria. First, studies must have evaluated a TVET intervention. Second, studies must have investigated outcomes for young people aged 15-24 years. Third, the geographical location of the studies must have been a low-or middle-income country. Fourth, studies must have utilised an experimental or quasi-experimental research design, including random assignment, quasi-random assignment (and groups generated were shown to be equivalent, or there was sufficient information to permit calculation of pre-treatment group equivalence), non-random assignment with matching, or non-random assignment with statistical controls. Fifth, studies must have reported at least one eligible outcome variable measuring employment (e.g., gaining paid employment) or employability (e.g., changing attitudes to work, or gaining job search skills). Finally, the date of publication or reporting of the study must have been between 2000 and 2011. No language restrictions were applied.

DATA COLLECTION AND ANALYSIS

The electronic literature search yielded a total of 8072 potentially relevant reports, 145 of which were retrieved for full-text screening and nine were judged relevant. Handsearching identified a further 46 eligible reports. A total of 30 studies, reported in 55 publications, met the eligibility criteria. However, due to resources limitations, four of the eight eligible Spanish language papers we identified were not included in the review. Of the 26 studies included in the review, 3 utilised a randomised controlled trial (RCT) design, and 23 utilised a quasi-experimental design. The studies were coded independently by pairs of reviewers using a structured coding tool. Descriptive analysis was undertaken to examine and describe data related to the characteristics of the included studies and interventions. Ten of the 26 studies had data that allowed calculation of effects sizes. The findings from these 10 studies were statistically combined using meta-analytic techniques. The effect sizes were calculated using the standardised mean difference, corrected for small sample bias (i.e., Hedges' g). Analysis of the mean effect size, the heterogeneity of effect sizes, and the relationship between effect size and characteristics of the studies, participants and interventions was conducted.

RESULTS

The 26 included studies assessed the effectiveness of 20 different TVET interventions from various countries in Latin America, the Caribbean, Europe, East Asia, South Asia and Sub-Saharan Africa. Publication dates ranged between 2001 and 2011. Study settings included ten upper-middle income countries (Argentina, Bosnia and Herzegovina, Brazil, Chile, China, Colombia, Dominican Republic, Latvia, Mexico, Panama and Peru); two lower-middle income countries (India and Bhutan); and one low-income country (Kenya).

The following summary of evidence focuses on the results of the statistical analyses of 10 studies included in the review.

Employment

- The overall mean effect of TVET on paid employment was positive and significant; however, significant heterogeneity was observed ($Q = 23.8$; df = 7; p = 0.00124; $I^2 = 70.6\%$; $tau^2 = 0.0153$).

- Four variables were tested for moderating effects.

- Evidence of a statistically significant relationship between study quality and effect size was observed ($Q_b = 6.49$; p = 0.0108). It is reasonable, therefore, to conclude that the overall mean effect may be inflated and that our conclusions about treatment effect on paid employment should be based only on those studies rated medium quality (g=0.06; 95% CI [-0.01, 0.12]).

- No significant differences in mean effects were observed between studies according to type of TVET intervention ($Q_b = 1.43$; p = 0.231), length of follow-up period ($Q_b = 0.273$; p = 0.601), or gender ($Q_b = 2.1$; p = 0.147).

Formal employment

- The overall mean effect of TVET on formal employment was positive and significant; however, significant heterogeneity was observed ($Q = 11.1$; df = 4; p = 0.0256; $I^2 = 63.9\%$; $tau^2 = 0.0131$).

- One variable was tested for moderating effects.

- Evidence of a statistically significant relationship between study quality and effect size was observed ($Q_b = 10.6$; p = 0.00116). It is reasonable, therefore, to conclude that the overall mean effect may be inflated and that our conclusions about treatment effect on formal employment should be based only on those studies rated medium quality (g=0.12; 95% CI [0.05, 0.19]).

Monthly earnings

- The overall mean effect of TVET on earnings was positive and significant; however, significant heterogeneity was observed (Q = 25.5; df = 8; p = 0.00128; I^2 = 68.6%; tau^2 = 0.00815).

- Four variables were tested for moderating effects.

- No evidence of a statistically significant relationship between study quality and effect size was observed (Q_b = 0.204; p = 0.652. It is reasonable, therefore, to conclude that the overall mean effect is not inflated and that our conclusions about treatment effect on monthly earnings should be based on all studies in the analysis (*g*=0.127; 95% CI [0.043, 0.21]).

- No statistically significant differences in mean effects were observed between studies according to type of TVET intervention (Q_b = 0.397; p = 0.529), length of follow-up period (Q_b = 0.186; p = 0.666), or gender (Q_b = 1.26; p = 0.262).

Self-employment earnings

- The overall mean effect of TVET on self-employment earnings was negative and non-significant (*g*=-0.025, 95% CI [-0.11, 0.061]). No significant heterogeneity was observed (Q = 0.206; df = 1; p = 0.65; I^2 = 0%; tau^2 = 0). This analysis was based on two medium quality studies.

- One variable was tested for moderating effects.

- No significant differences in mean effects were observed between studies according to gender (Q_b = 1.27; p = 0.259).

Weekly hours worked

- The overall mean effect of TVET on number of weekly hours worked was positive but non-significant. No significant heterogeneity was observed (Q = 1.8; df = 5; p = 0.876; I^2 = 0%; tau^2 = 0).

- Four variables were tested for moderating effects.

- No evidence of a statistically significant relationship between study quality and effect size was observed (Q_b = 1.41; p = 0.234). It is reasonable, therefore, to conclude that the overall mean effect is not inflated and that our conclusions about treatment effect on weekly hours should be based on all studies in the analysis (*g*=0.043; 95% CI [-0.017, 0.104]).

- Statistically significant differences in mean effects were observed between studies according to gender (Q_b = 10.1; p = 0.00151). Treatment effects for female youth were positive, *g*=0.16 (95% CI [0.04, 0.28]), while those for male youth were negative, *g*=-0.09 (95% CI [-0.2, 0.01]).

- No significant differences in mean effects were observed between studies

according to type of TVET intervention ($Q_b = 0.0677$; $p = 0.795$), or length of follow-up period ($Q_b = 0.109$; $p = 0.741$).

AUTHORS' CONCLUSIONS

The studies included in this systematic review represent the best empirical evidence currently available for the impact of TVET on youth employment outcomes. As the review improves upon prior work by statistically synthesising TVET intervention research, its findings strengthen the evidence base on which current policies and practices can draw. That being said, interpreting the evidence and drawing out the implications for policy and practice is nonetheless challenging.

Although this review provides some evidence of the causal impact of TVET on certain labour market outcomes, several limitations of both the included studies and the review itself mean that drawing strong inferences from the results of the analyses is not recommended and caution should be used when applying the findings of the review. A number of additional points are worth emphasising. First, attempts to explain the observed heterogeneity in overall mean effects suggest that methods matter. The low quality studies have consistently larger mean effects than the medium quality studies. In addition, for paid employment, and formal employment, statistically significant differences in mean effects were observed between studies according to study quality, suggesting that the overall mean is inflated and that the treatment effects should be based on the medium quality studies only. Second, effects are generally small and difficult to detect. The mean effects for paid employment (medium quality studies only), self-employment earnings, and working hours are relatively close to zero, and statistically insignificant. The mean effects for formal employment (medium quality studies only) and monthly earnings, although larger, are still relatively small, but they are statistically significant. Third, due to an insufficient number of studies reporting relevant data, only some of the variables for which moderator analyses had been planned a priori could be performed. Of the participant and intervention characteristics that were tested, only one demonstrated a significant relationship with treatment effect.. For weekly hours, statistically significant differences in mean treatment effects were observed between studies according to gender. It would be premature to conclude, however, that there are not in fact real differences between young men and women for other labour market outcomes, or between different types of TVET interventions, or that treatment effects do not diminish over time. We may not have had adequate statistical power to detect moderating effects of the variables tested in this review. There may be other moderating variables that could account for the differences in effects between studies that we were unable to test.

In summary, the existing evidence shows that TVET interventions have some promise. Overall, interventions included in this review were found to demonstrate a small, positive effect on all but one of the employment outcomes measured, with the strength of the evidence being stronger for formal employment and monthly

earnings than for the other outcomes measured. Furthermore, TVET appears to increase the number of hours worked in paid employment by young women but not young men. Thus, it is both important and worthwhile to continue to invest in TVET provision for youth in developing countries. Although, statistically, the overall effects of TVET may be small, even a small increase in the rate of paid employment, for example, could translate into large numbers of young people entering the labour market, where programmes are delivered nationally. A recommendation cannot be made either for, or against, any one type of TVET included in the review. So, in the absence of evidence in support of a particular, and possibly expensive, intervention, opting for the cheapest and/or most culturally acceptable models may be the best approach. At the same time, because the effects observed in this review are generally small and were difficult to detect, it is of some importance that future programmes are evaluated rigorously and that the different stakeholders involved think carefully about how to improve programmes to create larger effects on the outcomes. To build the evidence base further, many more of the TVET interventions currently in existence in developing countries need to be rigorously evaluated, and the results reported and disseminated efficiently.

There is a clear need for additional research in this area. The methodological inconsistencies and weaknesses of the current evidence base, and specific knowledge gaps, suggest a number of future research priorities. These include: (a) evaluating all types of TVET; (b) testing the effects of different intervention components, and analysing all other relevant variables that may influence the effect; (c) measuring all key intermediate outcomes, long-term outcomes, and net outcomes; (d) improving reporting (e.g., description of interventions and outcome measures, data needed to calculate effect sizes, information needed for risk of bias judgments and study replication); and (e) evaluating the application of quasi-experimental techniques. Acting on these will also require the various stakeholders engaged in TVET research taking a critical look at the barriers affecting research production and dissemination.

1 Introduction

1.1 RATIONALE

In the 21st century, both developed and developing nations are faced with the demands of a rapidly changing, more globally competitive world. Major forces are driving change in the world of work, including advances in information and communication technology (ICT), the introduction of new manufacturing processes, increased economic integration between countries, and increased competition due to trade liberalisation. The impact of economic globalisation, however, has been uneven. Whilst some developing countries, particularly China and India, have considerably improved their standing in the global economy, others have not fared so well. Many are seeing an expansion of the informal economy, characterised by a reliance on unskilled work combined with stagnation in the formal economy. Recent development progress in education has meant that there are more skilled workers in the world than available job prospects. Simultaneously, global unemployment is on the increase, as shocks provoked by the international financial crisis continue to reduce the capacity of the global economy to add new jobs.

Youth have been particularly hard hit by the recent global economic crisis. The youth unemployment rate rose sharply between 2008 and 2009, from 11.8% to 12.7%, reversing the pre-crisis trend of declining youth unemployment rates since 2002 (International Labour Organization [ILO], 2011). In 2011, 74.8 million youth aged 15–24 were unemployed; globally, young people are nearly three times as likely as adults to be unemployed (ILO, 2012). There is significant regional variation in youth unemployment. Countries of the Middle East, Africa, South Asia, and Latin America are particularly affected (United Nations [UN], 2012; United Nations Department of Economic and Social Affairs [UNDESA], 2011). Many young people worldwide are underemployed and/or work in low quality jobs that offer limited socio-economic security, few training opportunities, and poor working conditions. The majority of the world's youth work in small-scale, often family-based jobs in the informal economy, many of which are labour-intensive and require low levels of skills.

In many countries, young women are much more likely to be un/underemployed than young men (UN, 2012). The marginalisation of women in employment and training is a relevant issue globally given the potential impact on human capital, but particularly in those countries in which women constitute the majority of the

population (Misola, 2010). Other groups of young people more prone to unemployment and underemployment include youth with disabilities, those affected by HIV/AIDS, indigenous youth, demobilised young soldiers, and young migrant workers. Many developing countries, particularly those in Sub-Saharan Africa, Southern Asia, the Middle East, and the Pacific Islands, are experiencing a 'youth bulge' (that is, have two-thirds of their populations under the age of 30). One billion young people are predicted to reach employment age within the next decade (ILO, 2012), compounding what are already severely limited opportunities for integrating youth into the labour market.

For young people and their families, prolonged absences from the labour market, underemployment, and employment in poor quality jobs contribute to high levels of poverty. Over 40% of all young people live on less than $2 a day; in developing countries, youth are disproportionately among the working poor (ILO, 2012). This enormous unlocked potential represents a substantial loss of opportunity for economic growth. Increasing numbers of youth are moving to urban areas in search of employment; simultaneously, however, many cities in the developing south lack the infrastructure and resources to support large bursts of population growth. There are also concerns that rising levels of youth un/underemployment, and the social exclusion which results from prolonged frustration in the search for status and livelihood, may be a source of social and political instability and conflict, often in already unstable countries.

The labour productivity gap between developing and developed regions, although decreasing, continues to be significant (ILO, 2012). Education and training are widely perceived to be relevant to debates about productivity and competitiveness, with increasing emphasis being given to work- and skills-based solutions to economic competition and poverty. Following a decline in interest from the mid-1990s to the mid-2000s, technical and vocational education and training (TVET) has returned to the agenda of governments and donor agencies internationally, particularly in sub-Saharan Africa and South Asia (King & Palmer, 2010). The political and policy communities in many low-and middle-income countries (LMICs) remain attracted by the assumed link between TVET and a reduction in unemployment, through its equipping of individuals with relevant skills and knowledge, thus enabling them to respond to employment opportunities (see, for example, African Union, 2007). UNESCO is amongst those highlighting the policy importance now being placed on higher-order skills and their central role in the global knowledge-based economy, particularly with regard to poverty reduction, economic growth and social stability (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2010a). This shift in priorities is evident in the 2012 Education for All (EFA) Global Monitoring Report, which strengthens the focus on TVET and skills development that might expand opportunities for marginalised groups (UNESCO, 2012). TVET has become a key area for investment in developing countries and many initiatives have been implemented to address unemployment issues and improve economic growth. Local and national

governments, private organisations and companies, national and international non-governmental organisations (such as the Asian Development Bank, the International Labour Organisation, and the World Bank), and, on a more personal level, trainees themselves, have all made varying levels of investments in TVET. For instance, the national expenditure for TVET activities in the Philippines in 2002 was estimated at $200 million, or 0.3% of GDP (Péano, Vergel de Dios, Atchoaréna, & Mendoza, 2008).

1.2 TVET INTERVENTIONS

1.2.1 Definition

There is no universally accepted definition of technical and vocational education and training (TVET). As a field, it is continually changing, usually in response to the demands made upon it (Maclean & Wilson, 2009). Broadly defined, TVET is concerned with the acquisition of knowledge and skills for the world of work. Here, we follow UNESCO's definition of TVET as '….a comprehensive term referring to those aspects of the educational process involving, in addition to general education, the study of technologies and related sciences, and the acquisition of practical skills, attitudes, understanding and knowledge relating to occupants in various sectors of economic and social life' (UNESCO, 2010b).

A great diversity of TVET models can be found worldwide. Various terms are used to describe the diverse elements of the field that are now conceived as comprising TVET, many of them specific to particular geographical areas (for example, in the United States, the current term is career and technical education). Furthermore, the organisation of TVET varies widely, both between and within countries. With no internationally accepted set of definitions of the different types that can be distinguished, the following definitions have been used for the purposes of this systematic review:

Technical education: theoretical vocational preparation of students for jobs involving applied science and modern technology; compared to vocation education (which focuses on the actual attainment of proficiency in manual skills), technical education emphasises the understanding of basic principles of science and mathematics and their practical applications; usually delivered at upper-secondary and lower-tertiary levels to prepare students for occupations that are classified above the skilled crafts but below the scientific or engineering professions (although diploma- and degree-level courses also exist).

Vocational education: organised activities designed to bring about learning as preparation for jobs in designated (manual or practical) trades or occupations; traditionally non-theoretical and focused on the actual attainment of proficiency in manual skills; usually considered part of the formal education system and thereby falling under the responsibility of the Ministry/Department of Education.

Vocational training: prepares learners for jobs that are related to a specific trade or occupation; but, compared to vocational education, is better linked to the labour market and employment development system, and therefore usually falls under the responsibility of the Ministry/ Department of Labour.

On-the-job training: workplace-based training that uses real jobs as a basis for instruction and for practical purposes.

Apprenticeship training: combines on-the-job training for a highly skilled craft or trade (from someone who is already a skilled leader in the field) with academic/ theoretical instruction; ranges from informal work-based 'learning-by-doing' to formal structured programmes sponsored by large industrial firms.

1.2.2 How the intervention might work

The following logic model provides a very simple representation of the relationships among (a) the resources that are invested; (b) the activities that take place; and (c) the benefits or changes that result, as a sequence of events.
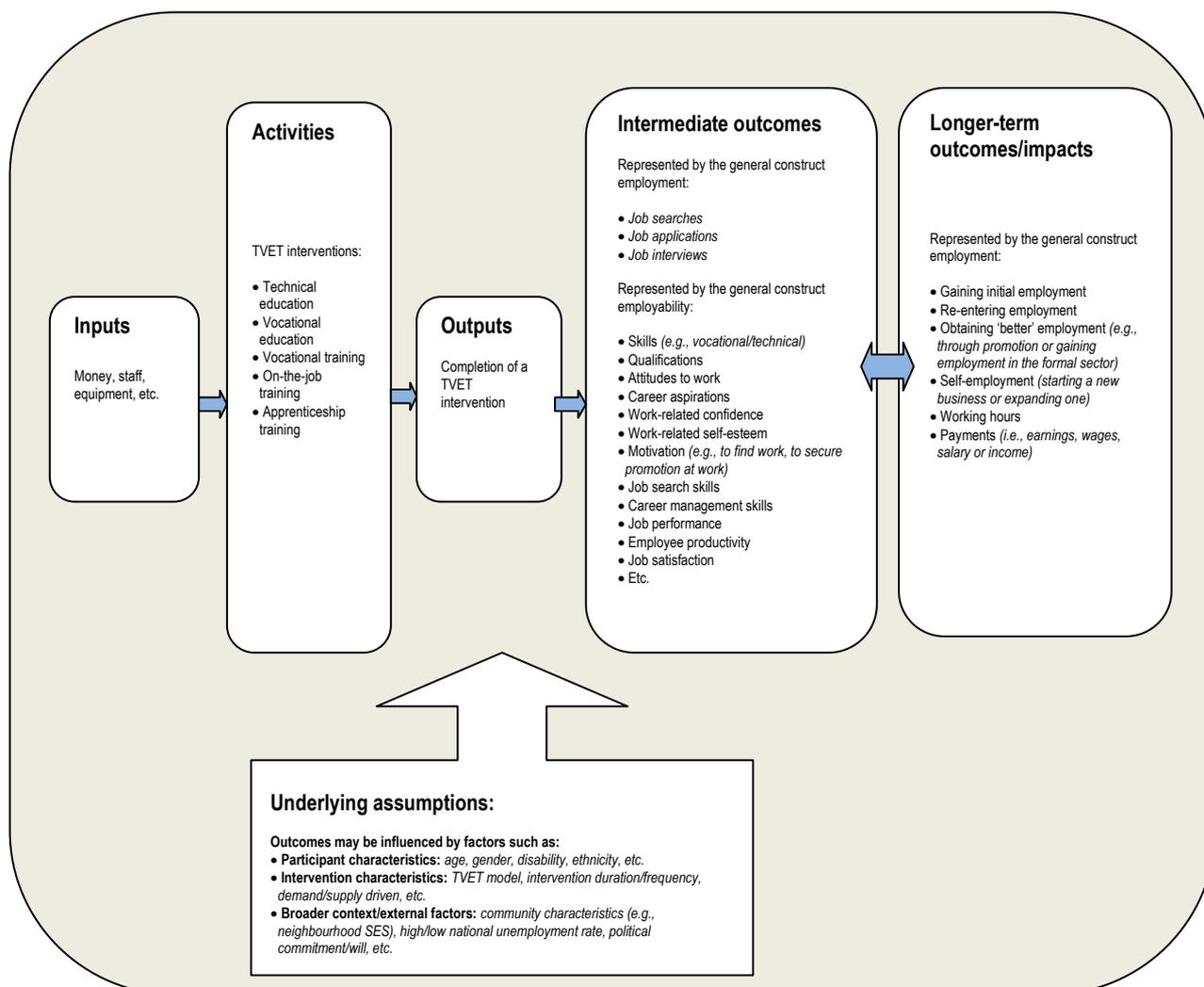
Preliminary thoughts on some of the underlying assumptions are indicated. Employability refers to a person's capability of gaining initial employment, maintaining employment (including the ability to make transitions between jobs and roles within the same organisation to meet new job requirements) and/or obtaining new employment if required (Hillage & Pollard, 1998). It is therefore a concept that can be applied to both employed people seeking alternative jobs or promotion and unemployed people seeking work. The concept of employability has become a cornerstone of labour market policies and employment strategies internationally, with many policymakers viewing the development of individual employability as a crucial step towards improving access to employment and as a means of offering workers the opportunity to develop the skills allowing self-sufficiency within the labour market (McQuaid & Lindsay, 2005). There is increasing recognition that employability is dependent not only on individual characteristics but also the environmental, social, and economic context in which work is sought (Department for Employment and Learning, Northern Ireland [DELNI], 2002).

Most interventions will have positive and negative impacts, both of which should be taken into account to assess the net difference that results from the intervention, over and above what would have taken place anyway. A key question concerning labour market interventions, including those offering TVET opportunities, is whether job creation is additional or not. It is changes in the net employment rate that are of primary interest to policy-level organisations and departments. Therefore, estimates of gross employment outcomes should, ideally, be adjusted to take into account displacement and substitution effects.[1] In instances where no adjustments have been made, it will be important to remember that, even if found to be effective, the TVET intervention may not generate any additional employment; it may only be affecting who gets employed, not the level of employment. It is recognised that determining the 'additionality' of any employment effects is methodologically very challenging.

1.3 PREVIOUS REVIEWS

The body of literature taking stock of the evaluation evidence on TVET in relation to young people is relatively limited. In many of the existing reviews, evaluations of training and retraining are presented alongside other typical active labour market programmes (ALMPs), such as employment services, public works, wage and employment subsidies, and self-employment assistance. Few reviews have focused specifically on young people and/or developing or transition countries.

Kluve and Schmidt (2002) compared the results of a sample of European impact evaluations of ALMPs implemented between 1983 and 1999 to programmes from

---

[1] Displacement effects: for example, where the setting up of new businesses has displaced less productive informal enterprises; substitution effects: for example, where a person who has received training obtains a job at the expense of other potential employees.

the United States previously studied by Heckman, LaLonde, and Smith (1999). Their analysis suggests mixed effects across different categories of intervention and target population. Young workers were found to be the most difficult group to assist among the unemployed. Kluve (2006) followed this up with a meta-analysis of European ALMPs in the later 1990s and 2000s. More recently, Card, Kluve, and Weber (2010) present the results of a meta-analysis of evaluations of ALMP impacts from 97 studies conducted between 1995 and 2007 (the vast majority set in high income countries). The sample is derived from responses to a survey of academic researchers affiliated with the Institute for the Study of Labour (IZA) and the National Bureau of Economic Research (NBER). The authors report that, when comparing across different participant groups, interventions specifically targeted at youths are less likely to yield positive impacts than untargeted programmes, although in contrast to some earlier reviews, they find no large or systematic differences by gender. Within-country, cross-programme comparisons were undertaken by Greenberg, Michalopoulos, and Robins (2003) who synthesised findings from 15 publicly-funded training programmes in the United States to measure effects on participants' earnings. Results of their meta-analysis suggest highly heterogeneous earning effects among assisted groups. The overall training effect on youth was negligible, but some control variables showed small positive effects: (i) across training types, classroom skills training courses yielded consistently better effects than on-the-job training, while (ii) gender and race controls suggested lower effectiveness of training for white and female beneficiaries than for all other participants. A global review of skills development and transition to work (Van Adams, 2007) reports positive findings from evaluations of TVET interventions for youth, although again these findings are mostly from advanced countries.

There have been a number of reviews based on information in the Youth Employment Inventory (YEI), the first comprehensive database to provide comparative information on youth employment interventions worldwide. Originally initiated by the World Bank, the YEI comprises more than 400 youth employment interventions from around 90 countries (see, for example, Betcherman, Godfrey, Puerto, Rother, & Stavreska, 2007; Betcherman, Olivas, & Dar, 2004; Fares & Puerto, 2009; Katz, 2008; Puerto, 2007; Stavreska, 2006). Betcherman et al. (2007) summarised information on a large number of international programmes supporting young people in their early years in the labour market. A substantial number of the reviewed interventions were from countries in Eastern Europe and Central Asia, Latin America, and the Caribbean (primarily middle-income countries). The review identified training as the most common form of intervention used to help young people improve their employment situation, and suggested that such programmes have a more positive impact in developing counties than in developed countries. A more recent review by Angel-Urdinola, Semlali, and Brodmann (2010) analysed the main design features of ALMPs targeted at youth in Arab-Mediterranean Countries. Interventions from nine countries were examined:

Morocco, Algeria, Tunisia, Egypt, Lebanon, Syria, Jordan, West Bank and Gaza, and Yemen.

Although there is growing consensus that TVET is important for economic growth and social cohesion, it is still not clear who should fund, provide, and regulate it, or who should take it. Collecting evidence from studies that have analysed these issues is crucial for purposes of policy making. Since most prior reviews have focused on high-income countries and/or adults of all ages, there are grounds for concentrating this review solely on the effects of TVET interventions on youth in low- and middle-income countries. There is also motivation for this systematic review from a methodological perspective. Few reviews in this area are based on a systematic search for literature, and several use a 'vote-counting' approach to synthesis. These are problems that this review aims to remedy, thereby adding value to the existing body of research on this topic.

1.4 OBJECTIVES OF THE REVIEW

This systematic review aims to answer the following questions:

- What are the effects of different models of technical and vocational education and training (TVET) interventions on the employment and employability outcomes of young people, aged 15-24 years, in low- and middle-income countries?

- What do the findings suggest about moderating effects?

To help in decisions as to whether and what kind of intervention should be undertaken, the main objective of the review is to systematically gather and synthesise the relevant evidence, and to show variation in treatment effects, magnitude of effects, and the relationship between magnitude and mode of TVET. In addition, evidence of differential effects for youth with different characteristics will be explored (e.g., in relation to gender). Possible reasons for varying or conflicting results will be discussed. A final objective of the review is to identify gaps in the literature and highlight potential avenues for future research.

2 Methods

2.1 TITLE REGISTRATION AND REVIEW PROTOCOL

The title for this systematic review was published in The Campbell Collaboration Library of Systematic Reviews on 14 November 2011. The review protocol was published on 1 September 2012. Both the title registration and protocol are available at: http://campbellcollaboration.org/lib/project/227/.

2.2 ELIGIBILITY CRITERIA

To be eligible for inclusion in the review, studies were required to meet the following criteria.

2.2.1 Interventions

Inclusion in the systematic review was restricted to TVET interventions[2] with the following characteristics:

- Technical education, vocational education, vocational training, on-the-job training, apprenticeship training (as defined in the Introduction);

- Formal and non-formal types of learning arrangements;

- All modes of delivery: e.g., online, face-to-face, distance learning, apprenticeship;

- All types of settings: e.g., schools, colleges, apprenticeship training centres, worksites, other private enterprises;

- All types of provider/regulator: public (e.g., government-funded schools and training centres); private (e.g., companies, churches, non-government organisations, private colleges) and traditional (e.g., craft guilds);

- TVET offered at secondary and post-secondary levels (including vocational diplomas and degrees);

---

[2] Where intervention is defined as: a policy, programme, or some other type of action that involves the intervention of a government, individual, group, or organisation in social affairs.

- Provision of (i) initial training for young people from the age of 15/16 years after compulsory school, but prior to entering work; (ii) continuing education and training for adults in the labour market leading to personal, flexible, and/or vocational competencies; or (iii) training for unemployed persons who are currently available for work and seeking work (including retraining for those made redundant);

- TVET delivered for any length of time or frequency.

Multi-service interventions (for example, combining on the-job training with wage subsidies) were eligible for inclusion in the review. Also eligible were studies focused on the provision of financial assistance to purchase training where trainees' participation in such training was then evaluated. Labour market programmes that did not incorporate any training, but were restricted to the provision of other services, such as job search assistance or financial subsidies, were not eligible. Interventions promoting entrepreneurship through, for example, technical and business training, were eligible, whilst those promoting self-employment by providing technical assistance only were not. Continuing professional development (CPD) interventions (i.e., those designed to upgrade the knowledge and skills of practitioners in the medical and other professions) were not eligible. Interventions targeted specifically at youth with particular special needs, such as learning disabilities, physical disabilities, emotional problems, or behavioural problems, were outside the scope of this review. However, studies in which youth with special needs participated in mainstream TVET/skills training were considered eligible.

2.2.2 Participant characteristics

The focus of this review is on young people. Countries vary considerably in their definition of young people. The standard United Nations definition of youth as those belonging to the 15-24 years age group was applied to this review (UN, 1992).

Eligible participant populations included youth with the following characteristics:

- Age: Young people aged 15 - 24 years. In addition to samples in which all participants were aged 15-24 years, samples consisting of both young people and adults older than 25 years were eligible if (i) the average age of the sample lay between 15 and 24 years; (ii) the majority of participants were aged 15-24 years; or (iii) findings were disaggregated by age and reported for 15-24 year olds.

- Geographical location: From low- or middle-income countries (as defined by the World Bank: see Appendix 8.2);

- Gender: Male and/or female (i.e., both dual- and single-sex studies were eligible for inclusion in the review);

- Employment and education:

- o   Any employment status at time of service receipt (i.e., not in paid employment or in paid full- or part-time employment);

- o   Any skills level, prior experiences, achievements or level of qualification.

2.2.3   Research designs

To fulfil the eligibility criteria, studies were required to be impact evaluations that used an experimental or quasi-experimental design. Eligible research designs included those in which the authors used a control or comparison group, and in which: participants were randomly assigned (using a process of random allocation, such as a random number generation); a quasi-random method of assignment was used and pre-treatment equivalence information was available regarding the nature of the group differences (and groups generated were essentially equivalent); participants were non-randomly assigned but matched on pre-tests and/or relevant demographic characteristics (using observables, or propensity scores) and/or according to a cut-off on an ordinal or continuous variable (regression discontinuity design); participants were non-randomly assigned, but statistical methods were used to control for differences between groups (e.g., using multiple regression analysis, including difference-in-difference, cross-sectional (single differences), or instrumental variables regression).

For this review, the control or comparison conditions in eligible studies included youth receiving no treatment, treatment as usual, or an alternative treatment. No restriction was placed on duration of follow up.

2.2.4   Outcomes

To be included, a study had to assess intervention effects on at least one eligible outcome variable. Qualifying outcome variables were those falling into the following general construct categories: (a) employment and (b) employability.

- primary outcomes represented by the general construct employment: for example, gaining initial employment; re-entering employment; obtaining 'better' employment (e.g., through promotion or gaining employment in the formal sector); self-employment (starting a new business or expanding one); working hours; and payment levels (i.e., earnings, wages, salary or income)

- intermediate outcomes represented by the general construct employment: for example, job searches, job applications, job interviews

- intermediate outcomes represented by the general construct employability: for example, vocational or technical skills/knowledge/qualifications; attitudes to work; career aspirations; confidence; self-esteem; motivation (to find employment, secure promotion, etc); job search skills; career management skills; job performance; employee productivity; job satisfaction

Studies measuring either gross employment or net employment (i.e., where displacement and substitution effects have been taken into account) were eligible for inclusion.

Qualitative studies were not eligible for inclusion in the review, including any studies using perception measures only (i.e., those examining the views of employers and/or the workforce about their employability), regardless of whether or not they quantified their findings.

2.2.5 Other study characteristics

The date of publication or reporting of the study must have been in 2000 or later. This date was chosen due to time and funding limitations. The funder approved the cut-off for that reason. Eligible studies could be published in any language provided they met all other eligibility criteria. We did not exclude specific forms of publication, such as theses and dissertations.

2.3 LITERATURE SEARCH

A comprehensive search strategy was used to search the international research literature for qualifying studies. Different types of sources were searched, including sources with a particular focus on low- and middle-income countries (many of which were sourced from the Cochrane Effective Practice and Organisation of Care (EPOC) Group's list of sources relevant to LMICs: http://epocoslo.cochrane.org/lmic-databases). A number of European-focused sources were included to assist the capture of relevant literature from 'transition economies' (countries in Central and Eastern Europe and the Former Soviet Union) and Turkey (see Appendix 9.2 for the World Bank list of low- and middle-income economies, grouped by region). The use of a wide range of sources was intended to capture both academic and 'grey' literature and reduce the omission of relevant studies, to ensure that our search was as unbiased as possible.

2.3.1 Information Sources

A wide range of general and specialist electronic bibliographic databases was searched (see Appendix 9.3).

A tailored search query was developed for each bibliographic database relying on the database's index terms (where available) and/or free-text terms. In most cases, the search strategies combined a comprehensive list of search terms related to the intervention, outcomes, and research design. Database thesauri were consulted to ensure that all appropriate synonyms were included. Synonyms and wildcards were applied as appropriate. There were no country or language restrictions to the search. A publication year filter was used. The search strategy for ERIC is presented in Appendix 9.4.

To supplement the electronic bibliographic database search, we hand searched websites/gateways, checked the bibliographies of all included studies and relevant reviews, and performed forward citation tracking (through the ISI Web of Knowledge and Google Scholar). A list of websites/gateways searched is presented in Appendix 9.5. We emailed specialists in the field, including authors of included studies, for information about any potentially relevant studies. A specific request for assistance with the location of study reports published in languages other than English was made. Authors and funding sources were also contacted regarding the availability of translated versions of included studies. Requests for relevant literature were also made through the Network for Policy Research, Review and Advice on Education and Training (NORRAG) and the UNESCO-UNEVOC e-forum. Again, a specific request for assistance with the location of studies published in languages other than English was made. We did not undertake a keyword search using Google Scholar, hand search individual journals, or search for conference proceedings or dissertations separately.

2.3.2 Study selection

Selection of primary studies was based on the pre-developed selection criteria described above. These criteria were piloted by two researchers who screened a sample of reports independently and compared their results. Discrepancies were resolved by further review of the respective titles and abstracts. This process was repeated until consistency in application of the selection criteria was achieved. The study selection process then proceeded as follows:

- The review team manually examined the titles and abstracts of records identified through the searches of electronic databases to assess eligibility. The relevance of each item was assessed by an individual reviewer (i.e., single screening) and decisions recorded in the reviewing software, EPPI-Reviewer. Items were included at this stage if they appeared to meet the criteria on the basis of the information in the title and abstract, and excluded if they were clearly ineligible. Where there was any doubt as to their eligibility, items were marked as 'unsure.' 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. In cases where only the title of the study was available, reference within the wording of the title to (a TVET intervention) AND (a relevant employment-related outcome OR a term suggesting the study was an evaluation) automatically warranted a full length review of the article.

- Following the manual screening of all items from the electronic searches, the hand searches referred to above were conducted. Here, the searching and screening processes ran concurrently. Study eligibility was assessed by an individual reviewer, who kept a manual record of all items that appeared to meet the inclusion criteria and those over which there was any doubt. Where only the

title was available, and/or the information was not in the English language, the same procedures as for items identified through the electronic searches were followed.

- The full-length reports of all studies promoted from the first level of screening that either (a) appeared to meet the inclusion criteria, or (b) were marked as 'unsure,' were obtained.

- Detailed manual examination of the full-length reports was undertaken independently by pairs of reviewers. Reviewers then compared and discussed their assessments. Any 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 differences.

2.3.3 Data collection

Reviewers used a coding tool to capture both substantive and methodological characteristics. The tool modified an existing EPPI-Centre tool in accordance with Campbell Collaboration guidelines, and also drew upon previous work by Wilson, Lipsey, Tanner-Smith, Huang, and Steinka-Fry (2010). Piloting of the coding tool was undertaken by members of the review team who worked independently on a random sample of eligible studies before meeting to compare their decisions. Reviewers were retrained on any coding items that showed discrepancies during this process and the coding manual was adapted accordingly. 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. The remaining studies were double-coded.

For eligible studies published in English, different pairs of reviewers independently extracted information from each study report and then came together to compare their decisions. Any uncertainties and discrepancies were resolved by discussion, further review of the respective study reports, and consultations with a third reviewer (JT, MN or JH), where necessary. Guidance on advanced statistical issues was provided by JH. For eligible studies published in languages other than English, attempts were made to contact authors and funding sources regarding the availability of translated versions. Where these could not be obtained, two Spanish-speaking reviewers were identified and invited to join the research team (KS-F, EW). They used the coding tool to extract the relevant information and critically appraise the four studies in question.

Reviewers entered data directly into the EPPI-Reviewer 4 database (Thomas, Brunton, & Graziosi, 2010). Information was collected relating to general study characteristics (such as year of publication); participant characteristics; the nature of the intervention and its implementation; study methods; outcome variables; and findings. The coding process also incorporated a careful consideration of the

potential methodological limitations of the included studies, focusing on the following key domains: selection bias, confounding, spillovers, outcome reporting bias, and analysis reporting bias. This involved use of a tool developed by researchers at the International Initiative for Impact Evaluation (3ie) specifically for assessing risk of bias in experimental and quasi-experimental designs based on statistical methods. The coding tool was detailed in the review protocol.[3]

The approach taken to formulating summary assessments of risk of bias and study 'quality' is presented in Table 2.1 (adapted from the suggested framework in Higgins et al., 2011).

TABLE 2.1: ASSESSMENTS OF STUDY QUALITY

Study quality	Bias within a study	Interpretation
High	• low risk of bias on all key domains	Bias, if present, is unlikely to alter the results seriously
Medium	• low or unclear risk of bias on all key domains (and number of domains rated low > number of domains rated unclear)	A risk of bias that raised some doubts about the results
Low/medium	• low or unclear risk of bias on all key domains (and number of domains rated low < number of domains rated unclear)	Bias may alter the results seriously
Low	• high risk of bias on one or more key domains	Bias is very likely to alter the results seriously

To help identify implications of the review findings for policy and practice, we used an interpretation framework that considered the strength of evidence for each set of results. This interpretation framework has evolved over several reviews by staff at the EPPI-Centre, spanning different disciplines.

2.4 ANALYTICAL METHODS

2.4.1 Effect size indices

Where data allowed, effect sizes were computed for each study. Standardised mean differences (SMDs) were the metric used in the meta-analyses to synthesise the effects of TVET interventions on both continuous outcomes (e.g., earnings) and dichotomous outcomes (e.g., employment). Although other effect sizes such as risk ratios or odds ratios might be methodologically more adequate to synthesise

---

[3] See also
http://www.3ieimpact.org/media/filer/2012/12/26/jorge_hombrados_and_hugh_waddington_conference-session12-b_3ie_dhaka_colloquium.pdf

dichotomous outcomes (Higgins & Green, 2011), most of the studies included in the review using dichotomous outcomes do not report sufficient information to use risk ratios and odds ratio. Thus, the use of risk ratios or odds ratios may lead to a substantial loss of information for the analysis. Therefore, we followed Petrosino, Morgan, Fronius, Tanner-Smith, and Boruch (2012) in the use of SMD for synthesising both continuous and dichotomous outcomes.

Each meta-analysis included a single type of outcome measure. In the primary studies, the causal effects of the programmes were expressed as a:

- Mean difference: i.e., a difference in the mean outcomes from the treatment and control groups (for example, in the mean level of earnings, weekly hours, etc.). For earnings, this value was reported in local currency or USD($).

- Probability: this is also a difference, this time in the employment rates of participants and non-participants, or more precisely differences between the two probabilities of working. For example, a value of 0.065 indicated that those in the treatment group had ceteris paribus 6.5 percentage points more probability of working than control group individuals. Often, this value was also in the form of a mean (where, for instance, the probability of being employed for trainees in each participating institution, and their control counterparts, was calculated and then effects averaged over the sample).

In the majority of cases (see below for exceptions), SMDs were calculated using the following formulae.[4] The numerator represents the causal raw impact of the programme on the outcome. In matching-based studies, this was the average treatment effect on the treated (ATT): the difference in outcomes between groups after matching (i.e., Y treatment − Y control). In a regression analysis, this is the coefficient of interest (β).

For matching-based studies:

$$SMD = \frac{ATT}{S_p}$$

To calculate the pooled standard deviation (the standard deviation of the outcome variable for both treated and control individuals) we used the Hedges' approach described in Lipsey and Wilson (2001):

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

For regression-based studies we first used the formula described in Keef and

---

[4] Where Y treatment, Y control, n treatment and n control are the outcome levels in the treatment and control groups and the sample sizes of the treatment and control groups; ATT is the average treatment effect on the treated; β is the coefficient of interest (i.e., yielding the impact of the intervention); and t is the t statistic of the regression coefficient or of the treatment effect.

Roberts (2004):

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

The denominator σ is the standard deviation of the error term in a regression. Where σ was not reported, we used the following formula as an equivalent[5].

$$S_{pooled} = \sqrt{\frac{S_{total} * (n_{total} - 1) - \frac{(\beta^2 * (n_{control} * n_{treatment}))}{n_{control} + n_{treatment}}}{n_{total} - 1}}$$

Standard errors were calculated using the following formulae[6], where t is the t test associated with the treatment effect of a regression:

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

In a small number of cases[7], SMDs were computed using the following input data:

- means, standard errors, and sample size information

- t-test statistic, sample size information

- proportion with event/without event (in each group)

The following effect-size calculator was used:
http://gunston.gmu.edu/cebcp/EffectSizeCalculator/index.html

The review also corrected for sample bias in the effect sizes due to small sample sizes by using the correction for sample bias procedure developed by Hedges and Olkin (1985). All effect sizes were coded such that positive effect sizes represented positive outcomes (e.g., less unemployment, higher wages).

Basically corrected SMD and corrected SE were estimated as follows:

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

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

---

[5] This formula was derived from the one available in Lipsey and Wilson (2001).

[6] The formula is from a presentation by David Wilson (2011), available at:
http://www.campbellcollaboration.org/artman2/uploads/1/2_D_Wilson__Calculating_ES.pdf

[7] In the case of matching-based studies, where the study reported the post-intervention treatment and control group outcomes for the matched samples of participants and non-participants, then these data were selected as the input data for computing the effect sizes. In the case of regression-based studies, where the unadjusted outcome data were available, then this was used to compute the effect sizes (since the effect size calculation methods are themselves an approximation, on balance this was felt to produce a more reliable estimate).

2.4.2 Synthesis methods

A meta-analysis was conducted on a set of studies answering a particular question where there was a minimum of two studies rated medium quality (see Table 2.1). The data synthesis was carried out using random effects statistical models. To account for differences in sample sizes for individual studies, effect sizes were averaged across studies by using inverse variance weighting of the individual effect sizes. This weighting resulted in the individual effect sizes from larger *n* studies being given more weight in the combined effect size. To visibly examine between-study variability in the effect size estimates, forest plots are used to display the estimated effect sizes from each study along with their 95% confidence intervals. Heterogeneity tests were used to examine whether variation in effect size estimates was larger than expected from sampling error alone (Deeks, Altman, & Bradburn, 2001). Heterogeneity was explored using both the Q test and the I^2 index. The Q test reveals the presence versus absence of heterogeneity as indicated by a p value of <0.05, while the I^2 index quantifies the degree of heterogeneity (Higgins & Green, 2011). If significant heterogeneity was found, possible reasons for the differences between studies was explored through analysis of sub-groups of studies (see below).

2.4.3 Missing data

If we had studies that were missing essential data, our approach involved thorough attempts to contact the original investigators and funding sources, and discussion of the potential impact of missing data on the findings of the review (Higgins & Green, 2011).

2.4.4 Moderator analyses

Moderator analyses were performed to examine potential variability in effects due to study, participant, and intervention characteristics. An analogue to the ANOVA analysis (univariate) approach was used, as described in Lipsey and Wilson (2001). It was not possible to conduct multivariate moderator analysis using meta-regression models, as we did not have the minimum of 10 studies for each individual moderator variable (Borenstein, Hedges, Higgins, & Rothstein, 2009). Moderator analyses were used both to explore heterogeneity and also to answer questions about the effectiveness according to specific characteristics of interventions and population groups. The categorical variables that identified the sub-groups used in this exercise were specified in advance of the meta-analysis.

2.4.5 Publication bias analyses

Due to an inadequate number of studies, we did not attempt to detect or exclude the existence of publication bias using statistical methods, such as funnel plots or 'trim and fill' analyses (Duval & Tweedie, 2000; Egger, Davey Smith, Schneider, & Minder, 1997; Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). This issue is discussed further in Chapter 5.

2.5 SELECTING DATA FOR ANALYSIS

Many of the studies included in the review used several estimation methods within the same study, principally matching and regression (covariance) adjustment. When there are large differences in the covariate distributions between the groups, standard model-based adjustments are known to rely heavily on extrapolation and assumptions. In response, matching has become a widely used non-experimental method of evaluation over the past three decades (D'Agostino, 1998; Rosenbaum & Rubin, 1983). Matching is done with the aim of creating treated and control groups with similar observed covariate distributions, thereby increasing robustness in observational studies by reducing reliance on modelling assumptions. Since the work of LaLonde (1986), many have investigated whether non-experimental methods can yield results similar to those from randomised experiments. The work of Dehejia and Wahba (1999), in particular, generated great interest regarding the ability of (propensity score) matching methods to potentially produce unbiased estimates of a programme's impact. A number of authors have specifically evaluated matching methods (Glazerman, Levy, & Myers, 2003; Heckman, Ichimura, & Todd, 1997; Heckman, Ichimura, Smith, & Todd, 1998; Heckman, Ichimura, & Todd, 1998; Michalopoulos, Bloom, & Hill, 2004), with many supporting the use of methods as a means of limiting reliance on inherently untestable modelling assumptions and the consequential sensitivity to those assumptions (for a discussion, see Stuart & Rubin, 2007). Others who have compared estimates from propensity score matching with different regression (covariance) adjustment analyses have found that no method is consistently better than the others (e.g., Michalopoulos et al., 2004). This presents major challenges for reviewers faced with assessing the potential of a wide range of matching and covariance adjustment methods for reducing bias in observational studies.

Drawn from some of the available practical guidance on this topic (e.g., Stuart & Rubin, 2007), the following outlines our approach for choosing between different methodologies when extracting outcome data.

Combining methods (i.e., matching and regression-based model adjustment) was judged to be more efficient in reducing bias in the estimate of the treatment effect than using those methods individually (Cochran & Rubin, 1973; Glazerman et al., 2003; Ho, Imai, King, & Stuart, 2007; Rubin, 1973a, 1973b, 1979; Rubin & Thomas, 2000). Combination could take the form of either:

- a two-step procedure in which matching is followed by regression analysis (linear regression, logistic regression, hierarchical modelling, and so on) to remove any remaining differences between groups. (Here, results should be less sensitive to the modelling assumptions and thus should be fairly insensitive to the model specification, as compared with the same analysis on the original unmatched samples.)

- a model incorporating a polynomial of the propensity scores (i.e., regression adjustment on matched sample)

Where matching and covariate adjustment were both used in a single study and then the findings from each method of estimation compared (i.e., the methods were not used in combination) matching was usually judged to be the more efficient estimator (especially in cases where the difference-in-differences version had been implemented). However, there was potential for model-based adjustment methods to be considered more efficient if:

- there was substantial bias between the groups in the matched samples (e.g., imbalance in the propensity score of more than 0.5 standard deviations), and the model-based approach used high-quality data with a rich set of covariates (Glazerman et al., 2003);

- matching was undertaken using a small set of covariates and the model-based approach involved the use of a rich set of covariates; or

- the matching procedure resulted in very small sample sizes (furthermore, much better balance is achieved when there are many controls available for the matching) (Rubin, 1976), and the model-based approach involved the use of a rich set of covariates.

In short, particularly in cases of cross-sectional versions of matching, if a model was correctly specified then it tended to be judged as more efficient than matching.[8]

In deciding which outcome data to select, making a choice between different matching techniques was sometimes required. Matching techniques differ in both the way they define similarity and the way weights are computed.[9] Where different techniques for constructing a matched sample (using the propensity score) were used in a single study included in the review, our approach to the selection of data was as follows:[10]

- If the authors reported which technique led to the most closely related/matched samples (i.e., best balance between the covariates in the treated and control groups) the outcome data based on this technique were extracted.

- Where no such information was presented by the authors, the following hierarchy applied:

---

[8] For matching, the use of the same data source for the participants and non-participants was also regarded as important, as this would help ensure similar covariate meaning and measurement.
[9] Traditional matching estimators pair each participant with a single matched non-participant (Rosenbaum & Rubin, 1983), whereas more recently developed estimators pair participants with multiple non-participants and use weighted averaging to construct the matched outcome.
[10] To some extent, the best method depends on the individual data set and where relevant this was also taken into consideration.

- o local linear (most efficient)[11]

- o kernel[12]

- stratified

  - o nearest neighbor[13] (also called pair-wise matching) (least efficient)

- In situations where different numbers of nearest neighbours were used, the general principle was that we extracted outcome data relating to the technique using the greatest number of neighbours (unless the authors reported better balance between the covariates in the treated and control groups using a different number of neighbours, or reviewers determined this).

- For kernel regression matching conducted using more than one bandwidth (0.1, 0.2 and so on), our approach was to extract the outcome data relating to the highest bandwidth.[14]

A large body of techniques for carrying out regression analysis has been developed. In cases where the authors reported several models with different combinations of control variables in the same paper, our approach was to focus on the effect estimates that were derived from the most similar models across studies. In so doing, the aim was to minimise (although not eliminate) the differences in what was adjusted across studies.

For studies using cross-sectional and difference-in-differences estimation strategies, we extracted the outcome data for both. For studies reporting different estimation parameters (e.g., average treatment effects, marginal treatment effects, and so on) we extracted the outcome data relating to each of these.

Many of the included studies reported results separately for different cohorts and/or different sub-groups of participants. When it could be established that the different cohorts or sub-groups contained no overlapping subjects, we treated them as independent samples. Where different sized samples were used (and these samples were overlapping), the general principle was that the impact effects for the largest sample would be used. In practice, however, decisions about which to include in the meta-analysis were made on a case-by-case basis (taking into consideration relevant

---

[11] Local linear regression is a non-parametric regression technique that improves on the more traditional kernel regression estimator (Fan, 1992, 1993). It differs from kernel regression in terms of weights.

[12] In Kernel matching, all the individuals of the sample are used. In the estimation of treatment effects, more weight is assigned to those matches that are more similar.

[13] In Nearest Neighbour matching, each individual in the treatment group is matched with the most similar individual or individuals in the control group. However, this process does not guarantee that the matched individuals are sufficiently comparable in terms of propensity scores if the samples do not overlap. The Nearest Neighbour matching can be improved by the use of a caliper, although this strategy may conduct to losses of observations from the treatment group. If a sufficiently small caliper is used, Nearest Neighbour approaches are preferred to stratification approaches.

[14] The literature is not homogeneous on this point.

issues relating to the selection of the sample, such as whether there was likely to be more overlap between control and treatment individuals in terms of observable and unobservable characteristics). Where no such issues were noted by the authors or identified by the reviewers, the approach was to select the largest sample.

3 Results of searches

This chapter outlines the results of the literature search and the key characteristics of the included studies. Appendix 8.6 shows the flow of literature through the review process. Additional details about each included study are provided in Appendices 8.7, 8.8 and 8.9. The term 'publication' in this chapter refers to a report of the methods and outcomes of a research study, and the term 'study' refers to an instance/piece of research work. Where more than one study evaluated the same intervention, and/or the study data overlapped, they were treated as unique studies if they involved the use of different methods of evaluation. No two studies drawing on the same data are used in any individual meta-analysis presented in Chapter 5.

3.1 LITERATURE SEARCH AND STUDY IDENTIFICATION

Electronic searches of bibliographic databases identified a total of 8514 citations. After removal of 442 duplicates, the titles and abstracts of the remaining 8072 items were manually screened for relevance. This process resulted in the exclusion of 7925 items due to their not meeting the inclusion criteria and the retrieval of 145 full-length publications for closer examination. Two publications were unobtainable. Full text screening of the 145 publications resulted in 136 being excluded. Nine publications were judged as meeting the inclusion criteria. Hand searches as described above were then undertaken, leading to the identification of a further 46 eligible publications. On closer inspection, a number of publications was found to be linked to others, in that they described the same study (occasionally reporting on different aspects of it). A total of 25 publications were consequently coded as companion publications. At the end of this process, 30 unique studies, reported in 55 publications, had been identified for inclusion in the review. Four studies, however, could not be included (see below). In total, therefore, 26 studies, reported in 51 publications, were included in the review.

Of the 55 publications detailing relevant studies, 8 were published in Spanish, and we were unable to identify English language versions of the publications. For 4 of these 8 studies, additional resources were obtained in order to extract data, and they were included in the review. One eligible Spanish-language study could not be included in the review, as efforts to obtain a copy of the publication were

unsuccessful.[15] The 3 remaining eligible non-English language studies were not included in the review because financial resources had been exhausted.[16]

Publication dates ranged between 2001 and 2011. The majority of studies were published as technical reports, many of which were described as working papers. Most were published by corporations such as the World Bank or regional development banks (primarily, the Inter-American Development Bank). A small number were published by non-governmental organisations and independent research institutes. Occasionally, this information was not reported.

3.2   INCLUDED STUDIES

Information about the methodological and participant characteristics of the 26 included studies is presented in the remainder of this chapter, including a brief overview in Tables 3.1 and 3.2. Details about individual studies are presented in Appendix 8.9.

3.2.1   Methodological characteristics

The majority of the 26 studies were conducted independently by teams based in universities and/or other research organisations; others were carried out exclusively by researchers who were closely related to the funding body or by teams comprising both independent and related evaluators.

The 26 studies employed different methodologies for evaluating the impacts of the interventions. Three studies used a randomised experimental design, and 23 used a non-experimental design (of these, 2 were natural experiments and 21 were quasi-experiments). There was heterogeneity within the 23 non-experimental studies in terms of the selection of the comparison group. In around half of the non-experimental studies, the comparison group was constructed ex-ante, for example, from registered applicants who did not start the course or from eligible non-applicants. The remaining non-experimental studies involved ex-post selection of the comparison group, for example, from similar individuals identified in household or labour force survey data. The non-experimental evaluations used different econometric techniques to address selection bias and net out the impacts of other factors (each of which imposes different assumptions and have different strengths with respect to internal validity). These techniques are broadly classified into two main types: matching and covariate/regression adjustment. The most common matching method used was propensity score matching (Rosenbaum & Rubin, 1983); other matching methods included log-odds ratio matching and the use of non-parametric reweighting techniques. Studies using covariate adjustment methods

---

[15] The study was judged as meeting the review selection criteria on the basis of detailed descriptions in previous literature reviews.
[16] The four eligible non-English language studies that were not included in the review evaluated ProJoven, a Peruvian programme that is evaluated by a number of studies included in the review.

used different model specifications, most commonly ordinary least squares regression. The majority of studies used more than one estimation method, often as a means to verify robustness of the results.[17]

In terms of measurement of outcomes, the majority of studies utilised a cross-sectional impact estimator; that is, they compared the outcomes for treatment and comparison group persons measured at some time period after the intervention ended. By exploiting the panel structure of the data, around a third of the included studies attempted to purge time-invariant unobservables through the use of the difference-in-difference estimator. This involves subtraction of the before and after change in outcomes for comparison group members from the before and after change for treatment outcomes (where the change is measured relative to some pre-intervention benchmark time period).[18]

Most studies measured average treatment effects on the treated (ATT): the differential impact that the treatment showed for those individuals who actually participated in the intervention. A minority of studies measured average treatment effects (ATE), which can be defined as the average effect for the population, and/or conducted intention-to-treat (ITT) analyses which were based on the initial treatment intent, not on the treatment eventually administered.[19] Other parameters of interest included marginal treatment effects (MTE) and average treatment effects on the untreated (ATU).

There was variation between studies in relation to the time that had elapsed between completion of the TVET intervention and the measurement of outcomes (see Table 3.1). Nine of the 26 studies investigated the sustainability of treatment effects over time, with a maximum of three post-test measurements. To examine whether the impact of the intervention was robust to the data collection period, 7 studies collected data for several cohorts covering a number of years (with results reported separately for the different year groups and/or pooled).

In the case of 11 studies, a single cohort of a multi-cohort intervention was used for the evaluation. Two studies focused on the entire universe of trainees. Two-thirds of all studies estimated heterogeneity of treatment effects, most commonly by gender. Sixteen studies involved analysis with a sample size of more than 250 participants, and 7 with a sample size of 250 participants or less. The sample size was not specified in 3 studies. Finally, a very small number of studies dealt specifically with

---

[17] For example, many of the matching studies tested the robustness of the results to different matching algorithms.

[18] It has been argued that combining propensity score matching and difference-in-differences can greatly reduce the bias found in other non-experimental evaluations (Heckman et al., 1997; Heckman et al., 1998; Smith & Todd, 2001), although, even with the use of these techniques, bias due to unobservables cannot be ruled out. Heckman et al. (1997) suggests that failure to compare participants and controls at common values of matching variables is a greater source of bias than the problem of selection bias due to differences in unobservables.

[19] ATT is the estimator computed in most non-experimental evaluations. ATE can also be defined as the weighted average of the effect on the treated and the effect on the untreated. In experimental designs (i.e., in 'perfect' RCTs) there is no distinction between ATE and ATT.

the issue of trainees with partial instruction.

3.2.2 Participant characteristics

The majority of studies included both male and female participants (fairly evenly balanced), with a single study focused exclusively on young women. In 19 studies, the average age of the study participants lay between 16 and 24 years. One study focused on 12-22 year olds, but no average age was provided. Of the 5 studies that did not report the average age, 2 reported the age range: 1 study focused on participants aged 16-35 years, and another included individuals aged 18-65 years. Each of these 5 studies conducted sub-group analyses by age. Finally, in 1 study, the average age of participants was 36 years; it too conducted sub-group analyses by age. The majority of study samples included some participants in employment and others who were not in employment. In a single study, none of the training beneficiaries were employed at the start of the study.

Outcomes

The studies measured a number of different labour market outcomes (see Table 3.2). The main outcomes fell into three main categories: employment, hours worked, and income. Most studies measured paid employment and/or monthly earnings. Studies examined gross employment only, with none measuring changes in net employment (where displacement and substitution effects have been taken into account). Intermediate outcomes were examined in only 2 studies. The focus of the evidence synthesis is on the outcomes that are not italicised in Table 3.2.

The ways in which outcomes were measured varied. For example, for earnings and weekly hours of work, a few studies restricted their analysis to participants who were working at the time of the follow-up interview, but most studies included everyone in their calculations (i.e., authors imputed zero earnings/hours etc., for those who reported being unemployed or otherwise out of the labour force).

TABLE 3.1: CHARACTERISTICS OF TVET STUDIES (N=26)

Study Characteristics					
General characteristics					
Publication year	k	%	**Sub-group impacts**	k	
2000-2005	8	31	By gender	17	
2006-2011	18	69	By age	2	
			By region	4	
Form of publication	k	%	By education	3	
Journal article	4	16	By income	2	
Technical report	20	77	By work experience	1	
Dissertation	1	3	By age and gender	1	
Conference paper	1	3	By education and region	1	
			By gender and region	1	
Method characteristics			Other	3	
Design	k	%			
RCT	3	11	**Outcome timing**	k	
Natural experiment	2	8	Short-term (approx. 0-8 mths)	12	
Quasi-experiment	21	81	Medium-term (approx. 9-17 mths)	18	
			Long-term (approx. 18+ mths)	10	
Estimation methods	k				
Unadjusted difference in means	5		**Study quality**	k	%
Regression	13		*Summary assessments*		
Matching	20		High	0	0
Cross-sectional	22		Medium	5	19
Difference-in-differences	9		Low/medium	10	39
			Low	11	42
Participant characteristics	k	%			
Mixed-sex samples	25	97			
Female only samples	1	3			

TABLE 3.2: OUTCOMES

Intermediate outcomes	Main outcomes
- Further training (1 study) - Time spent job searching (1 study)	**Employment** - Overall paid employment (22 studies) - Formal employment (10 studies) - Self-employment (3 studies) *Total months worked (1 study)* *Months in employment since training ended (1 study)* **Hours worked** - Weekly hours worked, amongst the employed (9 studies) - Weekly hours worked, amongst the self-employed (1 study) *Hours worked per month (1 study)* *Days worked per month (1 study)* **Income** - Monthly earnings (22 studies) - Self-employment earnings/profits (3 studies) - Hourly wages / rate of pay (5 studies) *Income (2 studies)* *Formal earnings (1 study)* *Informal earnings (1 study)*

4 Intervention characteristics

The 26 studies included in the review evaluated 20 different programme interventions (hereafter interventions) providing TVET opportunities to young people in terms of specific outcomes. The number of reviewed studies is greater than the number of interventions because three interventions were evaluated by more than one team of investigators. Four studies evaluated different programme components/modalities available to trainees, sometimes in addition to an examination of the intervention as a whole. One study evaluated several different interventions. One study evaluated an intervention comprised of different projects operating internationally.[20] This chapter outlines the key characteristics of the 20 interventions, with further details provided in Appendix 8.10.

4.1 SETTING AND COVERAGE

Fourteen of the 20 interventions were located in Central/South American countries, with three situated in Asia, one in Africa, and two in Europe. Settings include ten upper-middle income countries (Argentina; Bosnia and Herzegovina; Brazil; Chile; China; Colombia; Dominican Republic; Latvia; Mexico; Panama and Peru); two lower-middle income countries (India and Bhutan); and one low-income country (Kenya). Some were intended to provide a small-scale demonstration effect, whilst others were large-scale operations (sometimes involving complete transformation of previous training systems).

Twenty studies evaluated an intervention (and/or different sub-components or modalities) in one country only (Acero, Alvarado, Bravo, Contreras, & Ruiz-Tagle, 2011; Aedo & Nuñez, 2004; Aedo & Pizarro, 2004; Alzuá & Brassiolo, 2006; Analítica Consultores, 2006; Attanasio, Kugler, & Meghir, 2011; Benus, Rude, & Patrabansh, 2001; Bidani, Goh, & O'Leary, 2002; Card, Ibarraran, Regalia, Rosas-Shady, & Soares, 2011; Chong & Galdo, 2006; Chong, Galdo, & Saavedra, 2008; Chun & Watanabe, 2011; Delajara, Freije, & Soloaga, 2006; Díaz & Jaramillo, 2006; Dmitrijeva, 2009; Elías, Ruiz Núñez, Cossa, & Bravo, 2004; Espinoza, 2010; Hicks, Kremer, Mbiti, & Miguel, 2011; Ibarraran & Rosas-Shady, 2006; Jaramillo, Galdo, & Montalva, 2007; López-Acevedo, 2003; Mensch, Grant, Sebastian, Hewett, &

---

[20] This is treated as a single intervention in this review.

Huntington, 2004; Ñopo, Robles, & Saavedra, 2007; van Gameren, 2010).[21] One study evaluated several different interventions within a single country (Medina & Nuñez, 2005). One study evaluated an intervention in two different countries (Alzúa, Nahirñak, & Alvarez de Toledo, 2007).

Argentina:

1. Entra 21: an international programme (2001-ongoing) operating in 18 countries across South America and the Caribbean. The two Argentine projects evaluated are (i) Fundación SES (Sustentabilidad- Educación -Solidaridad), which provides training in five regions of the country; and (ii) Agencia para el Desarollo Económico de la Ciudad de Córdoba (ADEC), which trains students from the area of Rio Segundo. This intervention was evaluated by Alzuá et al. (2007).

2. Proyecto Joven: a national programme which operated between 1993 and 2001 (distribution of the training activities was determined in accordance with regional populations). One of a series of Latin American training programmes sponsored during this period by the Inter-American Development Bank. This intervention was evaluated by Aedo and Nuñez (2004); Alzuá and Brassiolo (2006); and Elías et al. (2004).

Bosnia and Herzegovina:

3. Emergency Demobilization and Reintegration Project: a national programme implemented over a four year period (1996-1999). This intervention was evaluated by Benus et al. (2001).

Bhutan:

4. Rural Skills Development Project: operating between 2007 and 2010, this was a regional programme covering all sub-districts across the three rural districts of Haa, Trashigan, and Bumthang. This intervention was evaluated by Chun and Watanabe (2011).

Brazil:

1. (see above) Entra 21: the two Brazilian projects evaluated are: (i) Centro de Ensino Profissionalizante Rotary (CEPRO) based in São Paulo; and (ii) Instituto de Hospitalidade (IH), which operates in the northeast of the country. This intervention was evaluated by Alzuá et al. (2007).

---

[21] Whilst some of these interventions may have entailed a number of different sub-components, they were not evaluated separately. Occasionally, authors provided results for more than one intervention site.

Chile:

5. <u>Chile Joven</u> (Programa de Capacitación Laboral de Jóvenes): one of the earliest in a series of Latin American training programmes sponsored during this period by the Inter-American Development Bank, Chile Joven was available between 1991-2002 (phase I: 1991-1995; phase II: 1996-2002). This intervention was evaluated by Aedo and Pizarro (2004).

6. <u>Jóvenes al Bicentenario</u>: the first cohort of this programme was in 2008, and since then at least 10 different regions have implemented the programme. This intervention was evaluated by Acero et al. (2011).

China:

7. <u>Retraining programmes</u> for laid-off workers piloted in 30 municipalities in 1994 and expanded to 200 cities by 1996. Training in the cities of Shenyang and Wuhan is evaluated. This intervention was evaluated by Bidani et al. (2002).

Colombia:

8. <u>Jóvenes en Acción (Youth in Action)</u>: operating between 2002 and 2005, this was a national programme offered in seven of the largest cities of the country: Barranquilla, Bogota, Bucaramanga, Cali, Cartagena, Manizales, and Medellin. This intervention was evaluated by Attanasio et al. (2011).

9. <u>SENA</u> (Servicio Nacional de Aprendizaje) job training programme is an ongoing national initiative. SENA is a government agency. The programme started over 50 years ago, in 1957. There are 20 regional offices in the main cities. This intervention was evaluated by Medina and Nuñez (2005).

10. <u>Public sector vocational training</u> that is not provided by SENA: national. This intervention was evaluated by Medina and Nuñez (2005).

11. <u>Private sector vocational training</u>: national (although concentrated in the main cities). This intervention was evaluated by Medina and Nuñez (2005).

Dominican Republic:

12. <u>Juventud y Empleo:</u> one of a series of Latin American training programmes sponsored during this period by the Inter-American Development Bank, this programme operated nationally between 1999 and 2007. (NB: A programme by the same name continues in a slightly different format, with evaluation results due later this year). This intervention was evaluated by Card et al. (2011).

India:

13. A <u>livelihoods intervention</u> implemented in 2001 as a year-long, city-wide pilot project (operating in Allahabad). This intervention was evaluated by Mensch et al. (2004).

Kenya:

14. <u>Technical and Vocational Vouchers Program</u>: a regional initiative operating since 2008 in the area of Busia in Western Kenya (still ongoing). This intervention was evaluated by Hicks et al. (2011).

Latvia:

15. <u>Occupational training (OT) programmes</u> for the unemployed implemented nationwide by the State Employment Agency of Latvia; operating since the early 1990s. This intervention was evaluated by Dmitrijeva (2009).

Mexico:

16. <u>Bécate (and the programme sub-component Capacitación en la Práctica Laboral)</u>: this ongoing initiative started in 2004, as a replacement for SINAT (see no.18 below). This intervention/sub-component was evaluated by Analítica Consultores (2006) and van Gameren (2010).

17. <u>CONALEP</u> (College of Professional Technical Education): an ongoing national programme in operation since 1978. All 31 states in Mexico have CONALEP schools, although the distribution of students by state remains uneven, with large numbers attending schools within the metropolitan zone of Mexico City. This intervention was evaluated by López-Acevedo (2003).

18. <u>PROBECAT-SINAT</u> (Programa de Becas de Capacitación para Trabajadores Desempleados_ Sistema de Capacitación para el Trabajo): as a national programme initiated in 1984, PROBECAT changed its name to SINAT in 2001 (and was replaced by Bécate in 2004—see above); supported by the Inter-American Development Bank since 1996. This intervention was evaluated by Delajara et al. (2006).

Panama:

19. PROCAJOVEN: an independent sub-programme of the Assistance Program for the Building of a Training and Employment System in Panama (one of a series of Latin American training programmes sponsored during this period by the Inter-American Development Bank). Approved in 2002, PROCAJOVEN operated nationally until 2009. This intervention was evaluated by Ibarraran and Rosas-Shady (2006).

Peru:

20. <u>ProJoven</u> (Programma de Capacitacion Laboral Juvenil): an ongoing large-scale intervention first implemented in 1996, and one of a series of Latin American training programmes sponsored during this period by the Inter-American Development Bank. Originally, the programme was intended to be implemented nationwide. In practice, reduced funding resulted in the programme starting in Lima (the capital) and progressively expanding to more cities (13 in total as of 2010). This intervention was evaluated by Chong and Galdo (2006); Chong et al. (2008); Díaz and Jaramillo (2006); Espinoza (2010); Jaramillo et al. (2007); and Ñopo et al. (2007).

4.2 TVET MODELS

This section outlines some of the key characteristics that differentiate between the different TVET interventions considered in the review.

4.2.1 Type of TVET intervention

The interventions involved different forms of TVET (see Table 4.1). The most common form of intervention was a two-phase TVET intervention that combined both theoretical and practical training (usually in the format of classroom-based vocational training followed by a period of on-the-job training to provide beneficiaries with work experience). Typically, these interventions were aimed at short-term semi-skill training in specific occupations demanded in the private sector, and provided basic job readiness skills and some trade-specific skills. Nine interventions were of this type. Two interventions consisted of different TVET-related sub-components, and young people seeking training could choose between the available options. The majority of the remaining interventions offered a single form of TVET.

4.2.2 Implementation

Various agencies were involved in the design and planning of the reviewed interventions, most commonly government agencies (Ministries of Labour, employment offices, etc.). Funding for the interventions came from a number of sources; many involved public/private partnerships between national and/or local government agencies, international development agencies (such as USAID) and multilateral organisations (most commonly, the Inter-American Development Bank). Social partners, in the form of employers, contributed to the financing of one intervention.

TABLE 4.1: CHARACTERISTICS OF TVET PROGRAMMES (N=20)

TVET programme characteristics		
Type of TVET (either programme as a whole or sub-programme)	*n*	
Technical education	2	
Vocational education	1	
Vocational training	16	
On-the-job training	14	
Apprenticeship training	0	
Specific programme features	*n*	
Consists of sequential training modalities	9	
Offers beneficiaries a choice of different training modalities	2	
Incorporates other types of training (e.g., life skills)	3	
Incorporates labour intermediation services	6	
Incorporates training specifically for self-employment	4	
Vocational training, as part of a health-focused multi-component livelihoods intervention	1	
Education vouchers programme[22]	1	
Target group	*n*	
Youth	12	
Disadvantaged/at-risk	17	
Urban or rural	4	
At least secondary education	2	
Females	2	
Other	2	
Geographical location	*n*	*%*
Central/South America	14	70
Asia	3	15
Africa	1	5

[22] The Technical and Vocational Vouchers Program (TVVP) based in Kenya provided vocational education tuition vouchers, which in turn facilitated access to educational providers. This review focuses on the impact of the training that was received by recipients of the vouchers.

TVET programme characteristics		
Europe	2	10

Participation of the private sector in the provision/delivery of training was a feature in around two-thirds of the reviewed interventions. Most of these adopted a 'demand driven' approach whereby the content of the courses they offered were customised to meet the needs of the local labour market: the assumption being that there would be job vacancies for the trainees when they graduate. In these cases, the government selected the training providers and courses competitively, through a bidding process where usually both private and public firms and/or training institutions could participate. For example, one intervention was executed by private sector NGOs from across 18 countries, with each NGO entirely responsible for eligibility criteria, obtaining the internships, defining the course contents, and so on. Of the remaining interventions, a small number adopted a centre-based model whereby the government was responsible for not only the financing and regulation of training, but also its content and provision (through a national training institution).[23] A notable exception was an intervention that employed peer-to-peer instructors.

4.2.3   Target group

The target groups for the reviewed interventions fell into two main categories. The majority were specifically targeted at disadvantaged young people, based on criteria such as household income, education level, and employment experience. The remaining interventions tended to be occupation- rather than age-focused (targeting, for example, demobilised soldiers or unemployed former employees of state-owned enterprises). A small number of interventions were directed exclusively at either urban or rural residents.

4.2.4   Intervention aims

The aims of the reviewed interventions were broadly uniform, although on the whole they were not clearly or consistently reported. Overall, the main emphasis was at the individual-level. Most interventions sought to increase levels of employment/ employability for beneficiaries, with a small number having the joint aim of increasing the value of the wages received. The broader policy goal of tackling rising unemployment attributed to the global recession of the early- to mid-1990s underpinned many interventions. In a few cases, tackling a specific issue lay at the root of the intervention. For example, the stated aim of one programme was to mitigate the degree of poverty amongst rural residents who relied heavily on agriculture; another programme addressed the problem of displaced former employees of nationalised industries.

4.2.5   Intervention duration, frequency and format

---

[23] Training concentrated in ICT, handicrafts

The majority of the reviewed interventions lasted for periods of less than six months. The format for the two-phase interventions was most commonly three months classroom-based training followed by a further three months of on-the-job internships (with the shortest about three months overall, and the longest taking place over eight months). Interventions consisting solely of theoretical instruction delivered at training institutions ranged from one to six months. Those comprised solely of a period of practical on-the-job internship ranged from one to three months duration. For one intervention, the majority of trainees chose courses that lasted two years or more. Trainees on eight of the interventions were required to attend daily (Monday to Friday), and for one intervention the beneficiaries attended once a week. For the majority of the remaining interventions it was implicit that trainees attended daily. All interventions were delivered face-to-face.

4.3 OTHER FEATURES

All the interventions appeared to be voluntary. In only one instance was it reported that trainees were charged part of the training costs (e.g., they were required to purchase books and practice materials). Several interventions provided financial and other benefits to trainees, including a stipend to cover transportation costs, meals, childcare, and medical or accident insurance. No intervention appears to have offered financial support for undertaking job search activities. Trainees participating in three of the interventions involving on-the-job training were paid a wage. None of the interventions appeared to be linked to a national or international qualifications framework. One intervention incorporated a competency-based model for careers developed as part of the Education Modernisation Project financed by the World Bank. In general, study authors provided very little information on course content, curriculum, or exit qualifications. Finally, none of the interventions appeared to have incorporated a gender strategy.

5 Synthesis of results

5.1 INTRODUCTION

The synthesis examines the impact of technical and vocational education and training (TVET) delivered to young people in low- and middle-income countries. It sought to address two main review questions:

1. What are the effects of different models of technical and vocational education and training (TVET) interventions on the employment and employability outcomes of young people, aged 15-24 years, in low- and middle-income countries?

2. What do the findings suggest about moderating effects?

To address Review Question 1, we attempted to answer a number of sub-questions:

Employment
- Does participation in TVET have an effect on young people's chance of obtaining paid employment?

- Does participation in TVET have an effect on young people's chance of obtaining employment in the formal sector?

- Does participation in TVET have an effect on young people's chance of obtaining self-employment?

Income
- Does participation in TVET have an effect on young people's earnings?

- Does participation in TVET have an effect on young people's self-employment earnings?

- Does participation in TVET have an effect on young people's hourly wages?

Hours worked
- Does participation in TVET have an effect on the number of weekly hours worked by young people?

The synthesis is structured according to these three outcome categories, with findings reported in the following order: employment status (Section 5.2), income (Section 5.3), hours worked (Section 5.4), and other outcomes (Section 5.5).

Review Question 2 is concerned with finding out whether any observed relationship between TVET and these outcomes varies according to participant, intervention, or study characteristics. Using the analogue to the ANOVA approach, the following categorical variables were tested for moderating effects: study quality (medium/low quality), type of TVET intervention (two-phase models/other models), length of follow-up (short-/medium-term), and participant characteristics (female/male). The available data did not allow the use of this approach to assess the role of other potential effect size moderators, as outlined in the protocol. The moderator analysis results are presented in the text with additional statistics in Appendices 8.13 - 8.17. On the whole, given differences between studies in how many of the sub-groups were constructed, the small sample sizes in some of these analyses, and other sources of bias, we need to be wary of drawing strong inferences from the findings of these analyses.

In this review, authors of included studies often did not provide all the necessary data for calculating effect sizes. It was possible to calculate at least one effect size for only 10 of the 26 included studies. These 10 studies, and the interventions they evaluated, are detailed in Table 5.1. The intervention identification numbers correspond to the descriptions of the interventions in Chapter 4, and will be used throughout this Chapter. The findings from these 10 studies have been statistically combined using meta-analytic techniques to answer a range of sub-questions. The studies by Ibarraran and Rosas-Shady (2006) and Medina and Nuñez (2005) will occasionally appear in the same meta-analysis (as they evaluated more than one intervention). The studies by Aedo and Nuñez (2004) and Elías et al. (2004) both evaluated the Proyecto Joven programme using data from the fifth wave; however, the effect sizes from these studies are used in different meta-analyses.

The findings from the 16 studies not included in the meta-analyses will be discussed when they shed additional (or, in some cases, the only) light on a particular sub-question. Whilst meta-analysis is a more valid analysis strategy than narrative review, we decided to retain the 16 studies because their inclusion helps provide a clearer picture of the gaps in the knowledge base. We will not draw any conclusions regarding effective interventions based on these 16 studies.

TABLE 5.1: STUDIES/INTERVENTIONS INCLUDED IN THE META-ANALYSES

Study	Intervention
Acero et al. (2011)	6. Jóvenes al Bicentenario (Chile)
Aedo & Nuñez (2004)	2. Proyecto Joven (Argentina)
Aedo & Pizarro, (2004)	5. Chile Joven (Chile)
Attanasio et al. (2011)	8. Jóvenes en Acción (Colombia)
Card et al. (2011)	12. Juventud y Empleo (Dominican Republic)
Elías et al. (2004)	2. Proyecto Joven (Argentina)
Espinoza (2010)	20. ProJoven (Peru)
Hicks et al. (2011)	14. TVVP (Kenya)
Ibarraran & Rosas-Shady (2006)	19. PROCAJOVEN-Insertion Modality (IM) (Panama)
Ibarraran & Rosas-Shady (2006)	19. PROCAJOVEN-Transition Modality (TM) (Panama)
Medina & Nuñez (2005)	9. SENA (Colombia)
Medina & Nuñez (2005)	11. Private sector training (Colombia)

The interventions evaluated in the 10 studies included in the meta-analyses comprised three distinct types (see Table 5.2).

TABLE 5.2: INTERVENTION DETAILS

Intervention type	Intervention name (study authors)
Two-phase intervention involving classroom-based theoretical training followed by an internship providing practical on-the-job training	6. Jóvenes al Bicentenario (Acero et al., 2011) 2. Proyecto Joven (Aedo & Nuñez, 2004; Elías et al., 2004) 5. Chile Joven (Aedo & Pizarro, 2004) 8. Jóvenes en Acción (Attanasio et al., 2011) 12. Juventud y Empleo (Card et al., 2011) 20. ProJoven (Espinoza, 2010) 19. PROCAJOVEN (IM) (Ibarraran & Rosas-Shady, 2006)
On-the-job practical training only	19. PROCAJOVEN (TM) (Ibarraran & Rosas-Shady, 2006)
Technical and vocational education (mainly classroom-based theoretical instruction)	14. Technical and Vocational Vouchers Program (Hicks et al., 2011)
Vocational training	9. SENA (Medina & Nuñez, 2005) 11. Private sector training (Medina & Nuñez, 2005)

All analyses were inverse variance weighted using random effects statistical models. The results of each meta-analysis are presented graphically in a forest plot. Each forest plot shows (i) the standardised mean difference for each individual study (represented by the dots) and the confidence intervals for that effect size (the bars on each side of the dot), and (ii) the overall weighted mean effect size (diamond) and its confidence interval (the points of the diamond represent the width of the confidence interval), which was obtained by combining the individual effect sizes from each study.[24] A standardised mean difference greater than zero indicates that, on average, the group who received the TVET intervention had a better outcome than the group who did not (a positive effect). A standardised mean difference less than zero indicates that on average the group who received the TVET intervention had a worse outcome than the group who did not (a negative effect). Confidence intervals show the precision of the estimates of the effect size, by indicating the range within which the true mean is likely to be, given the observed data. For example, a 95% confidence interval of g=0.08 to g=0.51 around a mean effect size indicates a 95% probability that true mean effect size is somewhere between these two values. If the confidence interval does not cross zero (the 'line of no effect') the calculated difference between the intervention and control groups can be considered as statistically significant, suggesting that the impact of the intervention is, on average, either positive or negative (depending on the direction of effect). However, interpreting the findings from the meta-analyses was challenging, due largely to the small number of studies involved.

5.2 EMPLOYMENT STATUS

The majority of studies examined the impact of a TVET intervention on overall paid employment, but smaller numbers also considered their impact on formal employment, and/or self-employment (see Table 5.3).

5.2.1 Does participation in TVET have an effect on young people's chance of obtaining paid employment?

The reported data allowed the calculation of effect sizes for seven studies, and these were combined. The studies of interventions nos. 8, 12, 20, and 14 (Attanasio et al., 2011; Card et al., 2011; Espinoza, 2010; Hicks et al., 2011) were rated medium quality. The studies of interventions nos. 6, 5, and 19 (Acero et al., 2011; Aedo & Pizarro, 2004; Ibarraran & Rosas-Shady, 2006) were rated low. This analysis uses the data from each study that were closest in time to a 12-month post-training follow-up. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.1.

---

[24] Appendix 8.11 details the mean effect sizes for each study individually.

TABLE 5.3: EMPLOYMENT STATUS OUTCOMES

Outcome (n)	Studies	Analysis
Overall paid employment (22 studies)	Acero et al. (2011); Aedo & Nuñez (2004); Aedo & Pizarro (2004); Alzuá & Brassiolo (2006); Alzuá et al. (2007); Analítica Consultores (2006); Attanasio et al. (2011); Benus et al. (2001); Bidani et al. (2002); Card et al. (2011); Chong et al. (2008); Delajara et al. (2006); Díaz & Jaramillo (2006); Dmitrijeva (2009); Elías et al. (2004); Espinoza (2010); Hicks et al. (2011); Ibarraran & Rosas-Shady (2006); Jaramillo et al. (2007); López-Acevedo (2003); Ñopo et al. (2007); van Gameren (2010)	Section 5.2.1
Formal employment (10 studies)	Aedo & Pizarro (2004); Alzuá & Brassioli (2006); Alzuá et al. (2007); Attanasio et al. (2011); Card et al. (2011); Chong et al. (2008); Díaz & Jaramillo (2006); Espinoza (2010); Ibarraran & Rosas-Shady (2006); van Gameren (2010)	Section 5.2.2
Self-employment (3 studies)	Delajara et al. (2006); Hicks et al. (2011); López-Acevedo (2003)	Section 5.2.3

FIGURE 5.1: FOREST PLOT OF MEAN EFFECTS ON OVERALL PAID EMPLOYMENT
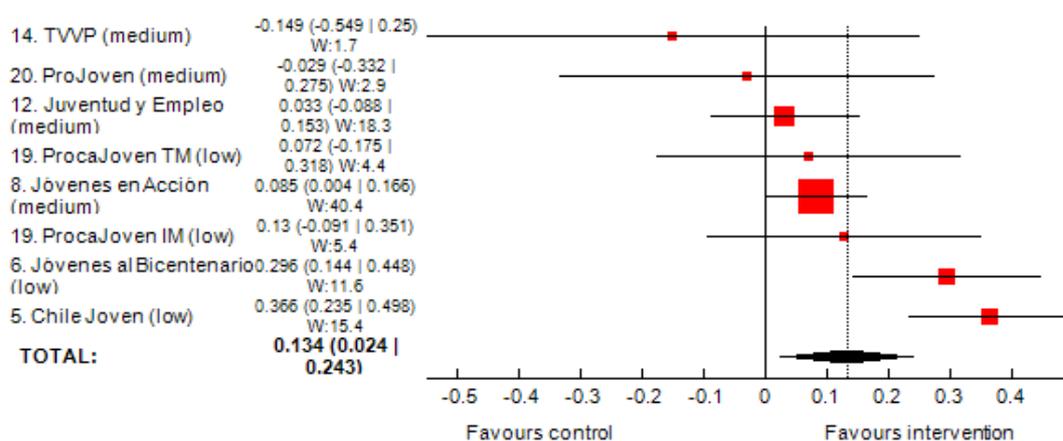
The pooled estimate of effect (g=0.134) suggests that the TVET interventions were, on average, effective; in other words, that those who experienced a TVET intervention had a greater chance of paid employment than those who did not. However, the high degree of heterogeneity between the studies (Q = 23.8; df = 7; p = 0.00124; I^2 = 70.6%; tau^2 = 0.0153) suggests differential effects across studies.

The first possible explanation that we considered for the pattern of variation seen in the meta-analysis was differences in study quality. Separate meta-analyses were conducted for (i) medium quality and (ii) low quality studies, and these were entered

into a sub-group analysis. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.2.

FIGURE 5.2: FOREST PLOT OF MEAN EFFECTS ON OVERALL PAID EMPLOYMENT (BY STUDY QUALITY)
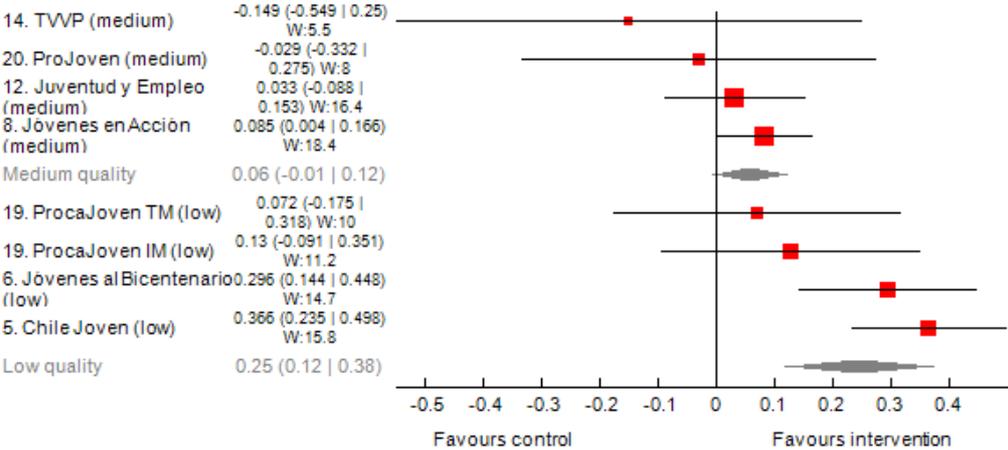
Treatment effects for the low quality studies, g=0.25 (95% CI [0.12, 0.38]), appear to be greater than for those rated medium quality, g=0.06 (95% CI [-0.01, 0.12]). Furthermore, the observed differences in mean effects were statistically significant (Q_b = 6.49; p = 0.0108).

Note: It was also observed that the majority of studies not included in this meta-analysis because effect sizes could not be calculated found that young people who had participated in TVET had a higher probability of being in paid employment than youth who had not participated (see Appendix 8.18).

5.2.1.1   *Does participation in different types of TVET have different effects on overall paid employment for young people?*

The TVET interventions evaluated in the seven studies included in the meta-analysis presented in Figure 5.1 comprised three distinct types (see Table 5.2). Separate meta-analyses were conducted for studies comprised of (i) two-phase TVET interventions and (ii) other TVET models. These meta-analyses were then entered into a sub-group analysis. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.3.
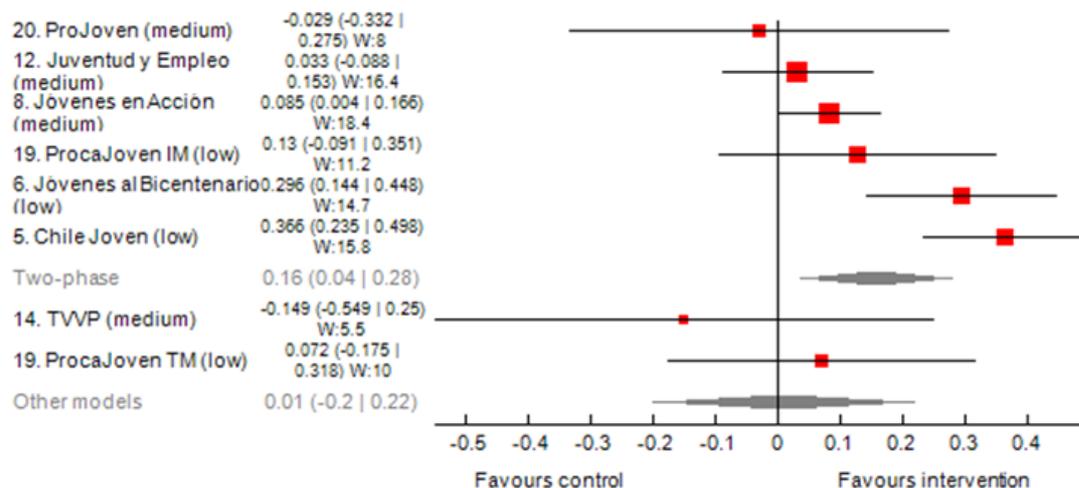
For two-phase TVET interventions, a weighted average effect size of g=0.16 was observed (95% CI [0.04, 0.28]). For other TVET models, the pooled estimate of effect was positive, but negligible (g=0.01), and the confidence intervals do not exclude a negative effect (95% CI [-0.2, 0.22]). However, although treatment effects for two-phase TVET interventions appear to be greater than for other TVET modalities, the observed differences in mean effects were not statistically significant (Q_b = 1.43; p = 0.231).

5.2.1.2 *Does participation in TVET have different effects on overall paid employment for young people in the short- and medium-term?*

In this section, we explore whether differences in time since completion of the training might be a cause of any observed variance in outcomes. Studies included in the review varied in the length of time that elapsed between completion of the intervention and measurement of its impact on paid employment. Six of the seven studies included in the meta-analysis for overall paid employment (see Figure 5.1) assessed the impact of the intervention at a single point in time after training had ended, either at approximately 6 months or around 12-15 months. The remaining study examined impacts over time, measuring outcomes at 6, 12, and 18 months (Espinoza, 2010). Information about individual studies is presented in Appendix 8.9. Separate meta-analyses were conducted for studies comprised of (i) short-term and (ii) medium-term follow-up periods. These meta-analyses were then entered into a sub-group analysis. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.4.
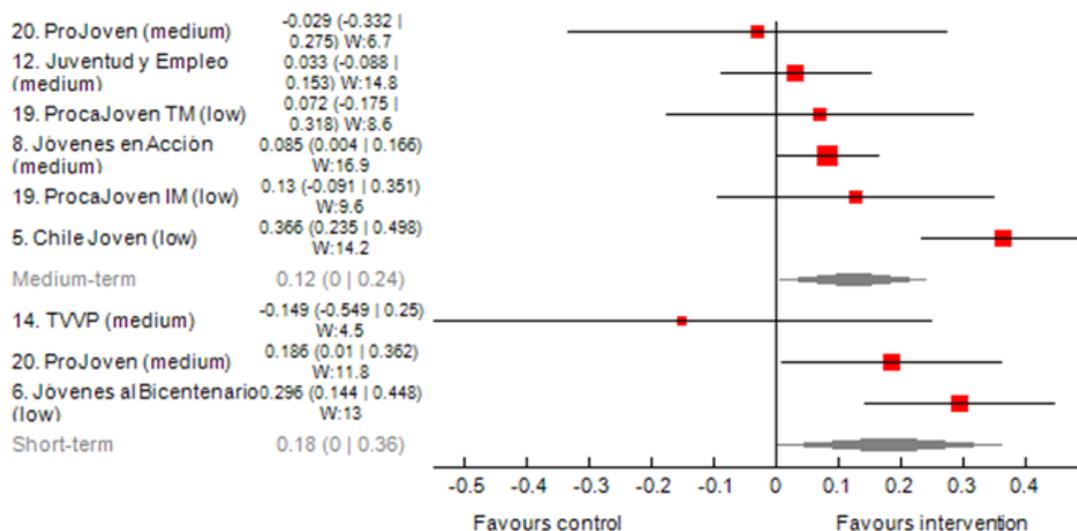
Short-term treatment effects, g=0.18 (95% CI [0, 0.36]), appeared to be greater than medium-term effects, g=0.12 (95% CI [0, 0.24]). However, the observed differences in mean effects were not statistically significant (Q_b = 0.273; p = 0.601). When the sub-group analysis was re-run with only those studies rated medium, the differences remained insignificant (Q_b = 0.000628; p = 0.98).

5.2.1.3 Does participation in TVET have different overall paid employment effects on different sub-groups of young people?

Several studies explored this issue. Fifteen studies disaggregated the employment effects of TVET on young people by gender; six studies estimated other sub-group impacts of TVET (see Appendix 8.12). Information about individual studies is presented in Appendix 8.9.

Gender differences

The data from the following five studies were amenable to meta-analysis. The studies of interventions nos. 8, 12, and 14 (Attanasio et al., 2011; Card et al., 2011; Hicks et al., 2011) were rated medium quality. The studies of interventions nos. 2 and 19 (Aedo & Nuñez, 2004; Ibarraran & Rosas-Shady, 2006) were rated low. Separate meta-analyses were conducted for studies comprised of samples with different gender compositions. These meta-analyses were then entered into a sub-group analysis. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.5.

FIGURE 5.5: FOREST PLOT OF MEAN EFFECTS ON OVERALL PAID EMPLOYMENT (BY GENDER)
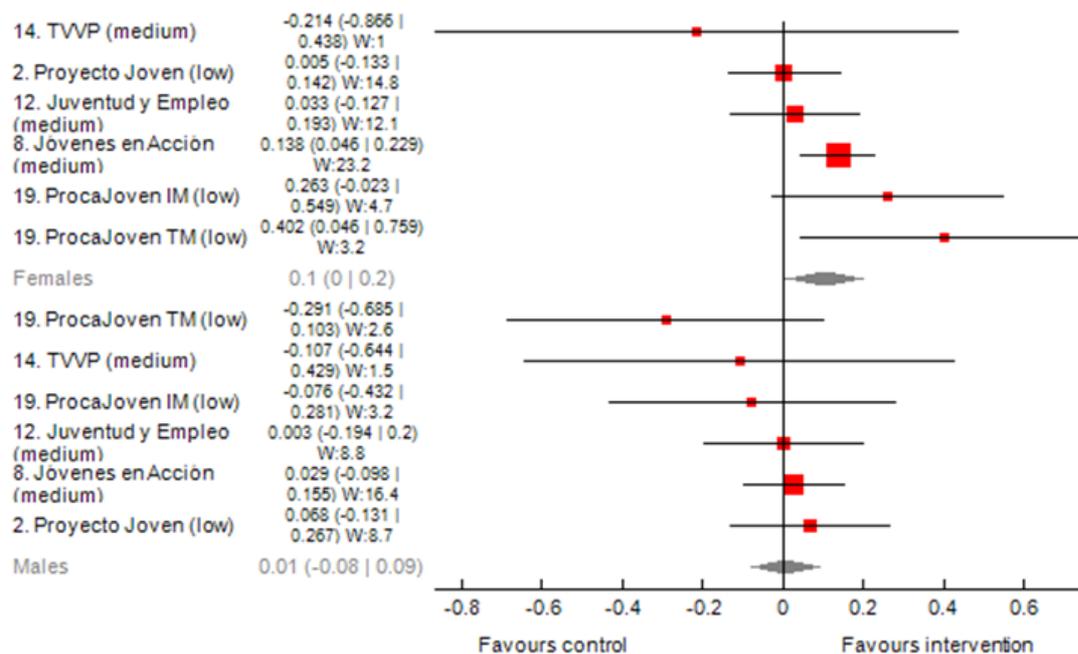
Treatment effects for female youth, g=0.1 (95% CI [0, 0.2]), appear to be slightly larger than for male youth, g=0.01 (95% CI [-0.08, 0.09]). However, the observed differences in mean effects were not statistically significant (Q_b = 2.1; p = 0.147). When the sub-group analysis was re-run with only those studies rated medium, the differences remained insignificant (Q_b = 1.49; p = 0.222).

Other population sub-group differences

The studies of interventions nos. 12 and 20 (Card et al., 2011; Espinoza, 2010) were rated medium quality. The studies of interventions nos. 18, 20, 19, and 20 (Delajara et al., 2006; Díaz & Jaramillo, 2006; Ibarraran & Rosas-Shady, 2006; Jaramillo et al., 2007) were rated low. As the two medium studies did not examine differential treatment effects for the same population sub-groups, no meta-analysis was performed.

One of the three studies that examined differences in impact by location (e.g., comparing impacts for young people from the capital city with those in other regions of the country) observed regional variation (Ibarraran & Rosas-Shady, 2006). Two of the three studies exploring treatment effect heterogeneity by level of education found larger positive effects for more educated workers, compared to the less educated (Card et al., 2011; Delajara et al., 2006). Two studies of the ProJoven intervention did not agree on the influence of poverty level. Espinoza (2010) observed that programme participation yielded no additional returns to individuals in the lowest household capita quartile prior to training, whereas Jaramillo et al.

(2007) concluded that that the strong treatment heterogeneity was not due to the variation in the initial poverty level of the beneficiaries. One of the two studies that were able to divide an evaluation sample of young people into two age groups found slightly higher point estimates among the youngest trainees in their sample (Díaz & Jaramillo, 2006). One study found that the programme increased the employment likelihood of individuals with no work experience prior to training (Espinoza, 2010).

5.2.2 Does participation in TVET have an effect on young people's chance of obtaining employment in the formal sector?

Several of the existing evaluations of TVET interventions have sought to capture not only whether trainees moved into employment as a result of the training, but also the quality of the position they secured. When talking about quality of employment, the focus is generally on the distinction between formal and informal employment. Whilst formal employment is government regulated, and workers are insured a wage and certain rights, informal employment tends to take place in unregistered enterprises, and often deprives people of financial stability and safe working environments.

In total, 10 studies assessed the impact of TVET interventions on formal employment (see Appendix 8.9 for details). Different benefit variables (proxies) were used to capture the quality of the employment position, including employment in a job with employer-provided health or social insurance, and/or formal written contract.

The reported data allowed the calculation of effect sizes from five of these studies, and these were combined in a meta-analysis. The studies of interventions nos. 8, 12, and 20 (Attanasio et al., 2011; Card et al., 2011; Espinoza, 2010) were rated medium quality. The studies of interventions nos. 5 and 19 (Aedo & Pizarro, 2004; Ibarraran & Rosas-Shady, 2006) were rated low. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.6.

The overall mean effect for formal paid employment is a standard mean difference of g=0.199. The confidence intervals do not cross the line of 'no effect' (95% CI [0.055, 0.344]). Although the analysis presented in Figure 5.7 provides evidence that the TVET interventions were, on average, effective, the heterogeneous nature of the distribution (Q = 11.1; df = 4; p = 0.0256; I^2 = 63.9%; tau^2 = 0.0131) suggests differential effects across studies.

FIGURE 5.6: FOREST PLOT OF MEAN EFFECTS ON FORMAL PAID EMPLOYMENT
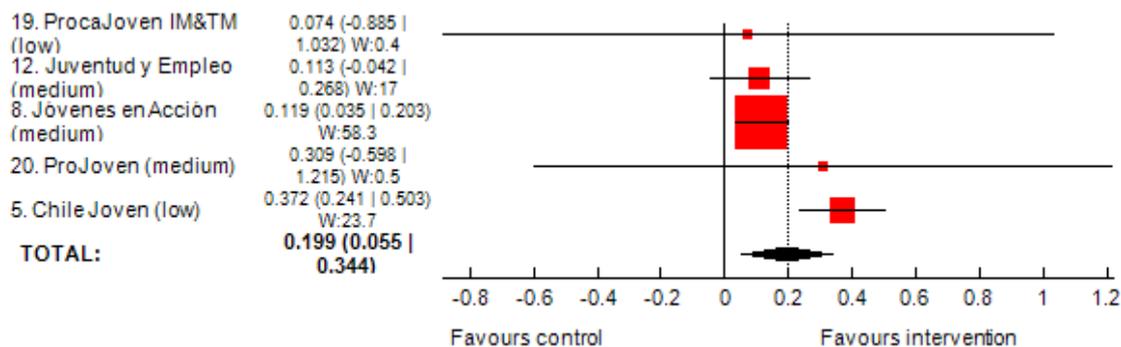
The first possible explanation that we considered for the pattern of variation seen in the meta-analysis was differences in study quality (see the forest plot in Figure 5.7). Treatment effects for the low quality studies, $g=0.37$ (95% CI [0.24, 0.5]), appeared to be greater than for those rated medium, $g=0.12$ (95% CI [0.05, 0.19]). Furthermore, the observed differences in mean effects were statistically significant ($Q_b = 10.6$; $p = 0.00116$).

FIGURE 5.7: FOREST PLOT OF MEAN EFFECTS ON FORMAL EMPLOYMENT (BY STUDY QUALITY)
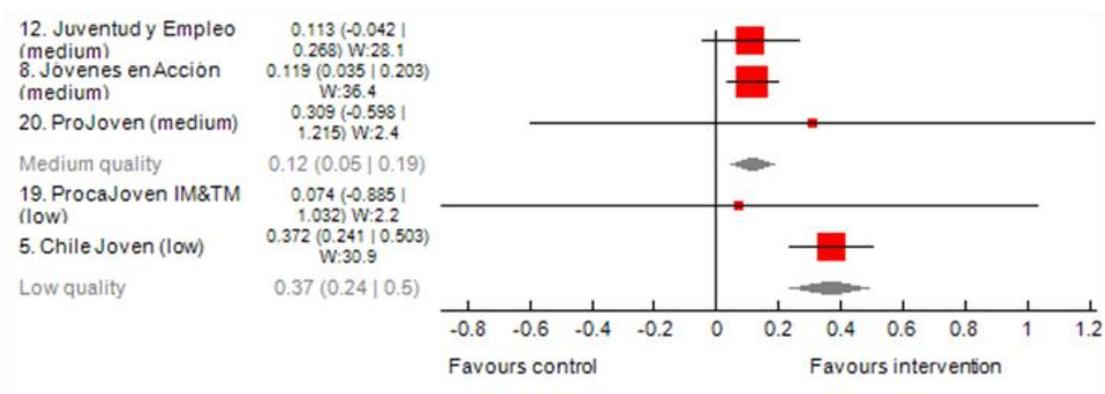
It was not possible to examine the variation in effect sizes (seen in Figure 5.6) by type of TVET or length of follow up due to an insufficient number of studies.

5.2.2.1 Does participation in TVET have different formal employment effects on different sub-groups of young people?

Several studies examined whether the impacts of TVET on formal employment differed according to population sub-group. Eight studies examined variation in treatment effects by gender, and three studies estimated differential effects for other sub-groups (see Appendix 8.12). Information about individual studies is presented in Appendix 8.9.

Gender differences

The studies of interventions nos. 8 and 20 (Attanasio et al., 2011; Espinoza, 2010) were rated medium quality. The studies of interventions nos. 5, 2, 20, 20, 19, and 16 (Aedo & Pizarro, 2004; Alzuá & Brassioli, 2006; Chong et al., 2008; Díaz & Jaramillo, 2006; Ibarraran & Rosas-Shady, 2006; van Gameren, 2010) were rated low. Effect sizes were computable for one medium quality study, and so no meta-analysis was performed. On the whole, studies observed relatively similar effects for males and females; although some authors found that young women benefitted most out of programme participation.

Other population sub-group differences

The study of intervention no. 20 (Espinoza, 2010) was rated medium quality. Two other studies of interventions nos. 19 and 20 (Díaz & Jaramillo, 2006; Ibarraran & Rosas-Shady, 2006) were rated low. The study by Espinoza (2010) found that training produced additional returns to those individuals with no work experience and those in the lowest household income quartile. One study found slightly larger programme effects on the likelihood of having a formal job among 16-20 year olds than among 21-25 year olds for some of the year cohorts, and for the others the larger effects were observed for the group aged 21-25 years (Díaz & Jaramillo, 2006). Ibarraran & Rosas-Shady (2006) reported that treatment effects were relatively evenly distributed across the sample.

5.2.3 Does participation in TVET have an effect on young people's chance of obtaining self-employment?

Three studies separated salaried and self-employed workers and examined the impact of TVET interventions on the probability of self-employment among young people. The study of intervention no. 14 (Hicks et al., 2011) was rated medium quality. The studies of interventions nos. 18 and 17 (Delajara et al., 2006; López-Acevedo, 2003) were rated low. As there was a single medium study, no meta-analysis was performed.

Two of the three studies (Hicks et al., 2011; López-Acevedo, 2003) found a slight positive treatment effect. The remaining study (Delajara et al., 2006) found an irregular effect on self-employment; in some years, treatment effects were positive, and in other years they were negative.

5.2.3.1 Does participation in TVET have different self-employment effects on different sub-groups of young people?

Two studies addressed this question. Delajara et al. (2006) reported that due to insufficient observations, many sub-groups could not be evaluated, and no clear pattern could be described. The study by Hicks et al. (2011) examined variation by

gender but found treatment effects to be relatively evenly distributed across the sample.

5.3 INCOME

The majority of studies examined the impact of a TVET intervention on monthly earnings, but smaller numbers also considered their impact on earnings from self-employment, hourly wages, or household/monthly income (see Table 5.4).

TABLE 5.4: INCOME-RELATED OUTCOMES

Outcome (n)	Studies	Analysis
Monthly earnings (22 studies)	Acero et al. (2011); Aedo & Nuñez (2004); Aedo & Pizarro (2004); Alzuá & Brassioli (2006); Alzuá et al. (2007); Analítica Consultores (2006); Attanasio et al. (2011); Benus et al. (2001); Bidani et al. (2002); Card et al. (2011); Chong & Galdo (2006); Delajara et al. (2006); Díaz & Jaramillo (2006); Elías et al. (2004); Espinoza (2010); Hicks et al. (2011); Ibarraran & Rosas-Shady (2006); Jaramillo et al. (2007); López-Acevedo (2003); Medina & Nuñez (2005); Ñopo et al. (2007); van Gameren (2010)	Section 5.3.1
Earnings from self-employment earnings/profits (3 studies)	Attanasio et al. (2011); Delajara et al. (2006); Hicks et al. (2011)	Section 5.3.2
Hourly wage (rate of pay) (5 studies)	Card et al. (2011); Díaz & Jaramillo (2006); Ibarraran & Rosas-Shady (2006); López-Acevedo (2003); Ñopo et al. (2007)	Section 5.3.3

5.3.1 Does participation in TVET have an effect on young people's earnings?

The available data allowed the calculation of effect sizes for eight studies. For this outcome, the studies of interventions nos. 8, 12, 2, and 14 (Attanasio et al., 2011; Card et al., 2011; Elias et al., 2004; Hicks et al., 2011) were rated medium quality, while studies of interventions nos. 6, 5, 20, and 19 (Acero et al., 2011; Aedo & Pizarro, 2004; Espinoza, 2010; Ibarraran & Rosas-Shady, 2006) were rated low.

Effect sizes from these eight studies were combined. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.8. The pooled estimate of effect is g=0.127 (95% CI [0.043, 0.21]).

FIGURE 5.8: FOREST PLOT OF MEAN EFFECTS ON EARNINGS
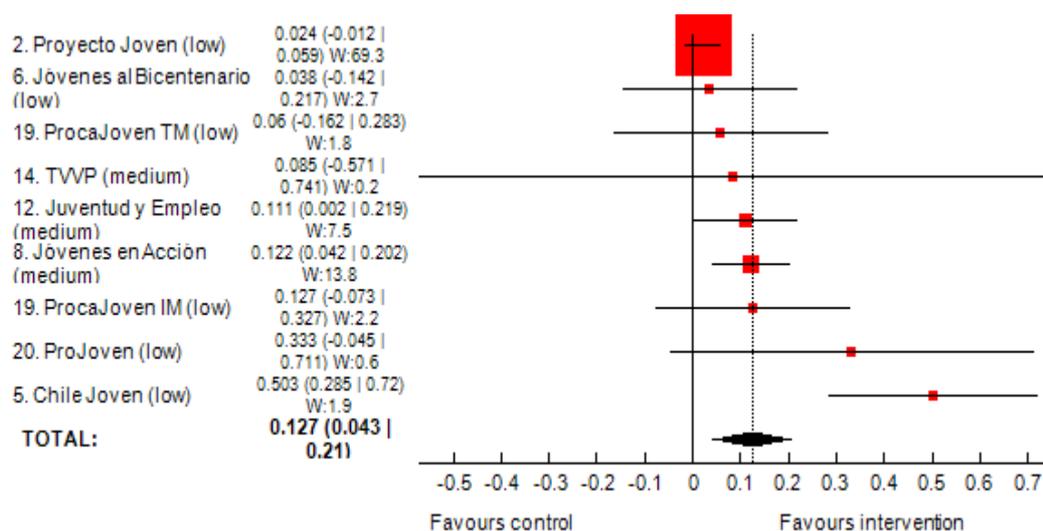
The overall mean effect size (*g*=0.127) suggests that, on average, young people who received a TVET intervention have higher monthly earnings than those who did not. The confidence interval for this point estimate (95% CI [0.043, 0.21]) is relatively precise and does not cross the line of 'no effect.' However, the results of the statistical tests for homogeneity (Q = 25.5; df = 8; p = 0.00128; I^2 = 68.6%; tau^2 = 0.00815) suggest differential effects across studies.

We first explored whether the pattern of variation seen in the meta-analysis could be explained by differences in study quality (see the forest plot in Figure 5.9). It looks as if treatment effects for the low quality studies, *g*=0.15 (95% CI [0.01, 0.3]), are very similar to those rated medium, *g*=0.12 (95% CI [0.05, 0.18]). The slight differences in mean effects were statistically insignificant (Q_b = 0.204; p = 0.652).

FIGURE 5.9: FOREST PLOT OF MEAN EFFECTS ON EARNINGS (BY STUDY QUALITY)
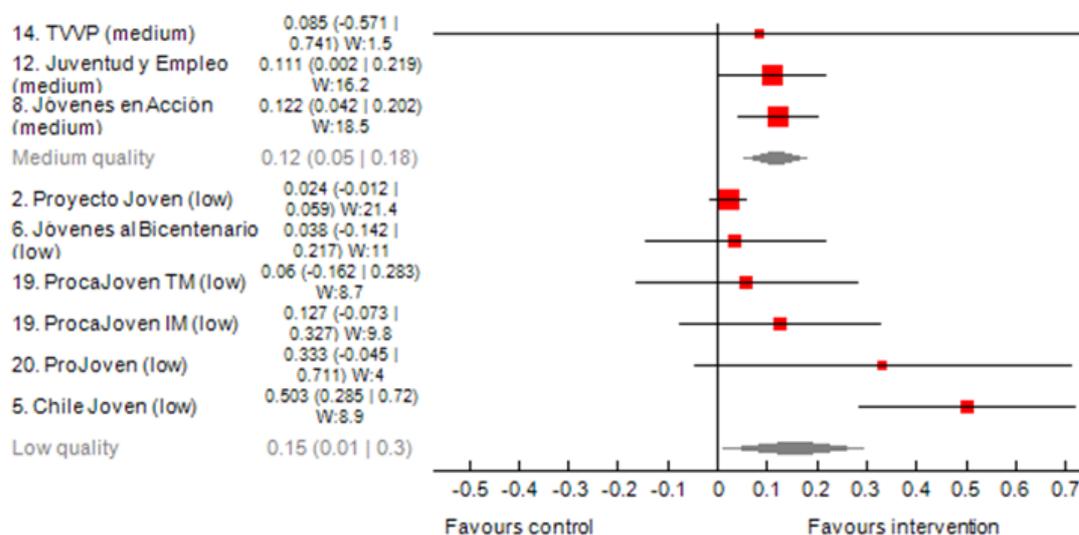
5.3.1.1 Does participation in different types of TVET have different effects on young people's earnings?

The TVET interventions evaluated in the eight studies included in the meta-analysis presented in Figure 5.8 comprised three distinct types (see Table 5.2). Separate meta-analyses were conducted for studies comprised of (i) two-phase TVET interventions and (ii) other TVET models. These meta-analyses were then entered into a sub-group analysis. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.10.

FIGURE 5.10: FOREST PLOT OF MEAN EFFECTS ON EARNINGS (BY PROGRAMME TYPE)
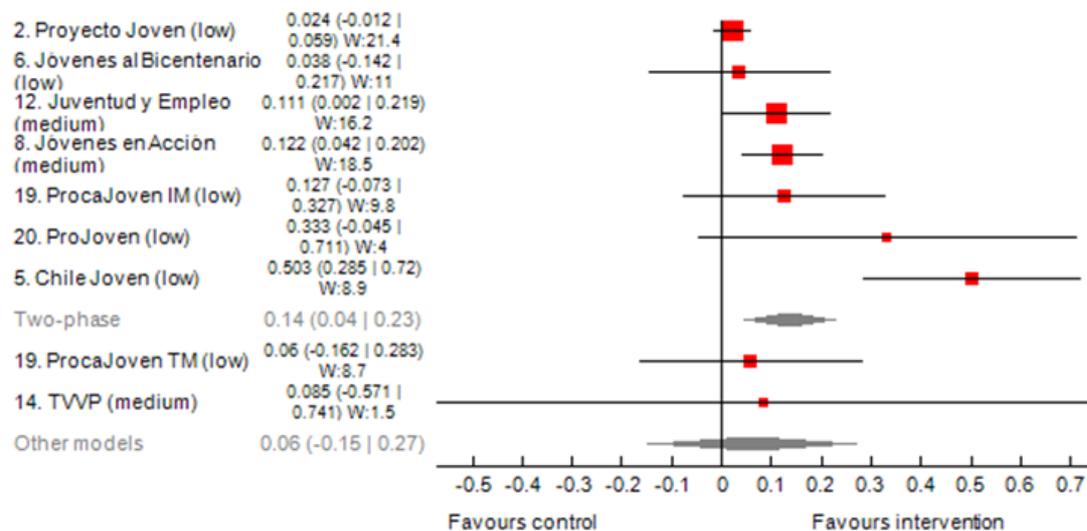
For two-phase TVET interventions, the individual effect sizes were meta-analysed to produce a weighted average effect size of g=0.14 (95% CI [0.04, 0.23]). For other TVET models, a positive pooled estimate of effect was also observed (g=0.06), but the confidence intervals do not exclude a negative effect (95% CI [-0.15, 0.27]). Although the treatment effect appears to be slightly larger for two-phase models of TVET than for other models, the observed differences in mean effects were not statistically significant (Q_b = 0.397; p = 0.529).

5.3.1.2 Does participation in TVET have different effects on young people's earnings in the short- and medium-term?

Seven of the eight studies included in the meta-analysis for monthly earnings (see Figure 5.8) followed up participants once, after training had ended. The remaining study examined the impact of the intervention over time, measuring outcomes at 6, 12, and 18 months (Espinoza, 2010). Information about individual studies is presented in Appendix 8.9.

Separate meta-analyses were conducted for studies comprised of (i) short-term and (ii) medium-term follow-up periods. These meta-analyses were then entered into a sub-group analysis. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.11. Short-term treatment effects, g=0.22 (95% CI [-0.13, 0.58]), appeared to be greater than medium-term effects, g=0.14 (95% CI [0.05, 0.24]). However, the observed differences in mean effects were not statistically significant (Q_b = 0.186; p = 0.666).

FIGURE 5.11: FOREST PLOT OF MEAN EFFECTS ON EARNINGS (BY LENGTH OF FOLLOW-UP)
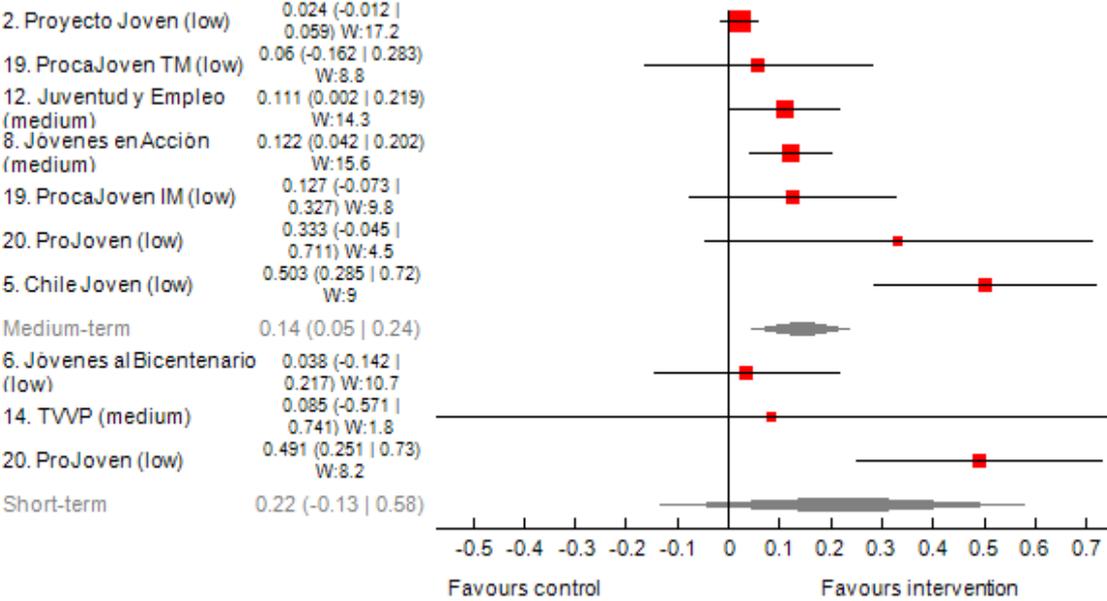
5.3.1.3 *Does participation in TVET have different earnings effects on different sub-groups of young people?*

Several studies examined whether some groups of young people benefited more than others, in terms of post-training earnings. Sixteen studies disaggregated the earnings effects of TVET on young people by gender, and seven studies estimated additional sub-group impacts (see Appendix 8.12). Information about individual studies is presented in Appendix 8.9.

Gender differences

Of these 16 studies, 6 were amenable to meta-analysis. The studies of interventions nos. 8, 12, and 14 (Attanasio et al., 2011; Card et al., 2011; Hicks et al., 2011) were rated medium quality. The studies of interventions nos. 2, 19, 9, and 11 (Aedo & Nuñez, 2004; Ibarraran & Rosas-Shady, 2006; Medina & Nuñez, 2005) were rated low. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.12.

FIGURE 5.12: FOREST PLOT OF MEAN EFFECTS ON EARNINGS (BY GENDER)
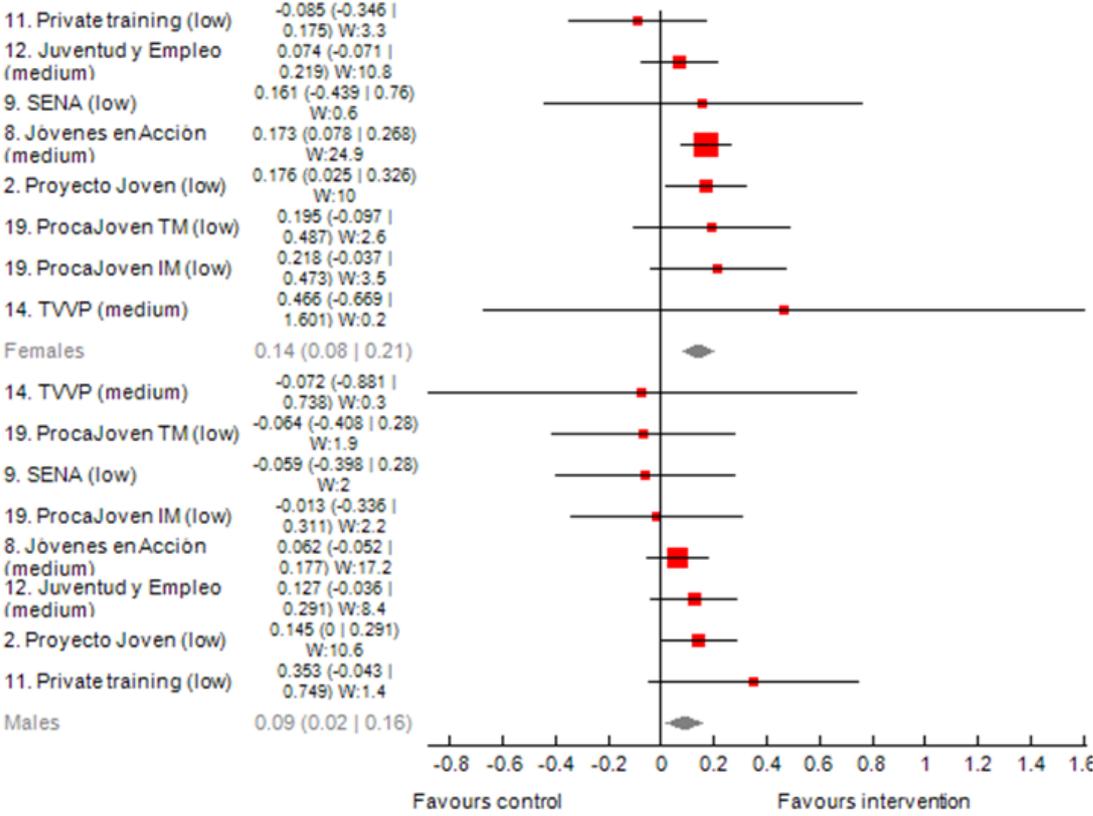
For female youth, the individual effect sizes were meta-analysed to produce a weighted average effect size of $g=0.14$ (95% CI [0.08, 0.21]). For male youth, the pooled estimate of effect was also positive ($g=0.09$) and again the confidence intervals do not include zero (95% CI [0.02, 0.16]). However, the observed differences in mean effects were not statistically significant ($Q_b = 1.26$; $p = 0.262$). When the sub-group analysis was re-run with only those studies rated medium, the differences remained insignificant ($Q_b = 1.01$; $p = 0.315$).

Other population sub-group differences

The study of intervention no. 12 (Card et al., 2011) was rated medium quality. The studies of interventions nos. 18, 20, 2, 20, 19, and 20 (Delajara et al., 2006; Díaz & Jaramillo, 2006; Elías et al., 2004; Espinoza, 2010; Ibarraran & Rosas-Shady, 2006; Jaramillo et al., 2007) were rated low. Three of the four studies that explored regional variation in impact observed some differences between sub-groups (Card et al., 2011; Elías et al., 2004; Ibarraran & Rosas-Shady, 2006). Two of the three studies exploring treatment effect heterogeneity by level of education found larger positive effects for more educated workers, compared to the less educated (Card et al., 2011; Delajara et al., 2006). Two evaluations of the ProJoven programme disagreed on the influence of poverty level on the earnings returns to young people. Espinoza (2010) found that ProJoven yields additional returns for those trainees in

the lowest household income per capita quartile. In contrast, Jaramillo et al. (2007) found the programme to be 'equity enhancing…as the evidence indicates similar returns for participants along varying poverty lines' (p. 43). One of the two studies exploring treatment effects by age found slightly higher point estimates among the youngest trainees in their samples (Díaz & Jaramillo, 2006), whereas the other study (Card et al., 2011) found the estimated impacts on monthly earnings to be fairly similar for younger and older workers. Finally, Espinoza (2010) found that ProJoven yields additional returns for those trainees with no work experience prior to training.

5.3.1.4 *Does participation in TVET have different effects on young people's earnings according to the quality of the training?*

Two studies examined the natural hypothesis that higher quality training will have a larger impact on participant outcomes (Card et al., 2011; Chong & Galdo, 2006). Card et al. (2011) reported that comparisons between treatment and control outcomes within each quality group 'showed no evidence of a large or systematic quality effect' (p. 290).[25] In contrast, Chong and Galdo (2006), who had a larger data set of trainee and training provider characteristics, found that young people attending high-quality training courses showed much higher impacts on monthly earnings than those attending low-quality courses.

5.3.2 Does participation in TVET have an effect on young people's self-employment earnings?

Two of the three studies that examined the impact of TVET on young people's earnings were amenable to meta-analysis. The studies of interventions nos. 8 and 14 (Attanasio et al., 2011; Hicks et al., 2011) were both rated medium quality. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.13.

FIGURE 5.13: FOREST PLOT OF MEAN EFFECTS ON SELF-EMPLOYMENT EARNINGS
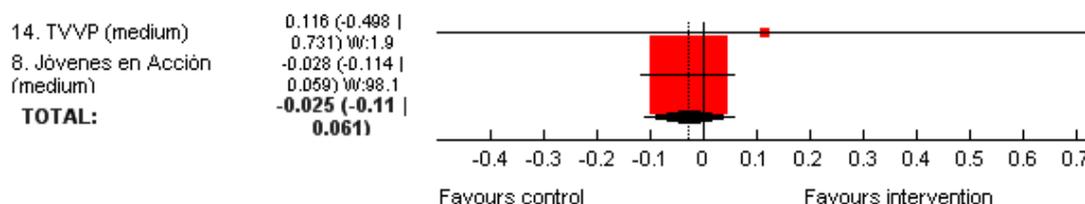
The overall mean effect size (g=-0.025) suggests that, on average, young people who received a TVET intervention had lower self-employment earnings than those who

---

[25] Although not explicitly stated in the paper, it is probable that Card et al. (2011) conducted this analysis for other outcomes in addition to monthly earnings (with this finding relating to them all).

did not, although this is a relatively imprecise estimate based on only two studies (95% CI [-0.11, 0.061]). The results of the statistical test for homogeneity was non-significant (Q = 0.206; df = 1; p = 0.65; I^2 = 0%; tau^2 = 0).

It was not possible to examine the variation in effect sizes (seen in Figure 5.13) by type of TVET or length of follow up due to an insufficient number of studies.

5.3.2.1 *Does participation in TVET have different self-employment earnings effects on different sub-groups of young people?*

Three studies examined whether the treatment effect on self-employment earnings varied across population groups. One study (Delajara et al., 2006) reported that, due to insufficient observations, many sub-groups could not be examined, and no clear pattern could be described. The remaining two studies examined gender differences only.

Gender differences

The studies of interventions nos. 8 and 14 (Attanasio et al., 2011; Hicks et al., 2011) were rated medium quality. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.14.

FIGURE 5.14: FOREST PLOT OF MEAN EFFECTS ON SELF-EMPLOYMENT EARNINGS (BY GENDER)
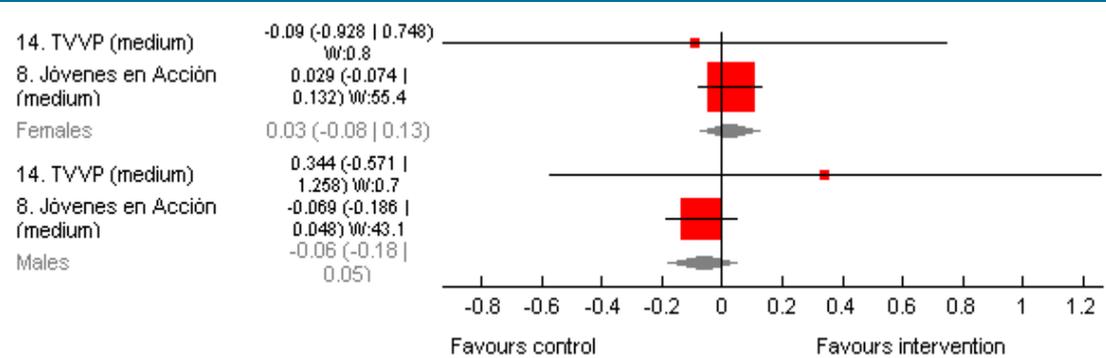

For female youth, the individual effect sizes were meta-analysed to produce a weighted average effect size of g=0.03 (95% CI [-0.08, 0.13]). For male youth, the pooled estimate of effect was negative (g=-0.06) although the confidence intervals cross zero (95% CI [-0.18, 0.05]). The observed differences in mean effects, however, were not statistically significant (Q_b = 1.27; p = 0.259).

5.3.3 Does participation in TVET have an effect on young people's hourly wages?

Five studies measured changes in the hourly wage received by workers. The study of intervention no. 12 (Card et al., 2011) was rated medium quality. The study of interventions nos. 20, 19, 17, and 20 (Díaz & Jaramillo, 2006; Ibarraran & Rosas-

Shady, 2006; López-Acevedo, 2003; Ñopo et al., 2007) were rated low. The findings of the five studies were generally consistent, with four studies observing that young people who had participated in TVET experienced higher hourly rates of pay than those who had not participated.

5.3.3.1 *Does participation in TVET have different hourly wage effects on different sub-groups of young people?*

Three of these five studies assessed whether TVET interventions have different hourly wage effects on different population sub-groups. Díaz and Jaramillo (2006) found higher programme effects for women. Ñopo et al. (2007) found that 12 months after training ended, males were benefiting more than females. However, at 6 and 18 months, gender differences in income per hour were minor. Ibarraran and Rosas-Shady (2006) reported that the positive effects they observed were relatively evenly distributed across the whole sample. Díaz and Jaramillo (2006) found that effects tended to be higher for 16-20 year old youths than for those aged 21-25 years.

5.4 HOURS WORKED

A third of all studies examined the impact of a TVET intervention on the number of weekly hours worked by young people, whilst a single study also measured this outcome amongst the self-employed (see Table 5.5).

TABLE 5.5: HOURS WORKED

Outcome (n)	Studies	Analysis
Weekly hours worked, among the employed (9 studies)	Acero et al. (2011); Attanasio et al. (2011); Card et al. (2011); Díaz & Jaramillo (2006); Hicks et al. (2011); Ibarraran & Rosas-Shady (2006); López-Acevedo (2003); Mensch et al. (2004); Ñopo et al. (2007)	Section 5.4.1
Weekly hours worked, among the self-employed (1 study)	Hicks et al. (2011)	Section 5.4.2

5.4.1 Does participation in TVET have an effect on the number of weekly hours worked in paid employment by young people?

The data from five studies allowed calculation of effect sizes. The studies of interventions nos. 8, 12, and 14 (Attanasio et al., 2011; Card et al., 2011; Hicks et al., 2011) were rated medium quality. The studies of interventions nos. 6 and 19 (by Acero et al., 2011; Ibarraran & Rosas-Shady, 2006) were rated low.

FIGURE 5.15: FOREST PLOT OF MEAN EFFECTS ON HOURS WORKED
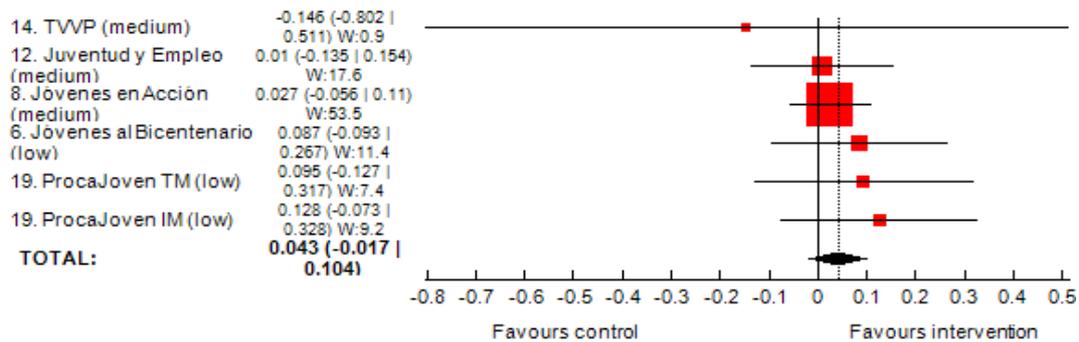
The forest plot in Figure 5.15 shows the mean effect size and the confidence intervals for each study. The majority of individual studies show a positive effect on the number of weekly hours worked. When the studies were pooled, the effect size favours the intervention (g= 0.043). However, as the confidence interval crosses zero (95% CI [-0.017, 0.104]) the result does not exclude a possible negative effect. Although the results of the statistical tests for homogeneity (Q = 1.8; df = 5; p = 0.876; I^2 = 0%; tau^2 = 0) suggest that variability in effect sizes between studies was not larger than expected from sampling error, visual indicators suggest that there is a degree of heterogeneity between studies.

We first explored whether the pattern of variation seen in the meta-analysis could be explained by differences in study quality (see the forest plot in Figure 5.16). Although treatment effects for the low quality studies, g=0.1 (95% CI [-0.01, 0.22]), appear to slightly larger than for those rated medium, g=0.02 (95% CI [-0.05, 0.09]), the observed differences in mean effects were statistically insignificant (Q_b = 1.41; p = 0.234).

FIGURE 5.16: FOREST PLOT OF MEAN EFFECTS ON HOURS WORKED (BY STUDY QUALITY)
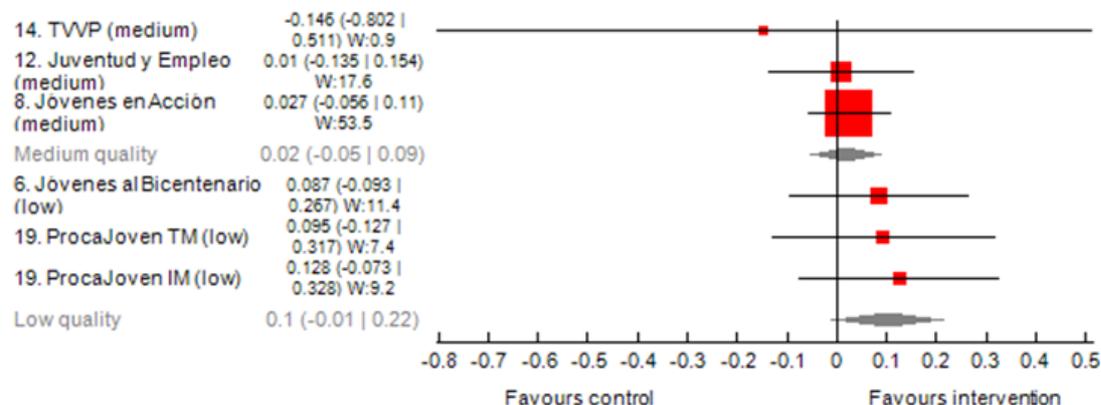
Note: The majority of the studies not included in the meta-analysis because effect sizes could not be calculated also found that the number of weekly hours worked by young people increased as a result of participating in TVET (see Appendix 8.20).

5.4.1.1 Does participation in different types of TVET have different effects on the weekly hours worked by young people?

The TVET interventions evaluated in the five studies included in the meta-analysis presented in Figure 5.15 comprised three distinct types (see Table 5.2). Separate meta-analyses were conducted for studies comprised of (i) two-phase TVET interventions and (ii) other TVET models. These meta-analyses were then entered into a sub-group analysis. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.17.

FIGURE 5.17: FOREST PLOT OF MEAN EFFECTS ON HOURS WORKED (BY PROGRAMME TYPE)
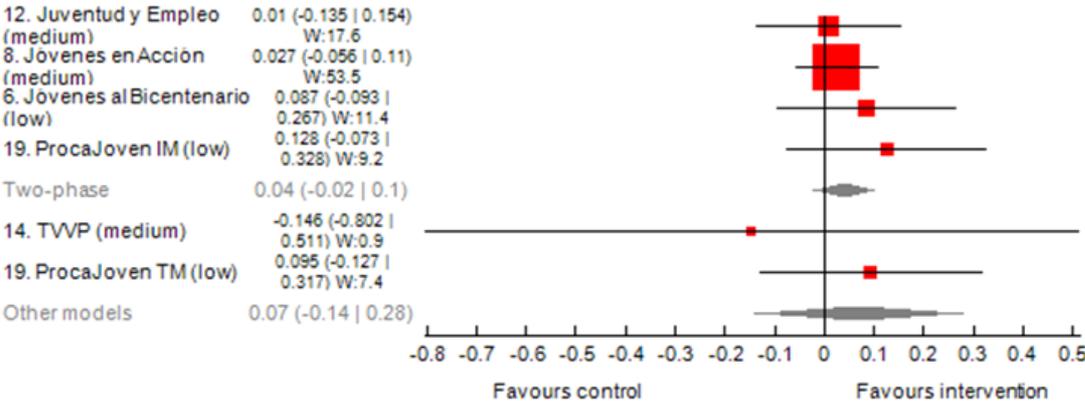
For two-phase TVET interventions, the individual effect sizes were meta-analysed to produce a weighted average effect size of $g=0.04$ (95% CI [-0.02, 0.1]). For other TVET models, a pooled estimate of effect of similar magnitude was observed ($g=0.07$), and again the confidence intervals do not exclude a negative effect (95% CI [-0.14, 0.28]). The observed differences in mean effects were not statistically significant ($Q_b = 0.0677$; $p = 0.795$).

5.4.1.2 Does participation in TVET have different effects on the weekly hours worked by young people in the short- and medium-term?

All five studies included in the meta-analysis for weekly hours worked (see Figure 5.15) measured outcomes at a single point in time after training ended. Information about individual studies is presented in Appendix 8.9.

Separate meta-analyses were conducted for studies comprised of (i) short-term and (ii) medium-term follow-up periods. These meta-analyses were then entered into a sub-group analysis. The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.18. Short-term treatment effects, $g=0.07$ (95%

CI [-0.1, 0.24]), were very similar to medium-term effects, g=0.04 (95% CI [-0.03, 0.1]). The observed differences in mean effects were not statistically significant (Q_b = 0.109; p = 0.741).

FIGURE 5.18: FOREST PLOT OF MEAN EFFECTS ON HOURS WORKED (BY LENGTH OF FOLLOW-UP)
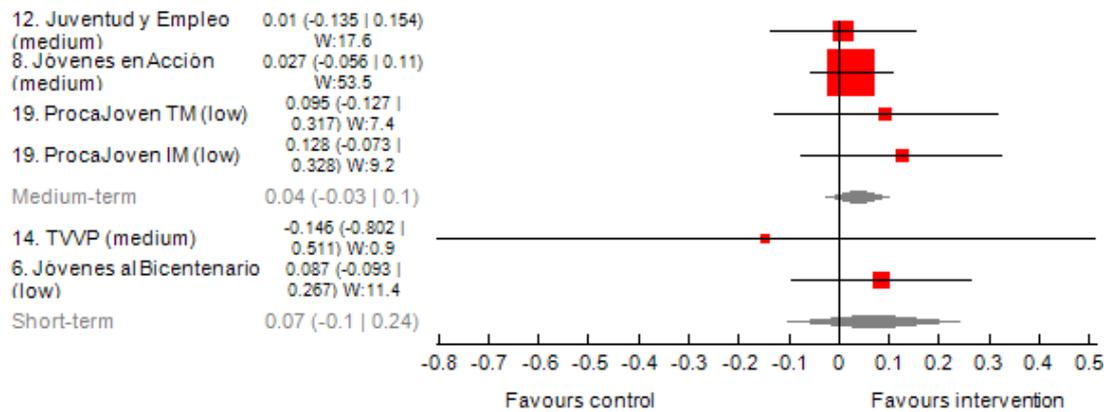
### 5.4.1.3 *Does participation in TVET have different weekly hours effects on different sub-groups of young people?*

Several studies examined whether the impacts of TVET on weekly hours were the same for different types of trainee. Five studies disaggregated the effects on weekly hours by gender, and two studies explored a number of additional subgroup impacts (see Appendix 8.12). Information about individual studies is presented in Appendix 8.9.

Gender differences

Three of the five studies had data that allowed effect size calculations. The studies evaluating interventions nos. 8 and 14 were rated medium quality (Attanasio et al., 2011; Hicks et al., 2011). The study of intervention no. 19 was rated low (Ibarraran & Rosas-Shady, 2006). The mean effect size and confidence intervals for each study are shown in the forest plot in Figure 5.19.

FIGURE 5.19: FOREST PLOT OF MEAN EFFECTS ON HOURS WORKED (BY GENDER)
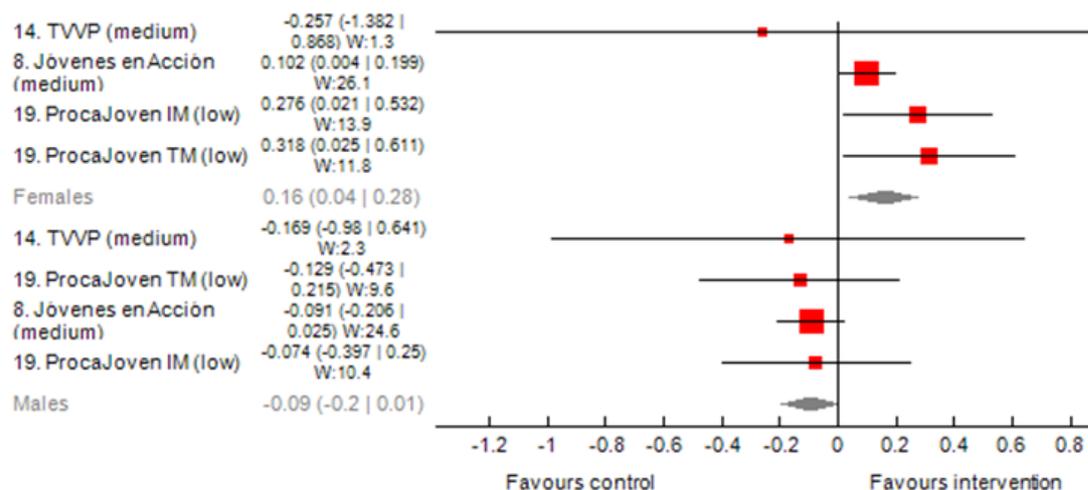
For female youth, the individual effect sizes were meta-analysed to produce a weighted average effect size of g=0.16 (95% CI [0.04, 0.28]). For male youth, the pooled estimate of effect was negative (g=-0.09; 95% CI [-0.2, 0.01]). The observed differences in mean effects were statistically significant (Q_b = 10.1; p = 0.00151). When the sub-group analysis was re-run with only those studies rated medium, the differences remained significant (Q_b = 6.28; p = 0.0122).

Other population sub-group differences

Two studies estimated additional sub-group impacts of TVET on the number of weekly hours worked by young people. Both studies were rated low. When splitting the sample by age, Díaz & Jaramillo (2006) did not observe a clear difference in favour of either group (16-20 years or 21-25 years). Ibarraran & Rosas-Shady (2006) examined a number of different sub-groups and found effects relatively evenly distributed across the sample.

5.4.2 Does participation in TVET have an effect on the number of weekly hours worked in self-employment by young people?

One medium quality study (Hicks et al., 2011) examined the impact that training (intervention no. 14) had on the number of hours worked in self-employment. It found that young people in the treatment group worked fewer hours than their control counterparts. A similar pattern was observed for males and females in the sample.

5.5 OTHER OUTCOMES

5.5.1 Does participation in TVET encourage job searching amongst young people?

One study examined the impact of participation in TVET (following receipt of a training voucher) on the length of time spent job searching. It found that young people in the intervention group spent less time on job search compared to their control counterparts, where this gap was especially pronounced among men (Hicks et al., 2011).

5.5.2 Does participation in TVET lead to further additional training?

One study examined whether recipients of training went on to participate in further training activities (López-Acevedo, 2003). It found that participants of TVET interventions were more likely to participate in future training than control group individuals who had received no training.

5.6 PUBLICATION BIAS

One of the great problems with systematic reviews is that not all studies carried out are published. Combining only published studies, which are more likely to have statistically significant results, may lead to an over-optimistic conclusion. As only 10 of the 26 studies meeting the eligibility criteria for the review were included in the meta-analyses, we did not attempt to detect or exclude the existence of publication bias using statistical methods. Arguably, a more fruitful 'assessment' of publication bias can be achieved through a discussion about the strengths and limitations of our search.

We attempted to minimise the possibility of publication bias by conducting an extensive systematic search to identify both published and unpublished research literature. An initial electronic search of major and specialist databases was supplemented with hand searches of relevant websites, contacts with authors and other relevant stakeholders (such as government agencies), reference checking of included studies and relevant reviews, and forward citation tracking. Included studies were not limited to those published in English, and additional project time and financial resources were secured from the funder in order to allow the inclusion of four studies published in Spanish. Only a third of included studies were identified through the electronic search. The large number of unpublished technical reports in the sample (only 4 of the 26 included studies were published in academic journals) also attests to the effort to minimise the possibility of publication bias. However, although not limited to studies written in English, language bias was not fully avoided as the literature search did not involve searching non-English language databases and websites. Considering the number of TVET interventions from Latin America evaluated by the included studies, it is possible that the review could have benefited from Spanish-language sources of documents. There may be eligible studies that are accessible only through such sources. Furthermore, as highlighted in Chapter 3, four studies published in Spanish were not included in the review (although they all examined an intervention that has been evaluated many times and is included in the review). Finally, we did not specifically search databases of theses; although it is unclear how fruitful such as strategy would have been, given the review's focus on literature from low- and middle-income countries. Taken together, there are a number of suggestions that the evaluation literature relating to TVET in developing countries is hard to locate; nonetheless, the possibility remains that some studies (in addition to the four known Spanish papers) may have been missed.

6 Discussion

6.1 SUMMARY OF EVIDENCE

This review set out to examine the potential of technical and vocational education and training to improve the employment and employability of young people in developing countries. A comprehensive search for published and unpublished studies yielded 3 RCT studies and 23 quasi-experimental studies that met the inclusion criteria. The 26 studies included in the review evaluated 20 different interventions providing TVET opportunities to young people. Fourteen of the 20 interventions were located in Central/South American countries, with 3 situated in Asia, 1 in Africa, and 2 in Europe. Settings include 10 upper-middle income countries (Argentina, Bosnia and Herzegovina, Brazil, Chile, China, Colombia, Dominican Republic, Latvia, Mexico, Panama, and Peru); 2 lower-middle income countries (India and Bhutan); and 1 low-income country (Kenya). The most commonly measured outcomes were paid employment and monthly earnings. Studies examined gross employment only, with none measuring changes in net employment. Intermediate outcomes were examined in only two studies.

A summary of the evidence detailed in the previous chapter is presented next. Although all 26 studies were reviewed, and there is reference in Chapter 5 to some of the findings from the 16 studies for which we were unable to calculate effect sizes, the synthesis was weighted towards the meta-analytic investigations. Utilising more of the findings from the 16 studies, and incorporating them more comprehensively into the synthesis, requires additional resources. At this stage, it would be premature to say more than the majority of these 16 studies—based on an assessment of the observed 'direction of effects'—appear to support the findings from the meta-analyses. The following summary of evidence, and the conclusions we draw about 'what works,' focuses solely on the results of the statistical analyses of the 10 studies. Our approach to summarising the evidence, which in turn forms the basis of identifying implications for policy and practice, drew on the interpretation framework presented in Appendix 8.21.

Employment

- The overall mean effect of TVET on paid employment was positive and significant (g=0.134, 95% CI [0.024, 0.243]).

- However, significant heterogeneity was observed (Q = 23.8; df = 7; p = 0.00124; I^2 = 70.6%; tau^2 = 0.0153).

- Four variables were tested for moderating effects.

    - Statistically significant differences in mean effects were observed between studies according to study quality (Q_b = 6.49; p = 0.0108). Treatment effects for the low quality studies, g=0.25 (95% CI [0.12, 0.38]), were greater than for those studies rated medium quality, g=0.06 (95% CI [-0.01, 0.12]).

    - No significant differences in mean effects were observed between studies according to type of TVET intervention (Q_b = 1.43; p = 0.231), length of follow-up period (Q_b = 0.273; p = 0.601), or gender (Q_b = 2.1; p = 0.147).

- As there is evidence of a statistically significant relationship between study quality and effect size, it is reasonable to conclude that the overall mean effect may be inflated and that our conclusions about treatment effect on paid employment should be based only on those studies rated medium quality. The mean estimate for these studies is very small (g=0.06) and non-significant (95% CI [-0.01, 0.12]).

- Summary: Our interpretation is that there is only weak evidence that TVET interventions are, on average, effective (relative to no intervention) at increasing the probability of having paid employment for young people in LMICs. Furthermore, the observed effect was very small.

Formal employment

- The overall mean effect of TVET on formal employment was positive and significant (g=0.199, 95% CI [0.055, 0.344]).

- However, significant heterogeneity was observed (Q = 11.1; df = 4; p = 0.0256; I^2 = 63.9%; tau^2 = 0.0131).

- One variable was tested for moderating effects.

    - Statistically significant differences in mean effects were observed between studies according to study quality (Q_b = 10.6; p = 0.00116). Treatment effects for the low quality studies, g=0.37 (95% CI [0.24, 0.5]), were greater than for those studies rated medium, g=0.12 (95% CI [0.05, 0.19]).

- As there is evidence of a statistically significant relationship between study quality and effect size, it is reasonable to conclude that the overall mean effect may be inflated and that our conclusions about treatment effect on formal employment should be based only on those studies rated medium quality. The

mean estimate for these studies (g=0.12) is slightly smaller than for all studies in the analysis (g=0.199) but remains significant (95% CI [0.05, 0.19]).

- Summary: Our interpretation is that there is evidence that TVET interventions are, on average, effective (relative to no intervention) at increasing the probability of having a job in the formal sector for young people in LMICs.

Monthly earnings

- The overall mean effect of TVET on earnings was positive and significant (g=0.127, 95% CI [0.043, 0.21]).

- However, significant heterogeneity was observed (Q = 25.5; df = 8; p = 0.00128; I^2 = 68.6%; tau^2 = 0.00815).

- Four variables were tested for moderating effects.

    o No statistically significant differences in mean effects were observed between studies according to study quality (Q_b = 0.204; p = 0.652), type of TVET intervention (Q_b = 0.397; p = 0.529), length of follow-up period (Q_b = 0.186; p = 0.666), or gender (Q_b = 1.26; p = 0.262).

- As there is no evidence of a statistically significant relationship between study quality and effect size, it is reasonable to conclude that the overall mean effect is not inflated and that our conclusions about treatment effect on monthly earnings should be based on all studies in the analysis (g=0.127). Although there is no evidence of a statistically significant relationship between study quality and effect size, the larger mean for the low quality studies for this outcome is at least consistent with the findings for study quality for the other outcomes.

- Summary: Our interpretation is that there is evidence that TVET interventions are, on average, effective (relative to no intervention) at increasing the monthly earnings of young people in LMICs.

Self-employment earnings

- The overall mean effect of TVET on self-employment earnings was negative and non-significant (g=-0.025, 95% CI [-0.11, 0.061]).

- No significant heterogeneity was observed (Q = 0.206; df = 1; p = 0.65; I^2 = 0%; tau^2 = 0).

- This analysis was based on two medium quality studies.

- One variable was tested for moderating effects.

    o No significant differences in mean effects were observed between studies according to gender (Q_b = 1.27; p = 0.259).

- Summary: Our interpretation is that there is only weak evidence that TVET interventions (relative to no intervention) decrease the monthly self-employment earnings of young people in LMICs. Furthermore, the observed effect was very small.

Weekly hours worked in paid employment

- The overall mean effect of TVET on number of weekly hours worked was positive, but non-significant (g=0.043, 95% CI [-0.017, 0.104]).

- No significant heterogeneity was observed (Q = 1.8; df = 5; p = 0.876; I^2 = 0%; tau^2 = 0).

- Four variables were tested for moderating effects.

  - No significant differences in mean effects were observed between studies according to study quality (Q_b = 1.41; p = 0.234), type of TVET intervention (Q_b = 0.0677; p = 0.795), or length of follow-up period (Q_b = 0.109; p = 0.741).

  - Statistically significant differences in mean effects were observed between studies according to gender (Q_b = 10.1; p = 0.00151). Treatment effects for female youth were positive, g=0.16 (95% CI [0.04, 0.28]), while those for male youth were negative, g=-0.09 (95% CI [-0.2, 0.01]). When the analysis was re-run with only those studies rated medium, the differences remained significant.

- As there is no evidence of a statistically significant relationship between study quality and effect size, it is reasonable to conclude that the overall mean effect is not inflated and that our conclusions about treatment effect on weekly hours should be based on all studies in the analysis (g=0.043).

- Summary: Our interpretation is that there is only weak evidence that TVET interventions are, on average, effective at increasing the number of weekly hours worked by young people in LMICs. Furthermore, the observed effect was very small. The evidence suggests that female youth may benefit more than male youth.

6.2 LIMITATIONS

A key strength of this study is its application of systematic review principles to improve upon prior work. However, there are several limitations to the review that should be acknowledged. First, we could only calculate effect sizes for 10 of the 26 included studies. Thus, the meta-analysis of quantitative results includes less than half of the studies included in the review. If the studies for which it was possible to compute effect sizes are systematically different from those for which it was not, the

pooled effect estimated may not be the effect of all the studies included in the review. Also, the exclusion of half of the studies may have affected the power of the meta-analysis, limiting the possibility of detecting significant programme effects. Second, although, as previously noted, we attempted to minimise publication bias by conducting an extensive literature search, it is possible that we did not identify all eligible studies. Unfortunately, the small number of studies for which effect size calculation was possible hampers any meaningful quantitative publication bias analysis, such as the Egger test, that would have enriched the discussion on the existence of publication bias in studies assessing TVET interventions. Third, the methods for calculating comparable effect sizes from studies using complex econometrics methods, as used in this review, are under-developed and require further research (for a complete discussion, see Becker & Wu, 2007; Duvendack, Hombrados, Palmer-Jones, & Waddington, 2012; Lipsey & Wilson, 2001). Finally, the meta-analyses in this review synthesised effect sizes from a wide range of methodological designs including experimental and quasi-experimental designs. Some of the methodological concerns associated with the lower quality of some quasi-experimental studies, such as those using propensity score matching, may mean the studies have yielded biased estimates of treatment effect. All conclusions from the current review are therefore sensitive to the possibility that the results from the meta-analysis may be over- or under-estimating the effects of TVET interventions on employment outcomes.

6.3 CONCLUSIONS

There is increasing international interest in TVET as a means of advancing sustainable development and addressing economic and social challenges. This review was undertaken to support donors, foundations, and other policy-level organisations and departments who are concerned about, and trying to take action and develop policy to improve, labour market outcomes for youth in LMICs. Unfortunately, the evidence base on which to base conclusions about the relative efficacy of TVET interventions has a number of limitations. First, a key finding of this review is the overall scarcity of robust evidence, as indicated by the relatively few studies that met the inclusion criteria. It would seem that only a very small proportion of the many TVET interventions currently in operation in developing countries around the world have been rigorously evaluated. In addition, the scarcity of well-executed RCTs, in particular, means that the body of evidence is not as robust as we would like when trying to answer a question about effectiveness. Second, the included studies cannot be generalised to the population of programmes in existence. For example, no eligible studies of apprenticeships for young people in LMICs were located in the search process, and the majority of programmes were set in Central/South America. Third, due to the lack of quantitative information reported in many of the included studies, the meta-analysis includes data from only 10 of the 26 studies that were reviewed. Finally, the observed heterogeneity of effect sizes was often statistically significant, indicating that different studies point to

somewhat different conclusions.

TVET interventions included in the synthesis were found to demonstrate an overall positive effect on paid employment, formal employment, monthly earnings, and weekly hours worked. In contrast, the overall effect on self-employment earnings was found to be negative. The heterogeneous nature of the distributions warranted further exploration, from which two important points can be made. Firstly, attempts to explain the observed heterogeneity suggest that methods matter. The low quality studies have consistently larger mean effects than the medium quality studies. For paid employment, and formal employment, statistically significant differences in mean effects were observed between studies according to study quality, suggesting that the overall mean is inflated and that the treatment effects should be based on the medium quality studies only. Secondly, effects are generally small and difficult to detect. The mean effects for paid employment (medium quality studies only), self-employment earnings, and working hours are negligible, and statistically insignificant. In contrast, the mean effects for formal employment (medium quality studies only) and monthly earnings are larger (though still relatively small) and statistically significant. Overall, the existing evidence shows that TVET interventions have some promise, with the strength of the evidence being stronger for formal employment and monthly earnings than for the other outcomes measured.

With one exception, moderator analysis found no significant relationships between variables tested and effect size. For weekly hours, statistically significant differences in mean effects were observed between studies according to gender. Average treatment effects for female trainees were positive, while those for male trainees were negative, suggesting that, at least in terms of increasing the number of hours worked, TVET works better for young women and young men. The within-group mean effect for male youth, however, was not statistically significant. Different types of TVET produced similar effects on the various outcomes that were measured in the primary studies. Outcomes measured at different time-points after training ended also produced statistically similar effects. It would be premature to conclude, however, that there are not in fact real differences between different types of TVET intervention or that treatment effects do not diminish over time. Due to the relatively small number of studies in the meta-analysis and the heterogeneity between studies, we may not have had adequate statistical power to detect moderating effects of the variables tested in this review, especially since some categories only contained two or three studies. There may be other moderating variables which could account for the differences in effects between studies that we were unable to test in this review due to lack of data, such as participants' socio-economic status, duration of treatment, whether the intervention was theoretically informed, implementation fidelity, and so on.

The studies included in this systematic review represent the best empirical evidence currently available for the impact of TVET on youth employment outcomes. As the review improves upon prior work by statistically synthesising TVET intervention

research, its findings strengthen the evidence base on which current policies and practices can draw. That being said, interpreting the evidence and drawing out the implications for policy and practice is nonetheless challenging. Although this review provides some evidence of the causal impact of TVET on certain labour market outcomes, several limitations of both the included studies and the review itself mean that drawing strong inferences from the results of the analyses is not recommended, and caution should be used when applying the findings of the review.

6.3.1 Implications for policy and practice

Overall, the findings from this review provide evidence that young people in LMICs gain some benefit from TVET interventions. Statistically, the effect size may be small, or even negligible, but even a small increase in the rate of paid employment can translate into thousands, if not tens or hundreds of thousands, of young people entering the labour market, where the programme is delivered on a large scale. Notwithstanding that the crucial issue will always be whether the net employment rate has increased, it is both important and worthwhile to continue to invest in TVET provision for youth in developing countries and encourage access. It would be premature, however, to recommend for, or against, the use of any of the interventions included in this review. Interventions that were classified as multi-component (i.e., two-phase sequential programmes combining classroom-based training with on-the-job practical experience) produced mean effects that were substantially similar to simpler interventions comprised of a single mode of TVET (e.g., technical education). Although multi-component programmes may be able to target several barriers to employment, such as low levels of vocational skills and lack of work experience, thereby increasing their likelihood of success, the potential benefits of more complex interventions may suffer where there are difficulties with implementation. In the absence, therefore, of evidence in support of one particular (possibly expensive) intervention, opting for the cheapest and/or most culturally acceptable models may be the best approach.

As the effects observed in this review are generally small and were difficult to detect, it is therefore of some importance that future programmes are evaluated rigorously and that policymakers think carefully about how to improve programmes to create larger effects on the outcomes. To build the evidence base further, many more of the TVET interventions currently in existence in developing countries need to be rigorously evaluated, and the results reported and disseminated efficiently. It is important that policymakers, donors, and other relevant stakeholders coordinate efforts to identify why such research is not being conducted and/or is not being disseminated in a way that can inform others. They should also lend their support, financial and otherwise, to the systematic design, implementation, and evaluation of the full range of TVET models in operation in LMICs, in accordance with clear hypotheses of change that make explicit assumptions about causal links, implementation, and contextual and external factors. Assessment of the evidence should be undertaken for each key hypothesis. Doing so is likely to aid

understanding of how an intervention is making a difference (i.e., though which channels or mechanisms), how feasible it is to extrapolate it to other settings, and ultimately help improvement of TVET interventions for the benefit of young people in developing countries.

6.3.2 Implications for research

Given the relatively small number of studies that met the criteria for inclusion in this review, there is a clear need for additional research in this area. The methodological inconsistencies and weaknesses of the current evidence base, and specific knowledge gaps, suggest a number of future research priorities. Acting on these will require the various stakeholders engaged in TVET research taking a critical look at the barriers affecting research production and dissemination.

6.3.2.1 Evaluate other types of TVET

Two-phase vocational training interventions, consisting of a period of school-based instruction followed by on-the-job training where trainees gain work experience, were found to be over-represented by the studies included in this analysis. This therefore limits the scope of the synthesis and has implications for the conclusions of the review. Additional studies examining technical education, vocational education, and apprenticeships are needed, in order to allow an examination of the effectiveness of all the different types of TVET interventions in existence.

6.3.2.2 Adequately describe interventions

Most study authors did not adequately describe the intervention. Clear information about the specifics of interventions is necessary to aid attempts by reviewers at explaining any heterogeneity in effects that are observed. In addition, attempts to replicate the intervention will be hampered if adequate descriptions of intervention characteristics are not provided. It is recommended that authors report information for each of the intervention components on the following aspects: duration; frequency of attendance; curriculum; setting; class size; teaching methods; education/credentials of personnel delivering training; trainee accreditation; incentives offered; cost; and funding. It is also recommended that authors provide full details of the above, and also clearly state their involvement in the development and/or implementation of the intervention.

6.3.2.3 Test effects of different intervention components

Many of the interventions were comprised of multiple components. A few offered potential trainees a choice between different types of training (e.g., classroom-based vocational training *or* on-the-job internships) within the bounds of a single intervention. A very small number of included studies utilised an additional treatment group that received only part of the intervention (e.g., in the situation where some trainees had participated in only the first phase of a two-phase

intervention). The vast majority, however, did not attempt to evaluate each of the components of the intervention. It is recommended that, wherever possible, authors evaluate each component separately.

6.3.2.4 Consistently measure and report study outcomes

To facilitate more meaningful comparisons across studies, as well as to allow for better transparency, it is recommended that future research defines, measures, and reports outcomes clearly and consistently. For example, authors should specify the minimum number of hours that someone had to work to meet the definition of 'in employment,' distinguishing between part-time and full-time employment as appropriate. Studies need to look at other studies when they design their own, building on existing evidence in a more systematic way than at present.

6.3.2.5 Measure, report, and analyse all relevant variables

Deciding which TVET programmes to implement with limited resources requires an understanding not only of which interventions are effective, but for whom they are effective. Although many of the included studies provided some information about gender differences in impact, relatively few explored how the impact of TVET interventions on young women and men might then vary according to other populations characteristics, such as age, socio-economic status, and location. In addition, there were some key variables that were not measured and/or reported, such as whether the intervention was theoretically underpinned. If future reviews are to fully account for differences in effects between studies, and help us understand the different factors/mechanisms that contribute to the success or failure of one intervention over another, for different groups of young people, then this data should be collected and reported.

6.3.2.6 Measure long-term outcomes

It is important to measure long-term employment outcomes. Very few studies of the included studies measured outcomes over the course of a second follow-up period. It is recommended that studies follow up participants over several years to examine whether, and at what magnitude, TVET interventions can sustain employment effects over time.

6.3.2.7 Measure key intermediate outcomes

An additional gap in the literature relates to the tracking of intermediate outcomes. These are variables that could be directly affected by the intervention and that are a first step towards achieving the final outcomes. In the context of TVET interventions, these intermediate outcomes might include skills, contacts in the labour market, and motivation to search for a job. Understanding the impact of the intervention on these intermediate outcomes can aid understanding of how the programme is making a difference. It is recommended that future evaluations of

TVET measure intermediate outcomes, according to the hypotheses in the theory of change underpinning the intervention. This will require future evaluations to consider more fully the overall logic of the intervention.

6.3.2.8 Measure net employment

A key question concerning TVET interventions is whether job creation is additional or not. The total absence of any studies in the review measuring net employment outcomes signals a missing aspect of analysis that is ideally needed in primary studies. Future research could explore whether this is possible to do with sufficiently powered cluster designs.

6.3.2.9 Report data needed to calculate effect size

We also found in this review that overall the reporting of the quantitative results of the included studies was very poor. Sixteen of the 26 studies that met the eligibility criteria for inclusion in this review did not provide adequate data to calculate effect sizes, and therefore could not be included in the meta-analyses. The basic statistical information necessary to compute comparable effect sizes that were most commonly not reported included (i) the standard deviation (pooled, treatment or control) for the outcome variable, (ii) sample sizes, and (iii) frequencies/probabilities of an event occurring in each of the groups (as opposed to the difference between the groups). It is recommended that authors provide this information for all outcomes measured, regardless of whether the results were statistically significant, or the results of other analyses were presented. Better reporting of statistical results in impact evaluations of TVET programmes are needed to make meta-analysis and synthesis studies more meaningful.

6.3.2.10 Report information needed for assessments of risk of bias and replication of study findings

Whilst the experimental technique of randomisation is considered the gold standard of evaluation techniques, not all programmes can be randomised. This is particularly the case in the field of economics. The use of matching techniques has a long and established tradition in the TVET literature. The majority of the studies included in this review used propensity score matching, with smaller numbers using an experimental design or regression techniques. Each study included in the review was subject to thorough assessment for potential sources of bias. However, we found that study reports often lacked important details that would allow us to confidently judge the appropriateness of reported analyses. Many studies were rated low overall, not because they were judged as having a high risk of bias, but because their overall risk of bias was unclear. Because researchers can use different analytical approaches (especially when using quasi-experimental techniques) it is important that methodological choices made in the process are clearly described in published work.

For randomised controlled trials, it is recommended that authors diligently adhere

to the checklist items of the CONSORT (CONsolidated Standards of Reporting Trials) Statement, which is used is used worldwide to facilitate clarity, completeness, and transparency of reporting of RCTs.[26] So, for instance, baseline demographic data for each arm in a trial should be reported.

For matching-based studies, it is recommended that authors report full information about methodological choices regarding matching procedures/method by which the matched pairs were formed, methods for comparing the distribution of the covariates between the treated and untreated subjects, and methods for estimating treatment effects after matching on the propensity score. At a minimum, details should be provided on (a) the calliper widths used in the matching process, (b) all the variables included in the matching equation, (c) whether matching used data collected at baseline or endline, (d) the number of subjects that failed to match, the number of observations in the control group matched with observations in the treatment group (where matching with replacement was used), (e) the balance in measured variables between treated and untreated subjects in the matched sample, and (f) the results of Rosenbaum's sensitivity test.

For regression-based studies, it is recommended that authors report full details of the method of adjustment, all variables used in the regression analysis, Hausman's test, and specification tests (e.g., multicollinearity test).

Poor reporting practices have other important consequences in that researchers who try to replicate the findings of a published article may not be able to do so because critical information about analytical choices is missing.

6.3.2.11 Evaluate the application of quasi-experimental techniques

The publication of results based on propensity score methods has increased dramatically in recent years. The validity of estimations of effects based on propensity score methods lies among other conditions on the assumption that the study accounts for all the unobservable characteristics associated with participating in the programme; which is a non-testable condition. Thus, further studies using rigorous experimental designs to assess the effectiveness of TVET programmes are required to inform the design of these programmes. Since the use of experimental evaluations is not always feasible, it is also important that empirical methods for measuring, and ultimately reducing, the bias incurred by the use of quasi-experimental evaluation methods continue to be developed; see, for example, recent work by Costa Dias, Ichimura, and van den Berg (2012).

---

[26] http://www.consort-statement.org/consort-statement/

7 Acknowledgements

8 References

8.1 INCLUDED STUDIES

Acero, C., Alvarado, M., Bravo, D., Contreras, D., & Ruiz-Tagle, J. (2011). *Evaluación de impacto del Programa Jóvenes al Bicentenario para la cohorte de participantes en el Año 2008*. Retrieved from: http://www.pnud.cl/proyectos/Documentos%20de%20proyectos/2011/Infor me%20Evaluaci%C3%B3n%20Impacto%20JB_FINAL.pdf

Aedo, C., & Nuñez, S. (2004). *The impact of training policies in Latin America and the Caribbean: the case of Programa Joven (Research Network Working Paper #R-483)*. Washington, DC: Inter-American Development Bank.

Aedo, C., & Pizarro, M. (2004). *Rentabilidad económica del programa de capacitación laboral de jóvenes 'Chile Joven'*. Santiago de Chile: INACAP and Mideplan.

Alzuá, M. L., & Brassiolo, P. (2006). *The impact of training policies in Argentina: an evaluation of Proyecto Joven* (Working Paper: OVE/WP-15/06). Washington, DC: Inter-American Development Bank.

Alzúa, M. L., Nahirñak, P., & Alvarez de Toledo, B. (2007, July). *Documento de Investigación: Año 27- Edición Nro. 47: Evaluation of Entra 21 using quantitative and qualitative data*. Paper presented at the Conference Q2: Combining Qualitative and Quantitative Approaches in Poverty Analysis, Hanoi.

Analítica Consultores. (2006). *Estudio de evaluacion de la estrategia Bécate*. Tlalpan: Secretaría del Trabajo y Previsión Social.

Attanasio, O., Kugler, A., & Meghir, C. (2011). Subsidizing vocational training for disadvantaged youth in Colombia: evidence from a randomized trial. *American Economic Journal: Applied Economics*, *3*(3), 188-220. doi: 10.1257/app.3.3.188

Benus, J., Rude, J., & Patrabansh, S. (2001). *Impact of the Emergency Demobilization and Reintegration Project* (Prepared for: U.S. Department of

Labor Bureau of International Affairs Office of Foreign Relations). Retrieved from: http://pdf.usaid.gov/pdf_docs/Pcaab161.pdf

Bidani, B., Goh, C., & O'Leary, C. J. (2002). *Has training helped employ Xiagang in China? A tale from two cities* (Report for the World Bank's Poverty Reduction and Economic Unit East Asia and Pacific Region). Washington, DC: World Bank.

Card, D., Ibarraran, P., Regalia, F., Rosas-Shady, D., & Soares, Y. (2011). The labor market impacts of youth training in the Dominican Republic. *Journal of Labor Economics*, *29*, 267-300. doi: 0734-306X/2011/2902-0003

Chong, A., & Galdo, J. (2006). *Does the quality of training programs matter? Evidence from bidding processes data* (IZA Discussion Paper No. 2202). Bonn: Institute for the Study of Labor.

Chong, A., Galdo, J., & Saavedra, J. (2008). Informality and productivity in the labor market in Peru. *Journal of Economic Policy Reform*, *11*, 229-245. doi: 10.1080/17487870802543480

Chun, N., & Watanabe, M. (2011). *Can skill diversification improve welfare in rural areas? Evidence from the Rural Skills Development Project in Bhutan* (ADB Economics Working Paper Series No. 260). Manila: Asian Development Bank.

Delajara, M., Freije, S., & Soloaga, I. (2006). *An evaluation of training for the unemployed in Mexico* (Working Paper: OVE/WP-09/06). Washington, DC: Inter-American Development Bank.

Díaz, J. J., & Jaramillo, M. (2006). *An evaluation of the Peruvian 'youth labor training program' – Projoven* (Working Paper: OVE/WP-10/06). Washington, DC: Inter-American Development Bank.

Dmitrijeva, J. (2009). *The employment effects of public training programme: Evidence from Latvian micro and macro data*. Retrieved from: http://www.univ-orleans.fr/leo/images/espace_commun/seminaires/semmar2009/WP_57.pdf

Elías, V. J., Ruiz Núñez, F. R., 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.

Espinoza, H. (2010). *Impact evaluation of a job-training programme for disadvantaged youths: the case of Projoven*. Maastricht: Boekenplan.

Hicks, J. H., Kremer, M., Mbiti, I., & Miguel, E. (2011). *Vocational education voucher delivery and labor market returns: a randomized evaluation among*

Kenyan youth (Report for Spanish Impact Evaluation Fund (SIEF) Phase II). Washington, DC: World Bank.

Ibarraran, P., & Rosas-Shady, D. (2006). *Impact evaluation of the job training component (PROCAJOVEN) of the assistance program for the building of a training and employment system in Panama (PN0125)* (Ex-Post Project Evaluation Report: OVE/EPPER-02/06). Washington, DC: Inter-American Development Bank.

Jaramillo, M., Galdo, J., & Montalva, V. (2007). *Do the poorest among the poor benefit from active labor market programs? Evidence from Peru's Projoven*. Retrieved from: http://www.depeco.econo.unlp.edu.ar/cedlas/ien/pdfs/ meeting2007/papers/jaramillo.pdf

López-Acevedo, G. (2003). A reassessment of technical education in Mexico. *Journal of Career and Technical Education*, *19*, 59-81.

Medina, C., & Nuñez, J. (2005). *The impact of private and public job training in Colombia* (Research Network Working Paper #R-484). Washington, DC: Inter-American Development Bank.

Mensch, B. S., Grant, M. J., Sebastian, M. P., Hewett, P. C., & Huntington, D. (2004). *The effect of a livelihoods intervention in an urban slum in India: do vocational counseling and training alter the attitudes and behavior of adolescent girls?* (Working Paper No. 194). New York, NY: The Population Council, Inc.

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. Retrieved from: http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=1736050

van Gameren, E. (2010). *Evaluación de impacto del programa de apoyo al empleo*. México, DF: Secretaría de Trabajo y Previsión Social (STPS). Retrieved from: http://www.stps.gob.mx/bp/secciones/transparencia/programas_sociales/PA E2010-2011/informe%20final%20Evaluaci%C3%B3n%20de%20Impacto%20del%20 PAE.PDF

8.2 COMPANION PUBLICATIONS

Aedo, C., & Nuñez, S. (2001). *The impact of training policies in Latin American and the Caribbean: the case of Programa Joven*. Retrieved from: http://fen.uahurtado.cl/wp-content/uploads/2010/07/inv131.pdf

Attanasio, O. P., Kugler, A. D., & Meghir, C. (2008). *Training disadvantaged youth in Latin America: evidence from a randomized trial* (NBER Working Paper No. 13931). Cambridge, MA: National Bureau of Economic Research.

Attanasio, O. P., Kugler, A. D., & Meghir, C. (2009). *Subsidizing vocational training for disadvantaged youth in developing countries: evidence from a randomized trial* (IZA Discussion Paper No. 4251). Bonn: Institute for the Study of Labor.

Bidani, B., Blunch, N. H., Goh, C. C., & O'Leary, C. (2005). *Evaluating job training in two Chinese cities*. Kalamazoo, MI: Upjohn Institute for Employment Research.

Bidani, B., Blunch, N. H., Goh, C. C., & O'Leary, C. J. (2009). Evaluating job training in two Chinese cities. *Journal of Chinese Economic and Business Studies, 7*, 77-94.

Card, D., Ibarraran, P., Regalia, F., Rosas, D., & Soares, Y. (2007). *The labor market impacts of youth training in the Dominican Republic: evidence from a randomized evaluation* (Working Paper 12883). Cambridge, MA: National Bureau of Economic Research.

Chong, A., & Galdo, J. (2006). *Training quality and earnings: the effects of competition on the provision of public-sponsored training programs* (Research Department Working Paper #555). Washington, DC: Inter-American Development Bank.

de Crombrugghe, D., Espinoza, H., & Heijke, H. (2008). *Is it worthwhile to finish? job-training programmes and low completion rate: the case of Projoven-Peru*. Retrieved from: http://www.roa.unimaas.nl/seminars/pdf2009/Espinoza_Pena.pdf

de Crombrugghe, D., Espinoza, H., & Heijke, H. (2010a). *Determinants of dropout behaviour in a job training programme for disadvantaged youths* (ROA-RM-2010/8). Maastricht: Maastricht University.

de Crombrugghe, D., Espinoza, H., & Heijke, H. (2010b*). Job-training programmes with low completion rates: the case of Projoven-Peru* (ROA Research Memorandum 2010/8). Maastricht: Research Centre for Education and the Labour Market, Maastricht University.

Dmitrijeva, J. (2009). *Does unemployed training increase individual employability: evidence from Latvian microdata*. Retrieved from: http://doku.iab.de/veranstaltungen/2008/lmp2008_papers_dmitrijeva.pdf

Galdo, J., & Chong, A. (2007). *Does the quality of public-sponsored training programs matter? evidence from bidding processes data*. Retrieved from:

http://www.depeco.econo.unlp.edu.ar/cedlas/ien/pdfs/meeting2007/papers/galdo.pdf

Galdo, J., & Jaramillo, M. (2007). *Household wealth and heterogeneous impacts of market-based training programs*. Retrieved from: http://www.iza.org/conference_files/worldb2008/galdo_j2620.pdf

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). Québec: Poverty & Economic Policy Research Network.

Ibarraran, P., & Rosas, D. (2007). *Impact evaluation of a labor training program in Panama*. Retrieved from: http://www.webmeets.com/files/papers/LACEA-LAMES/2007/617/Rosas_Shady.pdf

Ibarraran, P., Rosas, D., & Soares, Y. (2006). *Impact evaluation of a youth training program in the Dominican Republic: ex-post project evaluation report of the labor training and modernization project (DR0134)* (Ex-Post Project Evaluation Report: OVE/EPPER-03/06). Washington, DC: Inter-American Development Bank.

International Youth Foundation (2009). *Final reports of the Entra21 program: phase 1: 2001-2007*. Washington, DC: Inter-American Development Bank.

López-Acevedo, G. (2001). *An alternative technical education system in Mexico: a reassessment of CONALEP* (Policy Research Working Paper 2731). Retrieved from: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2002/01/18/000094946_02010804031950/Rendered/PDF/multi0page.pdf

López-Acevedo, G. (2004). *A duration analysis of CONALEP (Mexico's national technical professional school)* (World Bank Policy Research Working Paper 3327). Retrieved from: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2004/08/02/000160016_20040802155117/Rendered/PDF/WPS3327.pdf

López-Acevedo, G. (2005). *Evaluation of national school for professional technology education in Mexico* (World Bank Policy Research Working Paper 3572). Retrieved from: https://openknowledge.worldbank.org/bitstream/handle/10986/8981/wps3572.txt?sequence=2

Medina, C., & Nuñez, J. (2001). *The impact of public and private job training in Colombia*. Retrieved from: http://www.iadb.org/res/laresnetwork/files/pr201finaldraft.pdf

Ñopo, H., Robles, M., & Saavedra, J. (2002). *Una medición del impacto del Programa de Capacitación Laboral Juvenil PROJoven* (Documento de Trabajo, 36). Lima: Group for the Analysis of Development (GRADE).

Ñopo, H., & Saavedra, J. (2003). *Evaluacion del impacto de mediano plazo de Projoven: resultados de las mediciones realizadas a los seis, doce y dieciocho meses de culminado el programa*. Lima: Group for the Analysis of Development (GRADE).

Rosas-Shady, D. (2006a). *Impact evaluation of PROJoven youth labor training program in Peru* (Ex-Post Project Evaluation Report: OVE/EPPER-04/-06). Washington, DC: Inter-American Development Bank.

Rosas-Shady, D. (2006b). *Review of the youth worker training sub program in Colombia (CO0247)* (Ex-Post Project Evaluation Report: OVE/EPPER-01/06). Washington, DC: Inter-American Development Bank.

8.3 ELIGIBLE STUDIES NOT INCLUDED IN THE REVIEW

Arrospide, M., & Egger, P. (2000). *Capacitacion laboral y empleo de jóvenes en Peru: la experiencia del programa Projoven*. Retrieved from: http://www.oit.org.pe/WDMS/bib/publ/doctrab/dt_126.pdf

Burga, C. (2003). *Re-evaluation Projoven: propensity score matching y una evaluación paramétrica*. Lima: Consorcio de Investigaciones Económicas y Sociales (CIES). Retrieved from: http://cies.org.pe/files/active/0/Pb0112.pdf

Chacaltana, J., & Sulmont, D. (2004). Políticas activas en el mercado laboral peruano: El potencial de la capacitación y los servicios de empleo. In N. García, J. Chacaltana, D. Sulmont, J. Sierra, E. Sato, & M. Jaramillo (Eds.), *Políticas de empleo en el Perú* (pp.221-298). Lima: Consorcio de Investigaciones Económicas y Sociales (CIES).

Chacaltana, J., Guerrero, G., Espinoza, H., & Pain, O. (2003). *Qué funciona y qué no funciona en Projoven: Una evaluación de los procesos de capacitación y medición de impacto*. Washington, DC: Inter-American Development Bank.

8.4 OTHER REFERENCES

African Union. (2007, May). *Strategy to revitalize technical and vocational education and training (TVET) in Africa*. Presented at the Meeting of the Bureau of the Conference of Ministers of Education of the African Union (COMEDAF II+), Addis Ababa.

Angel-Urdinola, D. F., Semlali, A., & Brodmann, S. (2010). *Non-public provision of active labor market programs in Arab-Mediterranean countries: an*

inventory of youth programs (SP Discussion Paper No. 1005). Washington, DC: World Bank.

Becker, B., & Wu, M. J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science*, 22, 414-429. doi: 10.1214/07-STS243

Betcherman, G., Olivas, K., & Dar, A. (2004). *Impacts of active labour market programs: new evidence from evaluations with particular attention to developing and transition countries* (Social Protection Discussion Paper Series No. 0402). Washington, DC: World Bank.

Betcherman, G., Godfrey, M., Puerto, S., Rother, F., & Stavreska, A. (2007). *A review of interventions to support young workers: findings of the youth employment inventory* (Social Protection Discussion Paper No. 0715). Washington, DC: World Bank.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: John Wiley & Sons.

Card, D., Kluve, J., & Weber, A. (2010). Active labour market policy evaluations: a meta-analysis. *The Economic Journal*, *120*, F452-F477. doi: 10.1111/j.1468-0297.2010.02387.x

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: a review. *Sankhya: The Indian Journal of Statistics, Series A*, *35*, 417–446.

Costa Dias, M., Ichimura, H., & van den Berg, G. (2012). Treatment evaluation with selective participation and ineligibles. Retrieved from: http://eprints.ncrm.ac.uk/2973/1/JASA_Dias

D'Agostino, R. B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 17, 2265-2281.

Deeks, J. J., Altman, D. A., & Bradburn, M. J. (2001). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In M. Egger, G. Davey Smith, & D. G. Altman (Eds.), S*ystematic reviews in health care: meta-analysis in context* (pp. 285-312). London: BMJ Publishing Group.

Department for Employment and Learning, Northern Ireland. (2002). *Report of the taskforce on employability and long-term unemployment*. Belfast: DELNI.

Dehejia, R., & Wahba, S. (1999). Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94, 1053-1062.

Duval, S., & Tweedie, R. (2000). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89-98. doi: 10.1080/01621459.2000.10473905

Duvendack, M., Hombrados, J. G., Palmer-Jones, R., & Waddington, H. (2012). Assessing 'what works' in international development: meta-analysis for sophisticated dummies. *Journal of Development Effectiveness*, 4, 456-471. doi: 10.1080/19439342.2012.710642

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634. doi: http://dx.doi.org/10.1136/bmj.315.7109.629

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, *87*, 998-1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, *21*, 196-216.

Fares, J., & Puerto, O. S. (2009). *Towards comprehensive training* (Social Protection Discussion Paper No. 0924). Washington, DC: World Bank.

Glazerman, S., Levy, D. M., & Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. *The Annals of the American Academy of Political and Social Science*, *589*, 63–93. doi: 10.1177/0002716203254879

Greenberg, D. H., Michalopoulos, C., & Robins, P. K. (2003). A meta-analysis of government-sponsored training programs. *Industrial and Labor Relations Review*, *57*, 31-53. doi: 0019-7939/00/5701

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: evidence from evaluating a job training program. *Review of Economic Studies*, 64, 605-654.

Heckman, J. J., Ichimura, H., Smith, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrika*, *66*, 1017–1098.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, *65*, 261–294.

Heckman, J. J., LaLonde, R., & Smith, J. (1999). The economics and econometrics of active labor market programs. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3A) pp. 1865-2097. Amsterdam: North-Holland.

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

Higgins, J. P. T., & Green, S. (Eds.) (2011). C*ochrane handbook for systematic reviews of interventions*. Chichester: Wiley-Blackwell.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., … Cochrane Statistical Methods Group (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal*, *343*, d5928. doi: 10.1136/bmj.d5928

Hillage, J., & Pollard, E. (1998). *Employability: developing a framework for policy analysis* (Research Report RR85). London: Department for Education and Employment.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*, 199–236. doi:10.1093/pan/mpl013

International Labour Organization. (2011). *Global employment trends for youth: 2011 update*. Geneva: International Labour Organization.

International Labour Organization. (2012). *Global employment trends 2012: preventing a deeper jobs crisis*. Geneva: International Labour Organization.

Katz, E. (2008, October). *Programs promoting young women's employment: what works?* Paper presented at the World Bank Launch of the Adolescent Girls Initiative, Washington, DC.

Keef, S. P., & Roberts, L. A. (2004). The meta-analysis of partial effect sizes. *British Journal of Mathematical and Statistical Psychology*, *57*, 97-129.

King, K., & Palmer, R. (2010). *Planning for technical and vocational skills development*. Paris: UNESCO International Institute for Educational Planning.

Kluve, J., & Schmidt, C. (2002). Can training and employment subsidies combat European unemployment? *Economic Policy*, *35*, 409-448.

Kluve, J. (2006). *The effectiveness of European active labor market policy* (IZA Discussion Paper 2018). Bonn: Institute for the Study of Labor.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, *76*, 604-620.

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal,* 333, 597-600. doi: http://dx.doi.org/10.1136/bmj.333.7568.597

Lipsey, M. W., & Wilson, D. B. (2001). *Practical neta-analysis. applied social research methods series* (Vol. 49). London: Sage.

Maclean, R., & Wilson, D. N. (Eds.) (2009). *International handbook of education for the changing world of work: bridging academic and vocational learning* (Vol. 1). New York, NY: Springer.

McQuaid, R. W., & Lindsay, C. (2005). The concept of employability. *Urban Studies*, *42*, 197-219. doi: 10.1080/0042098042000316100

Michalopoulos, C., Bloom, H. S., & Hill, C. J. (2004). Can propensity-score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? *Review of Economics and Statistics*, *56*, 156–179. doi:10.1162/003465304323023732

Misola, N. K. (2010). *Improving the participation of female students in TVET programmes formerly dominated by males: the experience of selected colleges and technical schools in the Philippines* (UNESCO-UNEVOC Case Studies of TVET in Selected Countries No. 3). Bonn: UNESCO-UNEVOC.

Péano, S., Vergel de Dios, B., Atchoaréna, D., & Mendoza, U. (2008). *Investment in technical vocational education and training (TVET) in the Philippines*. Paris: International Institute for Educational Planning.

Petrosino, A., Morgan, C., Fronius, T. A., Tanner-Smith, E. E., & Boruch, R. F. (2012). Interventions in developing nations for improving primary and secondary school enrollment of children: a systematic review. *Campbell Systematic Reviews*, *19*. doi: 10.4073/csr.2012.19

Puerto, O. S. (2007). *Interventions to support young workers in Latin America and the Caribbean: regional report for the Youth Employment Inventory*. Washington, DC: World Bank.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55.

Rubin, D. B. (1973a). Matching to remove bias in observational studies. *Biometrics*, *29*, 159–183.

Rubin, D. B. (1973b). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, *29*, 185–203.

Rubin, D. B. (1976). Multivariate matching methods that are equal percent bias reducing: II. maximums on bias reduction for fixed sample sizes. *Biometrics*, *32*, 121–132.

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, *74*, 318–328.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95,* 573–585. doi: 10.1080/01621459.2000.10474233

Smith, J. A., & Todd, P. E. (2001). Reconciling conflicting evidence on the performance of propensity-score matching methods. *American Economic Review,* 91, 112-118. doi: 10.1257/aer.91.2.112

Stavreska, A. (2006). *Interventions to support young workers in South, East Asia and the Pacific: regional report for the Youth Employment Inventory.* Washington, DC: World Bank.

Stuart, E. A., & Rubin, D. B. (2007). Best practices in quasi-experimental designs: matching methods for causal inference. In J. Osborne (Ed.), *Best Practices in Quantitative Social Science* (pp. 155-176). Thousand Oaks, CA: Sage Publications.

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.

United Nations. (2012). *World youth report.* Geneva: United Nations.

United Nations. (1992). *United Nations: statistical charts and indicators on the situation of youth, 1970-1990.* New York, NY: United Nations.

United Nations Department of Economic and Social Affairs. (2011). *World economic situation and prospects 2012: global economic outlook.* New York, NY: United Nations.

United Nations Educational, Scientific and Cultural Organization. (2010a). *Reaching the marginalised. education for All Global Monitoring Report 2010.* Paris: UNESCO.

United Nations Educational, Scientific and Cultural Organization. (2010b). *Guidelines for TVET policy review* (ED/ESB/TVET/2010/02). Paris: UNESCO.

United Nations Educational, Scientific and Cultural Organization. (2012). *Youth and skills: putting education to work. Education for All Global Monitoring Report 2012.* Paris: UNESCO.

Van Adams, A. (2007). *The role of youth skills development in the transition to work: a global review* (HDNCY No. 5). Washington, DC: World Bank.

Wilson, S., Lipsey, M., Tanner-Smith, E., Huang, C., & Steinka-Fry, K. (2010). *Dropout prevention and intervention programs: effects on school completion*

and dropout among school-aged children and youth. Protocol submitted to Campbell Collaboration, August 23 2010. Retrieved from http://campbellcollaboration.org/lib/project/158/

9 Appendices

9.1 LIST OF ABBREVIATIONS

3ie - International Initiative for Impact Evaluation

ADB - Asian Development Bank

ADEC - Agencia para el Desarollo Económico de la Ciudad de Córdoba

ALMP - Active Labour Market Programme

ATT - Average Treatment Effect on the Treated

ATU - Average Treatment Effects on the Untreated

AusAID - Australian Agency for International Development

Bécate - Becas de Capacitación para el Trabajo

CEPRO - Centro de Ensino Profissionalizante Rotary

CONALEP - College of Professional Technical Education

CPD - Continuing Professional Development

CS – Cross-sectional

CVT - Centres for Vocational Training

DinD – Difference-in-Differences

EFA - Education for All

ENCOPE - Encuesta Nacional de Colocación y Permanencia del Empleo

ENOE - Encuesta Nacional de Ocupación y Empleo

EPOC - Effective Practice and Organisation of Care

EPPI - Evidence for Policy and Practice Information and Co-ordinating Centre

ERIC - Education Resources Information Center

Fundación SES - Fundación (Sustentabilidad- Educación -Solidaridad)

HIV/AIDS - Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome

ICAP - Instituciones de Capacitación

IDB – Inter-American Development Bank

IH - Instituto de Hospitalidade

ICT - Information and Communications Technology

ILAB - International Labor Affairs Bureau

ILO - International Labour Organization

IMF - International Monetary Fund

INFOTEP - Instituto Nacional de Formación Técnico Profesional

IM – Insertion Modality

ITT - Intention-to-Treat

IZA - Institute for the Study of Labor

LMIC - Low- and Middle-Income Country

NBER - National Bureau of Economic Research

NORRAG - Network for Policy Research, Review and Advice on Education and Training

MTE - Marginal Treatment Effects

NGO - Non-Governmental Organisation

OECD - Organization for Economic Co-operation and Development

OLS - Ordinary Least Squares

OTEC - Organismos Técnicos de Capacitación

PROBECAT-SINAT - Programa de Becas de Capacitación para Trabajadores Desempleados- Sistema de Capacitación para el Trabajo

PROCAJOVEN - Programa de Apoyo para un Sistema Panameño de Capacitación y Empleo

ProJoven - Programa de Capacitación Laboral Juvenil

SEAL - State Employment Agency of Latvia

SEE - Servicios Estatales de Empleo

SENA - Servicio Nacional de Aprendizaje

SMD - Standardised Mean Difference

SNE - Sistema Nacional de Empleo

TEVETA - Technical Education, and Vocational Education and Training Authority

TM – Transition Modality

TVET - Technical and Vocational Education and Training

TVVP - Technical and Vocational Vouchers Program

UN - United Nations

UNDESA - United Nations Department of Economic and Social Affairs

UNESCO - United Nations Educational, Scientific and Cultural Organization

USAID - United States Agency for International Development

9.2   WORLD BANK LIST OF ECONOMIES

	Low-income economies	Lower-middle income economies	Upper-middle income economies
Europe and Central Asia	Kyrgyz Republic, Tajikistan	Armenia, Georgia, Kosovo, Moldova, Turkmenistan, Ukraine, Uzbekistan	Albania, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Kazakhstan, Latvia, Lithuania, Macedonia, Montenegro, Romania, Russian Federation, Serbia, Turkey
South Asia	Afghanistan, Bangladesh, Nepal	Bhutan, India, Pakistan, Sri Lanka	Maldives
Middle East and North Africa		Djibouti, Egypt, Iraq, Morocco, Syrian Arab Republic, West Bank and Gaza, Yemen	Algeria, Iran, Jordan, Lebanon, Libya, Tunisia
East Asia and Pacific	Cambodia, Democratic Republic of Korea, Fiji, Myanmar	Indonesia, Kiribati, Lao, Marshall Islands, Micronesia, Mongolia, Papua New Guinea, Philippines, Samoa, Solomon Islands, Timor-Leste, Tonga, Tuvalu, Vanuatu, Vietnam	American Samoa, China, Malaysia, Palau, Thailand
Sub Saharan Africa	Benin, Burkina Faso, Burundi, Central African Republic, Chad, Comoros, Democratic Republic of Congo, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Kenya, Liberia, Madagascar, Malawi, Mali, Mozambique, Niger, Rwanda, Sierra Leone, Somalia, Tanzania, Togo, Uganda, Zimbabwe	Angola, Cameroon, Cape Verde, Republic of Congo, Côte d'Ivoire (Ivory Coast), Gabon, Ghana, Lesotho, Mauritania, Nigeria, São Tomé and Principe, Senegal, Sudan, Swaziland, Zambia	Botswana, Mauritius, Mayotte, Namibia, Seychelles, South Africa
Latin America and Caribbean	Haiti	Bolivia, El Salvador, Guatemala, Guyana, Honduras, Nicaragua, Paraguay	Antigua and Barbuda, Argentina, Belize, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, Dominican Republic, Ecuador, Grenada, Jamaica, Mexico, Panama, Peru, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Suriname, Uruguay, Venezuela

18 July 2011 revision; in effect until July 2012

9.3   DATABASES SEARCHED

General bibliographic databases:

- AEI (Australian Education Index) (Dialog)

- ASSIA (Applied Social Sciences Index and Abstracts) (CSA)

- BEI (British Education Index) (Dialog)

- Econlit (Ovid)

- ERIC (Education Resources Information Centre) (CSA)

- IBSS (International Bibliography of the Social Sciences) (CSA)

- PsycINFO (CSA)

- Social Sciences Citation Index (WoK)

- Social Services Abstracts (CSA)

- Sociological Abstracts (CSA)

Specialist bibliographic databases:

- 3ie Database of Impact Evaluations
  www.3ieimpact.org/database_of_impact_evaluations.html

- AfricaBib: Bibliography of Africana Periodical Literature Database
  www.africabib.org/africa.html

- Africal Journals OnLine (AJOL) www.ajol.info/
- Bangladesh Journals Online (BanglaJOL) www.banglajol.info/
- Bioline International www.bioline.org.br/
- East View Information Service Online Databases www.eastview.com/
- Hrcak http://hrcak.srce.hr/index.php
- IDEAS Economics and Finance database (RePEc) http://ideas.repec.org/
- Indian Citation Index (ICI) www.indiancitationindex.com/
- International Labour Organization (ILO) library http://labordoc.ilo.org/
- JOLIS library catalogue - International Monetary Fund, World Bank and
  International Finance Corporation http://jolis.worldbankimflib.org/e-
  nljolis.htm
- National Centre for Vocational Education Research: VOCEDplus
  www.voced.edu.au./

- Nepal Journals OnLine (NepJOL) www.nepjol.info/

- OpenGrey www.opengrey.eu/

- Philippines Journals OnLine (PhilJOL) www.philjol.info/philjol/index.php

- SciDev Net (Science and Development Network) www.scidev.net/en/

- Thai Research http://thesis.stks.or.th/

- VET-Bib European Centre for the development of vocational training (CEDEFOP) http://libserver.cedefop.europa.eu/F?RN=100966697
- Vietnam Journals Online (VJOL) www.vjol.info/
- Youth Employment Inventory www.youth-employment-inventory.org/

9.4 SEARCH QUERY TERMS

The following string was used to search ERIC and other bibliographic databases using the CSA platform (the KW function searches the descriptor field, title and abstract). It was adapted accordingly for the remaining databases.

#1 (KW=(TVET or "technical education" 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 education" or "vocational training" 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 "industrial training" or "apprenticeship*" or traineeship* or "day release" 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 program*" or "skill* development training" or "skill* development cours*" or "staff development program*" or "work* learning" 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 intervention*" or "workforce development program*" or "labour market program*" or "labor market program*" or "labour force development" or "labor force development" 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 "occupational home economics" or "occupation* course*" or "cooperative education" or "farmer education" or "agricultural education" or "agricultural training" or "business education" or "business training" or "entrepreneurship training" 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"))

#2 (KW=(employment or employability or unemployment or underemployment or "underemployment" or "self-employment" or wage* or income* or salaries or salary or earning* or "re-enter work" or "work reentry" or "work re-entry" or "return* to work" or "stay* at work" or "remain* in work" or "stay* in work" or "attitude* to work*" or "economic impact*" or "labor economics" or "labour economics"))

#3 (KW=(job* or work* or employ* or staff or occupation* or vocation* or career* or "lab* force " or "lab* market*") **within 5** (hour* or retention or recruitment or performance or placement* or search* or security or interview* or application* or satisfaction or dissatisfaction or qualification* or skill* or attitude* or ethic* or promotion or "self-esteem" or "self esteem" or confidence or motivation or aspiration* or mobility or transition or behavio* or knowledge or opportunit* or productivity or change* or adjustment* or ambition* or development* or abilit* or efficiency or proficien* or efficacy or competen* or qualit* or skill*))

#4 (KW=(random* or RCT* or "non-random*" or "non random*" or "quasi-random*" or "control* study" or "control* studies" or "control* design*" or "control* trial*" or "control* group*" or "comparison group*" or "treatment group*" or "equivalent group*" or "two group*" or intervention* or experiment* or "quasi-experiment*" or quasiexperiment* or counterfactual or "cohort analytic" or "case-control" or "observational study" or "observational studies" or "comparative study" or "comparative studies" or "comparative design*" or "prospective allocation" or "retrospective allocation" or "comparative analys*" or "match* group*" or "propensity score matching" or PSM or "statistical matching" or "statistical control*" or "matching on observable*" or "covariate matching" or "matching with replacement*" or "kernel matching" or "nearest neighbo* matching" or "stratified matching" or stratification or regression or "multivariate analys*" or "multi-variate analys*" or "statistical model*" or "Heckman model*" or "Probit model*" or "Tobit model*" or "ordinary least squares" or "log linear" or "linear probability model*" or "least squares estimat*" or "difference in difference*" or "difference-in-difference*" or effect* or efficacy or impact* or assess* or evaluat* or econometric* or "time series" or "instrumental variable*" or "economic research" or "co-efficient*" or coefficient* or "pretest* posttest*" or "pre test* post test*" or "post-intervention" or "pre and post" or "post test" or "post-test" or "pre-intervention" or "pre- post-test" or "pre post test" or "before and after" or "baseline" or "intention-to-treat" or "difference between group*" or estimator*))

#1 and (#2 or #3) and #4

9.5   WEBSITES/GATEWAYS SEARCHED

- African Development Bank www.afdb.org/en/
- Asian Development Bank: Education www.adb.org/Education/default.asp
- Association for the Development of Education in Africa (ADEA) www.adeanet.org
- Australian Council for Educational Research (ACER) www.acer.edu.au/
- Australian Education International (AEI) www.aei.gov.au/Aei/Default.aspx
- British Library for Development Studies (BLDS) www. blds.ids.ac.uk/
- CIARIS www.ciaris.org/
- Eldis www.eldis.org/
- European Training Foundation www.etf.europa.eu
- Google scholar http://scholar.google.co.uk/
- Institute for Fiscal Studies (IFS) www.ifs.org.uk
- Institute for the Study of Labor www.iza.org
- Institute of Development Studies (IDS) www.ids.ac.uk
- Institute of Southeast Asian Studies, Singapore (ISEAS) www.iseas.edu.sg/
- Inter-American Centre for the Knowledge and Development of Vocational Training (ILO/CINTERFOR) www.cinterfor.org.uy/public/english/region/ampro/cinterfor/index.htm
- Inter-American Development Bank www.iadb.org
- Inter-American Development Bank Office of Evaluation and Oversight www.iadb.org/en/office-of-evaluation-and-oversight
- National Bureau of Economic Research www.nber.org/papers.html
- Overseas Development Institute (ODI) www.odi.org.uk
- Poverty Action Lab www.povertyactionlab.org
- UNESCO: Asia and Pacific Regional Bureau for Education www.unescobkk.org/education/
- UNESCO-UNEVOC International Centre for Technical and Vocational Education and Training www.unevoc.unesco.org/pubs.php
- United Nations Development Programme www.beta.undp.org/undp/en/home.html
- United States Agency for International Development (USAID) www.usaid.gov/index.html
- World Bank www.worldbank.org/education

9.6   FLOW OF LITERATURE THROUGH THE REVIEW

```
┌─────────────────────────────┐
│     Database searches       │
│  Potentially relevant       │         ┌─────────────────────────────┐
│  studies identified and     │────────▶│ Excluded, with reason of    │
│  screened for retrieval:    │         │ not meeting the inclusion   │
│  8072 reports               │         │ criteria: 7925 reports      │
└─────────────────────────────┘         └─────────────────────────────┘
             │
             │                           ┌─────────────────────────────┐
             │─────────────────────────▶ │ Un-retrievable: 2 reports   │
             │                           └─────────────────────────────┘
             ▼
┌─────────────────────────────┐
│  Full texts screened for    │
│  eligibility: 145 reports   │         ┌─────────────────────────────┐
└─────────────────────────────┘         │ Excluded, with reason of    │
             │                  ────────▶│ not meeting inclusion       │
             │                           │ criteria: 136 reports       │
             ▼                           └─────────────────────────────┘
┌──────────────────────┐  ┌─────────────────────────────┐
│   Handsearching      │  │   Eligible: 9 reports       │
│  Eligible: 46 reports│  └─────────────────────────────┘
└──────────────────────┘
          \                /
           \              /
            ▼            ▼
┌─────────────────────────────┐
│  Total number of eligible   │         ┌─────────────────────────────┐
│  studies: 30 (in 55 study   │────────▶│ Studies unable to include   │
│  reports)                   │         │ in the review: 4            │
└─────────────────────────────┘         └─────────────────────────────┘
             │
             ▼
┌─────────────────────────────┐
│  Included in the review: 26 │
│  unique studies (in 51      │
│  study reports)             │
└─────────────────────────────┘
```

Study	Design*		Estimation techniques**				
	E	NE	U	R	M	CS	D-in-D
Acero et al. (2011)		✓			✓	✓	
Aedo and Nuñez (2004)		✓			✓	✓	
Aedo and Pizarro (2004)		✓			✓	✓	✓
Alzuá and Brassioli (2006)		✓			✓	✓	
Alzuá et al. (2007)		✓			✓	✓	
Analítica Consultores (2006)		✓			✓	✓	
Attanasio et al. (2011)	✓			✓		✓	
Benus et al. (2001)		✓		✓		✓	
Bidani et al. (2002)		✓		✓	✓	✓	
Card et al. (2011)	✓		✓	✓	✓		
Chong and Galdo (2006)		✓		✓	✓	✓	✓
Chong et al. (2008)		✓			✓		✓
Chun and Watanabe (2011)		✓		✓		✓	
Delajara et al. (2006)		✓		✓	✓	✓	
Díaz and Jaramillo (2006)		✓			✓	✓	✓
Dmitrijeva (2009)		✓	✓	✓	✓	✓	
Elías et al. (2004)		✓		✓	✓	✓	✓
Espinoza (2010)		✓		✓			✓
Hicks et al. (2011)	✓		✓			✓	
Ibarraran and Rosas-Shady (2006)		✓	✓	✓	✓	✓	
Jaramillo et al. (2007)		✓		✓	✓	✓	✓
López-Acevedo (2003)		✓			✓	✓	
Medina and Nuñez (2005)		✓			✓	✓	
Mensch et al. (2004)		✓	✓	✓		✓	
Ñopo et al. (2007)		✓			✓		✓
van Gameren (2010)		✓			✓	✓	✓

* E: randomised experiment; NE: non-experimental design (natural or quasi-)
** U: unadjusted comparison of means; R: regression-based adjustment; M: matching (including reweighting techniques for matching groups/ a simple semi-parametric alternative to regression-based adjustment); CS: cross-sectional; D-in-D: difference-in-difference

9.8 QUALITY ASSESSMENT OF INCLUDED STUDIES

Study / bias[27]	Selection bias	Confounding	Spillovers	Outcome reporting bias	Analysis reporting bias	Summary assessments[28]
Acero et al. (2011)	U	U	U	L	U	low/medium
Aedo and Nuñez (2004)	U	U	U	L	L	low/medium
Aedo and Pizarro (2004)	U	U	U	L	L	low/medium
Alzuá and Brassiolo (2006)	H	H	U	L	L	low
Alzuá et al. (2007)	H	H	U	L	L	low
Analítica Consultores (2006)	U	U	U	L	L	low/medium
Attanasio et al. (2011)	L	L/U[29]	U	L	L	medium
Benus et al. (2001)	H	H	U	L	U	low
Bidani et al. (2002)	H	U	U	L	U	low
Card et al. (2011)	U	L	U	L	L	medium
Chong and Galdo (2006)[30]	U	L	U	L	L	medium
Chong et al. (2008)	U	U	U	L	U	low/medium
Chun and Watanabe (2011)	H	H	U	L	U	low
Delajara et al. (2006)	H	H	U	L	U	low
Díaz and Jaramillo (2006)	U	U	U	L	U	low/medium
Dmitrijeva (2009)	H	H	U	L	L	low
Elías et al. (2004)[31]	U	U	U	L	U	low/medium
Espinoza (2010)	U	L/U	U	L	L	medium[32] low/medium[33]
Hicks et al. (2011)	L	U	U	L	L	medium
Ibarraran and Rosas-Shady (2006)	U	U	U	L	L	low/medium
Jaramillo et al. (2007)	U	U	U	L	U	low/medium
López-Acevedo (2003)	H	H	U	L	U	low
Medina and Nuñez (2005)	H	H	U	L	L	low

27 H: high risk of bias; L: low risk of bias; U: unclear risk of bias

28 <u>High quality overall</u>: judgment based on low risk of bias on all key domains (interpretation: if present, bias is unlikely to alter the results seriously); <u>medium quality overall</u>: judgment based on low or unclear risk of bias on all key domains, and number of domains rated low > number of domains rated unclear (interpretation: a risk of bias that raised some doubts about the results); <u>low/medium quality overall</u>: judgment based on low or unclear risk of bias on all key domains, and number of domains rated low < number of domains rated unclear (interpretation: bias may alter the results seriously); low quality overall: judgment based on high risk of bias on one or more key domains (interpretation: interpretation: bias is very likely to alter the results seriously)

29 Low risk of bias from confounding (women); unclear risk of bias from confounding (men)

30 Same rating regardless of method

31 Same rating regardless of method

32 For employment outcomes (low risk of bias for confounding)

33 For earnings outcomes (unclear risk of bias for confounding)

Study / bias[27]	Selection bias	Confounding	Spillovers	Outcome reporting bias	Analysis reporting bias	Summary assessments[28]
Mensch et al. (2004)	H	H	U	L	U	low
Ñopo et al. (2007)	U	U	U	L	U	low/medium
van Gameren (2010)	U	H	U	H	L	low

9.9 INCLUDED STUDIES (N=26)

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Acero et al. (2011) This paper evaluates Jóvenes al Bicentenario, a training programme operated by Chile's National Service of Training and Employment.	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Groups were initially created before training took place by the National Service of Training and Employment office. The sample of beneficiaries was on the basis of SENCE priority criteria. The comparison group was selected from non-participants who registered and were eligible for the programme (every tenth registered candidate who arrived at the national SENCE office was excluded from the programme participation). Acero et al. later redefine the comparison group through the use of matching techniques based on propensity score estimates. For the treatment group, only individuals who completed all phases of the training were considered for 'completer analysis' and matched with the control group.	Data corresponds to the first wave of the training programme. Two different data sources used: • SENCE administrative records. • Survey: sample of beneficiaries (1991) and controls (992) before and after the programme completion. All participants aged 18-29 years: average age 23. At time 1: treatment, n =1991; control group, n=992. At time 2: treatment, n =1797; control group, n=864. However, some control group members had entered some phase of the treatment programme (see Table 25). Numbers used in analysis: the most extensive analyses use those treatment participants who finished the programme (n = 291) and the control group (n = 827).	Two waves of surveys (i.e., panel data) which were collected during the internship phase of the programme and approximately 4-6 months after the courses ended. Evaluation period: 2008-2009. Matching techniques: Propensity score matching based on demographic characteristics. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Intention-to-treat (ITT) • Average treatment on the treated (ATT)	• Employment • Hours • Monthly earnings At 4-6 months Results reported for whole sample (average age 23 years); no subgroup analysis provided

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Aedo and Nuñez (2004) This paper evaluates Proyecto Joven, a training programme conducted by Argentina's Ministerio del Trabajo (Ministry of Labor). Companion publications: Aedo and Nuñez (2001) The programme Proyecto Joven has also been evaluated by: • Alzuá and Brassiolo (2004) • Elías et al. (2007)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Groups were initially created before training took place by the Ministry of Labor who conducted their own evaluation of the programme. The sample of beneficiaries was selected according to region and gender (a kind of stratified sampling). The comparison group was selected from non-participants who registered and were admitted to the programme but did not receive training (a number of variables were taken into consideration to determine the counterparts. For the treatment group, individuals who at least completed the first phase of the programme (technical knowledge) were considered. In this study, Aedo and Nuñez redefine the comparison group through the use of matching techniques based on propensity score estimates.	Data corresponds to the fifth wave of the training programme. Three different data sources used: 1. PSTOT: data for all individuals (139,732) who registered and qualified to take the training programme; 2. PSUN: sample of beneficiaries (1,670) and controls (1,670) used by the Ministry of Labor (see opposite) to evaluate the programme with information gathered at their registration; 3. PMSU: same sample of 3,340 individuals, but with information gathered at a survey conducted one year after completion of the programme. Most participants aged 16-35 years: young males and females (<21 years); adult males and females (21+ years; average age not reported). Numbers used in analysis: PSMU sample • Young females: total 709; 355 beneficiaries; 353 controls • Young males: total 1026; 513 beneficiaries; 513 controls	Two surveys (both collected by the Ministry of Labor): baseline and approximately 12 months after the courses ended. Evaluation period: 1996-1998 Matching technique(s): Propensity score matching (different specifications) • Nearest neighbours (5, 10, 20 and 30), using each of the 3 data sources (PSTOT, PSUN, PMSU) Matching based on characteristics collected at baseline. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Monthly earnings At 12 months Results reported for whole sample (average age unknown) and disaggregated by • Age • Gender Only two sub-groups (young males, young females) are relevant to this review.

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Aedo and Pizarro (2004) This paper evaluates Chile Joven, a training programme operated by Chile's *Ministerio del Trabajo y Previsión Social de Chile (Ministry of Labour and Social Security)*.	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Groups were initially created before training took place by the Ministry of Labor and Social Security who ordered their own evaluation of the programme some years earlier. The original sample of beneficiaries was selected according to region and subprogram using as simple random sampling. A number of variables were taken into consideration to determine the counterparts. In this study, Aedo and Pizarro redefine the treatment and comparison group through the use of matching techniques based on propensity score estimates. For the treatment group, only individuals who at least completed the first phase of the programme (classroom-based) and participated in the subprogramme Capacitación y Experiencia Laboral en Empresas were considered.	Data corresponds to the 1997 cohort of the training programme. Source of data used: Survey conducted by Santiago Consultores Asociados: data for a sample of 1246 individuals who registered and participated in the programme and 658 controls. Majority of participants are aged 16-24 years. Numbers used in analysis: CEL completers: total 916 (504 beneficiaries; 412 controls)	One post-test survey (collected by the Santiago Consultores Asociados in 1998) in the year following the end of the training. This survey gathered both pre-programme and post-programme information. Evaluation period: 1998 Matching techniques: Propensity score matching (different specifications) • Nearest neighbours with replacement • Kernel Epanechnikov estimation Matching based on variables related to the programme eligibility criteria. Impact estimator(s): • Cross-sectional • Differences-in-Difference Evaluation parameter(s): • Average treatment on the treated (ATT)	• Employment • Formal employment • Monthly earnings At 0-12 months. Results reported for whole sample and disaggregated by • Age • Gender Only three sub-groups (young people aged <=21 years, young males aged <=21 years, young females aged <=21 years) are relevant to this review.

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Alzuá and Brassioli (2006) This paper evaluates Proyecto Joven, a training programme conducted by Argentina's Ministerio del Trabajo (Ministry of Labor). Proyecto Joven also evaluated by: • Aedo and Nuñez (2004) • Elías et al. (2007)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Groups were initially created before training took place by the Ministry of Labor who conducted their own evaluation of the programme. The sample of beneficiaries was selected according to region and gender (a kind of stratified sampling). The comparison group was selected from non-participants who registered and were admitted to the programme but did not receive training (a number of variables were taken into consideration to determine the counterparts. For the treatment group, individuals who at least completed the first phase of the programme (technical knowledge) were considered. In this study, Alzuá and Brassioli redefine the comparison group through the use of matching techniques based on propensity score estimates.	Data corresponds to the 2nd/3rd and 5th wave of the training programme. 2nd/3rd calls: at 11 months: 3001 observations, of which 1512 (treatment) and 1489 (controls); at 29 months: 2370 observations, of which 1213 (treatment) and 1157 (controls) 5th call: 3340 observations, of which 1670 (treatment) and 1670 (controls) Average age: 2nd/3rd calls: treated 24.83 years; control 24.97 years; 71% of individuals are younger than the average age. 5th call: treated 24.67 years; 25.28 years; 68% of individuals are younger than the average age.	Study used post-intervention information collected by the Ministry of Labor. 2nd/3rd calls: surveys conducted 11 and 29 months after courses ended. 5th call: survey conducted 12 months afterwards. Evaluation period: 1994-1995 (2nd/3rd call); 1996-1997 (5th call) Matching technique(s): Propensity score matching (3 versions) • Nearest neighbor • Kernel • Stratified Matching based on characteristics collected at endline. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Formal employment • Monthly earnings 2nd/3rd calls: at 11 and 29 months 5th call: at 12 months Results reported for whole sample and disaggregated by • Gender • Cohort (2nd/3rd and 5th calls)

The Campbell Collaboration | www.campbellcollaboration.org

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Alzuá et al. (2007) This study evaluates Entra21 training projects: two in Argentina (SES and ADEC) and two in Brazil (CEPRO and IH).	Quasi-experiment (ex-post) The programme not did include any impact evaluation component in its design. Control groups were built ex-post using Household Surveys for the regions in which the courses took place. Argentina: treatment group data drawn from the EPH and entra21 database. Brazil: treatment group data drawn from the PNAD 2004 and entra21 database. No further details reported.	Argentina: four different samples corresponding to the two projects that took place there (assumed that results for the 2 projects were pooled): 2003 (2nd semester): treated, n=133; controls, n=843 2004 (2nd semester): treated, n=96; controls, n=616 2005 (1st semester): treated, n=125; controls, n=1087 2005 (2nd semester): treated, n=81; controls =, n=802 Brazil: two different samples corresponding to the 1st cohort of graduates for two of the four projects that took place there: 2004 (Cepro): treated, n=474; controls, n=542 2004(IH): treated, n=546; controls, n=921 Average age: Argentina: treated 19.36; 19.5; 20.1; 20.52 years; control 20.31; 20.39; 20.41; 20.6 years. Brazil: treated 16.41; 17.74 years; control 17.06; 19.23 years. Numbers used in analysis: Maximum number of observations as above, with loss of control individuals greatest for nearest neighbour PSM.	For each sample (four from Argentina and two from Brazil), one post-intervention measurement is used (not explicitly stated, but judged to be reasonably close to the end of training). Evaluation period: 2003-2005 (assumed) Matching technique(s): Propensity score matching (3 versions) • Nearest neighbor • Kernel • Stratified Matching based on characteristics collected at endline. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Formal employment • Monthly earnings At 0-3 months Results reported for the whole sample for each of the different cohorts participating in the Argentine and Brazilian programmes (6 in total).

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Analítica Consultores (2006) This paper evaluates the Bécate (Becas de Capacitación para el Trabajo) programme and its sub-programmes. This review is concerned with one sub-programme only, Capacitación en la Práctica Laboral. Bécate has also been evaluated by: • van Gameren (2010)	Quasi-experiment (ex-post) Groups were created after training courses were implemented from existing surveys (ENCOPE and ENOE) The sample of beneficiaries was selected independently for each sub-programme of Bécate. Selection was based on acquiring a proper distribution of demographic and socio-economic variables. Comparison group was selected based on the similarity with the treatment sample on the following variables: employment situation, location, gender, age, education, and other employment related characteristics.	Data corresponds to participants in the training programme during 2005. Two different data sources used: • ENCOPE (beneficiaries): Encuesta Nacional de Colocación y Permanencia del Empleo from a pilot group of 4,982, the sample size was restricted to 376. Bécate participants that were considered to be geographically, demographically, and socio-economically representative. Of these, 107 individuals participated in "Capacitación en la Práctica Laboral", the programme focused on young people. • ENOE (comparison) Encuesta Nacional de Ocupacion y Empleo – 930 individuals who did not participate in Bécate. Training programme targeted individuals aged 16-29. In total, 63% of the 107 participants were below the age of 25; of whom 75 were female and 32 were male.	Surveys were conducted by the National Institute of Statistics, Geography and Informatics (approximately one year after treatment). Evaluation period: 2005-2006. Survey details: 1) ENCOPE – A survey of the level of placement and tenure of employment of the beneficiary group. 2) ENOE – A national survey of occupation and employment given to a national sample. Matching techniques: Limited details reported. Results reported using the Heckman method to control for selection bias. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT) • Average treatment on untreated (ATU) • Average treatment effect (ATE)	• Employment • Monthly earnings At approximately 12 months Results of interest are those for: • Bécate participants (aged under 20 years) • Participants of the sub-programme Práctica Laboral

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Attanasio et al. (2011) This study is an evaluation of Jóvenes en Acción, a training programme in Colombia. Companion publications: Attanasio, Kugler, & Meghir (2008, 2009); Rosas-Shady (2006b)	Randomised experiment Availability of training was randomly assigned among those who chose to apply for training and who were selected by the training institutions. Randomisation process involved use of 'special information system set up to register applications into the programme'. Randomisation took place at the course level across many different courses (i.e., individuals were randomly assigned to each class at each training institution). There were therefore treatment and control subjects in each of the seven cities covered by the programme.	Study evaluated 4th cohort of the programme. Random samples were collected from the applicant lists provided by the training institutions, stratified by initial treatment offer, so that roughly half the sample is in the treatment group and half is in the control group (plus, also stratified by city and sex). Baseline: 2066 treatment; 2287 controls. In total, 3549 individuals were interviewed in the follow-up survey (81.5% of the initial sample). Descriptive statistics are provided for 3237 individuals (1769 females and 1468 males) observed at both baseline and follow-up (see Table 2). Average age at baseline: females (treatment and control combined) 21.3 years; males (treatment and control combined) 21 years. Numbers used in analysis: Females: depending on the outcome measured, the total number of observations ranged from 1202 to 1767 (47% treatment, 53% controls) Males: depending on the outcome measured, the total number of observations ranged from 791 to 1464 (47% treatment, 53% controls)	Two surveys: (i)baseline data were collected either before the beginning of the training programme or during the first week of classes (ii)follow-up interviews were carried out approximately 13-15 months after the end of training (NB: because there were concerns about attrition, telephone updates had been carried out four months after completion of the programme to verify personal information such as whether participants intended to move address etc) Evaluation period: 2005-2006 Treatment effects calculated as: 1. Weighted average of programme effects across different training courses (which gave greater weight to observations in training institutions where individuals were equally likely to be assigned to the treatment and control groups); 2. As above, but with regression-based adjustments for observable characteristics measured pre-treatment. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Intention-to-treat (ITT)	• Overall employment • Formal employment • Weekly hours • Days per month • Monthly earnings • Self-employment earnings • Formal earnings • Informal earnings • Tenure • Contract At 13-15 months Results reported for whole sample and disaggregated by • Gender

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Benus et al. (2001) This study evaluates the Emergency Demobilization and Reintegration Project (Bosnia and Herzegovina).	Quasi-experiment (ex-post) A list of enterprises and institutions that participated in the programme was drawn up. These were categorised into 3 groups and then 30% of the enterprises (in each of the 3 groups) were randomly selected. Next, a database of all the training participants who received services from the selected enterprises and institutions was created. Using this database, participants were randomly selected and grouped by municipality. The comparison sample was selected using quota sampling techniques (individuals were drawn from the municipal employment bureau's database of registered unemployed). The local research team responsible for constructing an evaluation dataset provided employment bureau staff with written details on how to select individuals from the database.	Sample size: participants 1714; non-participants 1743 (overall survey response rate was 58%) Age range 18-65+ years 18-24 years subgroup: treatment 17.4%; non-participants 17.2%	A single endline survey was used to measure the impacts of training that had been completed one to three years previously. Interviews were carried out between July and November 2000. Evaluation period: 1997-2000 Treatment effects calculated through the use of standard regression-based approaches: • OLS Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Monthly earnings* At 1-3 years Results are presented for the whole sample (outside age range) and by subgroup, including by age (<25 years; 25-40 years; 40+ years). Additional programme impacts by subgroup (e.g., by education level) are outside age range. *For the whole sample there are additional outcomes.

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Bidani et al. (2002) The study evaluates a reemployment project in China. Companion publications: Bidani, Blunch, Goh, & O'Leary (2009); Bidani, Blunch, Goh, & O'Leary (2005).	Quasi-experiment (ex-post) Beneficiaries and controls were selected after the programme had commenced from a list of laid-off workers who had received training (the treatment group) and those who had not received training (the control group) provided by Shenyang and Wuhan labour bureaus, as well as local training providers. Workers were randomly selected from the list. For each city, there was a treatment sample and a control sample.	Three types of samples were used; largest samples as follows: Shenyang: n=2165 (treatment 1099; control 1066) Wuhan: n= 1957 (treatment =832; control =1125) Average age: Shenyang (treated 36.75 years; controls 40.05 years); Wuhan (treated 36.98 years; controls 38.28 years) Numbers used in analysis (maximum number of observations for regression/matching estimates): Shenyang: n=2110 (treatment 1060; control 1050) Wuhan: n= 1729 (treatment =744; control =985)	Single post-intervention survey 22 months post-training. Survey designed and implemented by the Chinese Institute of Labor Studies and the World Bank. Evaluation period: 1998-2000. Treatment effects calculated using both standard regression-based approaches and matching. Regression technique(s): • OLS • Probit model Matching technique(s): Propensity score matching (4 specifications) • Nearest neighbour • Nearest five neighbours • Kernel regression • Local linear regression Log-odds ratio matching Matching based on characteristics collected at endline. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment (more specifically, *re-employment* of laid-off workers) • Earnings At 22 months Results presented for whole sample (outside age range) and by age (<25 years; 25-40 years; 40+ years); additional programme impacts by subgroup are outside age range.

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Card et al. (2011) This study is an evaluation of a job training programme – Juventud y Empleo - in the Dominican Republic. Companion publications: Card, Ibarraran, Regalia, Rosas, & Soares (2007).	Randomised experiment The ICAPs (training institutions) determined the eligibility of applicants who applied to receive training. Once a group of 30 eligible applicants was recruited, the ICAP submitted the list of names to the Ministry of Labor, which randomly selected 20 names to receive training, while the remaining 10 were assigned to the control group. ICAPs were allowed to reassign up to five people from the control group to the treatment group in the event that one or more of the original treatments failed to show up for training or dropped out within the first two weeks of the course. Over one third of the original control group was reassigned to treatment status (this process judged by authors to have been "essentially random").	Analysis is based on a sample of applicants for the 2nd cohort of the programme. Originally assigned: control 2,564; treatment 5,801. Of the original treatment group, 1,011 were 'no-shows', while 4,791 were recorded as having received training. To fill the places of the 'no-shows', 941 members of the original control group were re-assigned to the treatment group. Average age at baseline 22.3 years (both groups). Approximately a quarter of the sample are aged over 25 years. Although baseline information was collected on all applicants to the programme, follow-up information was only collected on a subsample (drawn by stratified sampling, using age, gender and education classes as strata, from administrative lists of the treatment and control subjects). Numbers used in analysis: Full sample: treatment 782; control 563 Subgroups (treatment and control combined): males 591; females 754; <22 years old 670; education <10 663; education 10+ 682; Santo Domingo 699; elsewhere 646; Santo Domingo & education 10+ 373; all others 972	Baseline data collected prior to random assignment. One follow-up survey approximately 10-14 months after most trainees has completed their initial course work. Evaluation period: 2004-2005 Estimation of treatment effects (*linked paper*): • unadjusted comparisons of mean differences between groups • reweighted difference (a technique for matching the groups; a simple semi-parametric alternative to a regression adjustment) Estimation of treatment effects (*main paper*): Randomised design was potentially compromised by the failure to include in the follow-up survey people who were originally assigned to receive training but did not attend or dropped out. Therefore, to calculate treatment effects, authors fit a series of regression models and parametric selection models: • OLS (no covariates) • OLS (with individual covariates and region effects) • OLS (with covariates and training institution effects) • Reweighting using a logit model • Joint models for participation in training if assigned to treatment and outcomes in follow-up survey (different specifications) Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment effects (ATE)	• Overall employment • Formal employment (i.e., a job with employer-provided health insurance) conditional on working • Months in employment since training ended • Weekly hours, conditional on working • Monthly earnings, conditional on working • Hourly rate of pay, conditional on working At 10-14 months Results reported for whole sample and disaggregated by • Gender • Age (<22 years; 22+ years) • Education • Location • Education and location

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Chong and Galdo (2006) This study evaluates the Peruvian training programme, ProJoven. Companion publications: Chong and Galdo (2006), Galdo and Chong (2007) ProJoven programme also evaluated by: • Chong et al. (2008) • Díaz and Jaramillo (2006) • Espinoza (2010) • Jaramillo et al. (2007) • Ñopo et al. (2007)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Beneficiaries and controls were initially selected ex-ante by ProJoven as part of the original evaluation design. The beneficiary sub-samples are selected from a stratified random sample of the population of participants. ProJoven constructs its official control group by selecting 3 comparable individuals per trainee, who are required to be eligible and satisfy enrolment requirements. The searches for potential controls take place in the same neighborhoods as the actual trainees. Matching of individuals is based on individual information at baseline, using a small set of characteristics (gender, age, poverty, education and employment status). In this study, a new control group using propensity score matching is formed. This new control group is a sub-sample of the official control group.	Dataset consists of panel data from the 1st, 2nd, 4th, 6th, and 8th calls of the programme. At baseline: Pooled data: whole sample, n=3467 (49.8% treated, 50.2% control) Treated: n=1,725; mean age 19.67 years; 42.7% males); Control: 1,742; mean age 19.73 years; 42.6% males) 1st call: n=599 (299 treated) 2nd call: n=627 (321 treated) 4th call: n=720 (343 treated) 6th call: n=732 (405 treated) 8th call: n=764 (421 treated) Numbers used in the analysis: Maximum number of observations: 1st call: n=585 (both groups) 2nd call: n=604 (both groups) 4th call: n=679 (both groups) 6th call: n=690 (both groups) 8th call: n=705 (both groups)	Baseline and three follow-up surveys taken 6, 12 and 18 months after the end of the programme. Evaluation period: 1996-2004 Treatment effects calculated using both matching and standard regression-based approaches. Matching technique(s): Propensity score matching • Local linear kernel Matching based on characteristics collected at baseline. Non-matching technique(s): OLS: regression-based estimator of the difference between the post-treatment earnings of treatment and control group members, holding constant the level of pre-treatment earnings and a set of control variables that includes the propensity score. Impact estimator(s): • Difference-in-Differences (matching) • Cross-sectional (non-matching) Evaluation parameter(s): • Average treatment effects (ATE) • Marginal treatment effects (MTE)	• Monthly earnings At 6, 12 and 18 months. Treatment impacts* reported for each cohort individually and for the pooled data. At each time interval, and for each of the cohorts and the pooled data, results are reported for the whole sample and disaggregated by • Gender *ATE on all the treated *ATE on those attending a high-quality course *ATE on those attending a low-quality course

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Chong et al. (2008) This study evaluates the Peruvian training programme, ProJoven. This programme also evaluated by: • Chong and Galdo (2006) • Díaz and Jaramillo (2006) • Espinoza (2010) • Jaramillo et al. (2007) • Ñopo et al. (2007)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Beneficiaries and controls were selected ex-ante by ProJoven as part of the original evaluation design. The beneficiary sub-samples are selected from a stratified random sample of the population of participants. ProJoven constructs its official control group by selecting 3 comparable individuals per trainee, who are required to be eligible and satisfy enrolment requirements. The searches for potential controls take place in the same neighborhoods as the actual trainees. Matching of individuals is based on individual information at baseline, using a small set of characteristics (gender, age, poverty, education and employment status). In this study, a new control group using propensity score matching is formed. This new control group is a sub-sample of the official control group.	Dataset consists of panel data from the 1st, 2nd, 4th, 6th and 8th calls of the programme (in one area only: Lima) 1st call: n=599 (299 treated) 2nd call: n=627 (321 treated) 4th call: n=720 (343 treated) 6th call: n=732 (405 treated) 8th call: n=764 (421 treated) Numbers used in the analysis not reported. Majority of participants are aged 16-24 years.	Baseline and three follow-up surveys taken 6, 12 and 18 months after the end of the programme. Evaluation period: 1996-2004 Matching technique(s): Propensity score matching • Local polynomial matching Matching based on characteristics collected at baseline. Impact estimator(s): • Difference-in-differences Evaluation parameter(s): • Not reported	• Overall employment • Formal employment (five proxies for 'quality of employment' are used: size of the firm; health benefits; accident insurance; social security; formal labour contracts At 6 and 12 months Results are presented disaggregated by • Gender

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Chun and Watanabe (2011) This study evaluates the Rural Skills Development Project (Bhutan).	Quasi-experiment (ex-post) Groups were created after programme ended. Trainees and non-trainees were identified through the use of survey techniques (see opposite). The survey sampling approach set out to obtain a census of all trainees between the ages of 18 and 40, covering 90% of the trainee population, and develop a control group that was characteristically similar in terms of age and gender profiles. Controls were derived from within the sub-districts covered by the project.	Total: 325 trainees and 414 controls Average age: trainees (28.24) and controls (31.63) Number used in analysis: Maximum number of observations, n=721. Number of <25 year olds in the sample not reported.	The evaluation used two survey instruments (a household survey and a village head survey; both post-test), conducted approximately 12 months after the programme ended. Evaluation period: 2007-2010. Estimation of treatment effects was through regression analysis: • OLS • Propensity score matching model (five nearest neighbours) • Treatment-effects regression (maximum likelihood estimator) Matching based on characteristics collected at endline. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Income (household) At 12 months Results presented for whole sample (outside age range) and by age (<25 years; >=25 years); additional programme impacts by subgroup are outside age range.

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Delajara et al. (2006) This study evaluates different modalities of the Mexican training programmes PROBECAT-SICAT. In 2001, PROBECAT (Programa de Becas de Capacitación para Trabajadores Desempleados) changed its name to SICAT (Sistema de Capacitación para el Trabajo. More recently, it was replaced by Becas a la Capacitación para el Trabajo (Bécate). Bécate has been evaluated by Analítica Consultores (2006) and van Gameren (2010).	Quasi-experiment (ex-post evaluation using previously collected survey data) ENCOPE is a survey that interviews a sample of PROBECAT-SICAT beneficiaries between three and six months after finishing their training. From ENCOPE, authors took as treatment observations those individuals that were unemployed at the moment of starting the program and completed the training course. From ENEU, they took as control observations those individuals that were unemployed two weeks or less at the moment the treatment group was starting the training course.	Six cohorts in total (1999-2004) For year 2000, for instance, there is labour market information for 14685 persons interviewed in ENCOPE. From ENEU, there is labour market information for 3122 and 1839 individuals that were re-interviewed 13 and 26 weeks after training, respectively. The authors use these 3122 and compare to the 14487 observations from ENCOPE that have information for their situation 13 weeks after beginning of training. They drop from the sample of treatments 198 individuals that were interviewed 12 weeks or less after beginning their training (because they do not have labour market information at 13 weeks so these cannot be compared to the controls from ENEU at 13 weeks). They compare the 1839 observations re-interviewed by ENEU 26 after their first interview, with 11384 from ENCOPE with at least 26 weeks after beginning their training. A similar exercise was done for all the years. The majority of the sample is aged 16-25 years (range 12 to 55+ years).	The study makes use of several surveys: the ENCOPE (Spanish acronym for Employment survey of PROBECAT-SICAT beneficiaries), the ENECE (Spanish acronym for National Training and Education Survey) and the ENEU (Spanish acronym for Urban Employment Survey). The ENECE is a special module introduced in the ENEU every second year from 1991 to 1999, and every year since 2001. ENCOPE (used for the treatment group) captures information of individuals at a particular point in time and asks the informant to recall information. Two methods for impact evaluation are used: • Matching-based estimator (propensity matching score with nearest neighbor controls) • Regression-based estimator (parametric method for dealing with the problem of selection on unobservable) Matching based on baseline characteristics reconstructed from data collected ex-post. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment effect (ATE) • Average treatment on the treated (ATT)	• Overall employment • Self-employment • Monthly earnings, for those who have a job • Self-employment earnings At 3 and 6 months For the different programme modalities, results are reported for each of the different cohorts (1999-2004), and by • Age (15-25 years; 26-35 years; 36+ years) • Gender • Education • Region • Year quarter

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Díaz and Jaramillo (2006) This study evaluates ProJoven, the Peruvian training programme. Companion publications: Rosas-Shady (2006a) This programme also evaluated by: • Chong and Galdo (2006) • Chong et al. (2008) • Espinoza (2010) • Jaramillo et al. (2007) • Ñopo et al. (2007)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Beneficiaries and controls were selected ex-ante by ProJoven as part of the original evaluation design. The beneficiary sub-samples are selected from a stratified random sample of the population of participants. ProJoven constructs its official control group by selecting 3 comparable individuals per trainee, who are required to be eligible and satisfy enrolment requirements. The searches for potential controls take place in the same neighborhoods as the actual trainees. Matching of individuals is based on individual information at baseline, using a small set of characteristics (gender, age, poverty, education and employment status). In this study, a new control group using propensity score matching is formed. This new control group is a sub-sample of the official control group.	Dataset consists of panel data from the 1st, 2nd, 4th, 6th, and 8th calls of the programme. Numbers of individuals in treated and untreated groups: not reported. Numbers used in analysis: not reported	Four measurements in total: surveys conducted (by the ProJoven programme operators) at baseline and 6, 12 and 18 months after both the learning and on-the-job training components of the treatment were concluded. Evaluation period: 1996-2004 Matching technique: Propensity score matching • Kernel regression Matching based on characteristics collected at baseline. Impact estimator(s): • Cross-sectional • Difference-in-differences Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Formal employment • Weekly hours • Monthly earnings • Hourly rate of pay At 6, 12 and 18 months For each call (1st, 2nd, 4th, 6th and 8th), results reported for whole sample and disaggregated by • Age (16-20 years; 21-25 years) • Gender

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Dmitrijeva (2009) This study evaluates national vocational (occupational) training programmes for the unemployed operating in Latvia.	Quasi-experiment (ex-post) Data provided by SEAL (State Employment Agency of Latvia) is used to construct the groups. The treatment group is composed of unemployed persons who have completed the training. The author withdraws from the sample those individuals who have completed more than one training programme. The comparison group consists of individuals who did not participate in the evaluated programme. The author again withdraws those who have participated in other active labour market programmes.	Treatment group: 9,773 unemployed individuals; control group: 250,792 individuals (based on 12 month measurements). The sample is further split into three sub-samples according to the year of unemployment registration with SEAL: 2003 (81,903 controls and 2,947 participants); 2004 (85,668 controls and 2,759 treated) or 2005/6 (83,221 controls and 4,040 participants). Average age not reported (around a quarter of the sample are aged < 25 years); proportion that are male: 2003 (48% controls, 39% treated); 2004 (48% controls, 26% treated); 2005/6 (49% controls, 28% treated).	The evaluation of the microeconomic impact of the programme on job seekers is performed using individual data provided by SEAL. It covers unemployed individuals (programme participants and non-participants) registered with SEAL as unemployed in the period between January 2003 and August 2006. Aggregate impact is analysed using monthly panel data (1999-2006) from 33 Latvian districts. Estimation technique: • 'Naïve' estimator: i.e., simple difference of means between the groups of treated and untreated individuals • Matching Propensity score matching methodology (nearest neighbour, 1st caliper) • Regression Probit model (using a set of covariates which include the socio-demographic characteristics used in propensity score estimation and pairing) Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Employment At 6, 9, 12, 18 or 24 months after registration with SEAL (unclear how long after training ended that data was collected on participants: assumed less than 6 months). Results are reported for whole sample and the author also assess the heterogeneity of effects across time and within various socio-demographic groups. Only the results reported for youth aged <25 years are relevant to this review. The results are sorted by year of inflow into the programme (2003, 2004, 2005/6). In the second part of the paper, the author focuses on the aggregate/ macro-economic effects of the programme.

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Elías et al. (2004) This paper evaluates Proyecto Joven, a training programme conducted by Argentina's Ministerio del Trabajo (Ministry of Labor). Proyecto Joven also evaluated by: • Aedo and Nuñez (2004) • Alzuá and Brassioli (2004)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Groups were initially created before training took place by the Ministry of Labor who conducted their own evaluation of the programme. The sample of beneficiaries was selected according to region and gender (a kind of stratified sampling). The comparison group was selected from non-participants who registered and were admitted to the programme but did not receive training (a number of variables were taken into consideration to determine the counterparts. For the treatment group, individuals who at least completed the first phase of the programme (technical knowledge) were considered. In this study, Elías and colleagues redefine this comparison group through the use of matching techniques based on propensity score estimates.	Dataset consists of panel data from the 5th call of the training programme Participants: n=1514, of whom 802 males and 712 females. Controls: n=1505, of whom 798 males and 707 females. Individuals are 35 years or younger, with mean age 22.71 (participants) and 23.18 (controls). Numbers used in analysis: Participants: n=1,514 Controls: original (n=1,505); 1 neighbour (n=1,356); 10 neighbours (n=13,560), 20 neighbours (n=27,120), 50 neighbours (n=67,800) See above for numbers for males and females. Numbers of individuals in other sub-groups not reported.	Two surveys: baseline and approximately 12 months after the courses ended. Evaluation period: 1996-1998 Treatment effects calculated using both standard regression-based approaches and matching. Non-matching technique(s): • Unconditional mean differences • Model with unobservable heterogeneity • Model with normal distribution of unobservables • Model incorporating a polynomial of the propensity score Matching technique(s): Propensity score matching (several versions) • Average nearest neighbor (1, 10, 20 and 50 neighbours) • Kernel regression (bandwidth 0.2, 0.3 and 0.4) • Local linear regression (10, 20 and 50 neighbours) Matching based on characteristics collected at baseline. Impact estimator(s): • Before and after (participants only) • Cross-sectional • Difference-in-difference Evaluation parameter(s): • Average treatment on the treated (ATT) • Average treatment effects (ATE) • Marginal treatment effects (MTE)	• Overall employment • Monthly earnings • Monthly income At 12 months Results reported for whole sample and disaggregated by • Age (<25 years; 25+years) • Gender • Region

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Espinoza (2010) This study evaluates ProJoven, the Peruvian training programme. Companion publications: de Crombrugghe, Espinoza, & Heijke (2008, 2010a, 2010b) ProJoven programme also evaluated by: • Chong and Galdo (2006) • Chong et al. (2008) • Díaz and Jaramillo (2006) • Jaramillo et al. (2007) • Ñopo et al. (2007)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Beneficiaries and controls were selected ex-ante by ProJoven as part of the original evaluation design. The beneficiary sub-samples are selected from a stratified random sample of the population of participants. ProJoven constructs its official control group by selecting 3 comparable individuals per trainee, who are required to be eligible and satisfy enrolment requirements. The searches for potential controls take place in the same neighborhoods as the actual trainees. Matching of individuals is based on individual information at baseline, using a small set of characteristics (gender, age, poverty, education and employment status). In this study, a new control group using propensity score matching is formed. This new control group is a sub-sample of the official control group. Four groups are compared: • Control group: no programme • Treatment group: completed both phases of the programme • Treatment group: completed the classroom phase but not the internship phase) • Treatment group: completed the classroom training phase but were not placed in training firms) This study assesses the effects of: • ProJoven participation • internship placement • internship completion	Dataset consists of panel data from the sixth call of the programme Control group (n=992: although only actually 488 individuals as some are repeated because study uses matching with replacement); Treatment group: completed both phases of the programme (n=578); Treatment group: completed the classroom phase but not the internship phase) (n=231); Treatment group: completed the classroom training phase but were not placed in training firms) (n=183); Average age at baseline 19 years	Four measurements in total: a baseline survey conducted two to three months prior to programme commencements and 3 follow-up surveys at intervals of 6, 12 and 18 months after the end of the internships. Evaluation period: 1999-2002 Estimation of treatment effects: A two-stage approach is taken, involving the use of propensity score matching (with replacement) to create a new control group, then a series of regression analyses: • Linear probability model - fixed-effects • Linear probability model - fixed-effects (with correction for sample selection) • Pooled linear probability model - two-stage least squares • Pooled linear probability model - two-stage least squares (with sample selection correction) • Linear probability model - random-effects • Linear probability model - random-effects (with correction for sample selection) Matching based on characteristics collected at baseline. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Formal employment • Monthly earnings At 6, 12 and 18 months Results for whole sample and disaggregated by • Gender • Working experience prior to training • Per-capita household income

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Hicks et al. (2011) This study evaluates the Technical and Vocational Vouchers Program (TVVP) operating in Kenya.	Randomised experiment Study participants (both treated and untreated) were drawn from the Kenya Life Panel Survey (KLPS). The entire KLPS sample was invited to apply for vocation education tuition voucher. Voucher winners were then randomly selected from this final pool of applicants using a computer random number generator. Among the voucher winners, a random half received vouchers that can be used only in government supported public vocational training institutes, while the other half received unrestricted vouchers that could be used in either public centers or in the private training sector. The remaining individuals served as the control group. Treatment group ('school completers') compared with control (no treatment). The majority of voucher winners who enrolled in a training institution were still in school at the time of this study. A survey of a representative subset of TVVP participants was undertaken to provide evidence of the short-term impacts (an evaluation of longer-term impacts is planned).	Total number of participants 2163 (treated subjects 1055, of whom 62.7% female; untreated subjects 1108, of whom 62.9% female) Age: treated 21.7 years; untreated 21.9 years (range 18-30 years) Numbers used in the analysis: Maximum number of observations 234 (female 142, male 92), consisting of (a) treated: 70 (female 43, male 27); (b) untreated: 164 (female 99, male 65)	One post-test measurement (not explicitly reported, but judged be within 3-6 months of the end of training as reference is made to 'short-term' impacts). Evaluation period: 2008-2010 Estimation of treatment effects: • unadjusted difference in means Impact estimator(s): • Cross sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Self-employment • Total months worked, among employed • Weekly hours, among employed • Weekly hours, among self-employed • Job search length, among employed • Earnings, among employed • Earnings (i.e., profit) from self-employment At 3-6 months Results reported for the whole sample and disaggregated by: • Gender

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Ibarraran and Rosas-Shady (2006) This study evaluates the training programme, PROCAJOVEN, based in Panama (two components were evaluated separately: the insertion modality and transition modality). Companion publications: Ibarraran and Rosas (2007)	Natural experiment (ex-post) The groups were created by the study authors after the training took place, from lists provided by the executing unit of the programme. The control groups were selected among the eligible participants that never received the treatment due to budgetary and administrative issues that affected the execution of PROCAJOVEN. The evaluation data was collected for a sub-sample of the beneficiaries (treatment) and the non-beneficiaries (control). These sub-samples were drawn by stratified sampling (using age, gender, and education classes as strata).	The total sub-sample used in the analysis includes 295 controls (insertion modality: 186; transition modality: 109) and 471 treatments (insertion modality: 199; transition modality: 272). Average age: treatment (23.1 years); control (23.1 years)	One post-test survey was undertaken. Treatment group members were asked to provide monthly information on their activities, starting from the month they completed (or left) their classroom-training programme. Because of variation in (i) the date of entry into the programme and (ii) the duration of classroom training, the number of months of post-classroom training data available for members of the treatment group ranges from 9 to 20 months. Information about survey of control group members not reported, but implicit it took place at the same time. Evaluation period: 2003-2005 Estimation of treatment effects: • unadjusted comparisons of mean differences between groups • reweighted difference (a technique for matching the groups; a simple semi-parametric alternative to a regression adjustment) Further analyses reported as being conducted, but results not reported: • regression adjusted comparison • probit model controlling by a propensity score Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Intention-to-treat (ITT)	• Overall employment • Formal employment (job with social protection) • Hours worked, conditional on working • Earnings, conditional on working • Hourly rate of pay, conditional on working At 9-20 months For each programme modality, results are reported for the whole sample (average age 23 years) and disaggregated by • Gender • Age (18-24 years; 25-33 years) • Education (secondary; more than secondary) • Region (Panama; other provinces) • Gender and region (males/Panama; females/Panama; males/other provinces; females/other provinces)

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Jaramillo et al. (2007) This study evaluates the Peruvian training programme, ProJoven. Companion publications: Galdo and Jaramillo (2007); Galdo, Jaramillo, & Montalva (2008). ProJoven programme also evaluated by: • Chong and Galdo (2006) • Chong et al. (2008) • Díaz and Jaramillo (2006) • Espinoza (2010) • Ñopo et al. (2007)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Beneficiaries and controls were selected ex-ante by ProJoven as part of the original evaluation design. The beneficiary sub-samples are selected from a stratified random sample of the population of participants. ProJoven constructs its official control group by selecting 3 comparable individuals per trainee, who are required to be eligible and satisfy enrolment requirements. The searches for potential controls take place in the same neighborhoods as the actual trainees. Matching of individuals is based on individual information at baseline, using a small set of characteristics (gender, age, poverty, education and employment status). In this study, a new control group using propensity score matching is formed. This new control group is a sub-sample of the official control group.	Dataset consists of panel data from the 1st, 2nd, 4th, 6th and 8th call of the programme (in one area only: Lima) Total number of treated individuals (pooled) 1602; total number of controls (pooled) 1660; average age 20 years Numbers used in analysis: Men: (pooled data/both treated and untreated subjects): 6 months: earnings (n=1294); employment (n=1294) 12 months: earnings (n=1294); employment (n=1294) 18 months: earnings (n=978); employment (n=978) Females: (pooled data both treated and untreated subjects): 6 months: earnings (n=1738); employment (n=1750) 12 months: earnings (n=1750); employment (n=1750) 18 months: earnings (n=1319); employment (n=1356)	Four measurements in total: surveys conducted (by the ProJoven programme operators) at baseline and 6, 12 and 18 months after the sample class had completed its internships. Evaluation period: 1996-2004 Treatment effects calculated using both standard regression-based approaches and matching. Non-matching technique(s): • OLS Matching technique(s): Propensity score matching • Kernel Impact estimator(s): • Cross-sectional • Difference-in-differences Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Monthly earnings At 6, 12 and 18 months Results are presented for the whole sample (pooled data, males & females combined) and disaggregated by • Gender • Income/earnings quartile • Type of training institution • Occupation 18 months after the programme

Study	Design	Sample	Data collection and analysis	Outcomes of interest
López-Acevedo (2003) This study evaluates the labour market performance of graduates of CONALEP, Mexico's National School for Technology. Companion publications: López-Acevedo (2001, 2004, 2005)	Quasi-experiment (ex-post) Groups created after training took place using CONALEP graduate tracer surveys and two national surveys from which non-participants were selected. Representative samples of participants and non-participants were drawn and then propensity score matching techniques used.	Treatment: CONALEP 1994: 1399 (unclear how many completed interviews) CONALEP 1998: 5574 (of who 84.5% completed interviews) Control: ENE98: approximately 200,000 ENECE99: 164,550 (sample size for 1999) Unclear how many remained in the analysis after matching. Age: Participants (all cohorts pooled) 22.1 years Controls (all workers) 17-65 years Controls (lower secondary education complete and 3 years experience) 18-19 years Controls (upper-secondary education complete and 1-5 years of experience) 22-26 years	This study utilises the following surveys: Treatment: • CONALEP graduate tracer surveys (CONALEP 1994; CONALEP 1998) – based on random samples of former students who graduated between 1991 and 1997 Control: • National Employment Survey (ENE98) • National Employment, Schooling and Training Survey (ENECE99) Both these surveys are representative at the national level. Evaluation period: approx. 1988-1998 Matching technique(s): Propensity Score Matching • Nearest neighbours (3 and 5) Matching based on characteristics collected at endline Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Overall employment by sector (including self-employment) • Hours worked • Earnings • Hourly rate of pay • Further training at work At 1 to 5 years Results presented for each of the different cohorts and for the pooled sample.

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Medina and Nuñez (2005) This paper analyses public and private job training programmes in Colombia: specifically • training provided by SENA (Servicio National de Aprendizaje), an agency of the Columbian government • training provided by public institutions (including SENA)* • training provided by private sector institutions • training taken more than a year earlier, irrespective of the provider** Companion publication: Medina and Nuñez (2001) *Data not used in the meta-analyses, due to the overlapping samples.* ***Not considered in this review.*	Natural experiment This study utilises general survey data (see opposite). The survey asked individuals whether they received training in the year before the interview. The treatment group is comprised of those who responded that they had ever participated in any training course. Those who responded that they had received no training were the control group used. The authors created sub-groups of participants based on the source of training (see information in left-hand column). Each of these four groups is compared against the control group.	The treatment group was made up of two samples: sample 1 (training only last year); sample 2 (training last year or ever before) These 2 treatment samples and the control group were divided into four demographic groups: adult male, adult female, male youth and female youth. Youth are aged 12 to 22 years and adults aged 22 to 55 years. This review focuses on male and female youth from sample 1 only. Numbers used in analysis: Male youth: Control: n=2907 Treatment (sample 1): SENA, n-39; public, n=46; private, n=58; ever, n=191 Treatment (sample 2): SENA, n-23; public, n=44; private, n=83; ever, n=203 Female youth: Control: n=2823 Treatment (sample 1): SENA, n-60; public, n=73; private, n=88; ever, n=243 Treatment (sample 2): SENA, n-34; public, n=67; private, n=141; ever, n=290	This study utilises the government designed survey Encuesta Nacional de Calidad de Vida (ECV). The survey, which was conducted in 1997, is random, based on the 1993 population census and comprises 10,000 households in 75 municipalities. It is reported as being representative at the national level, at the level of urban-rural regions, and at the level of some regions. Evaluation period: mainly 1996-1997 Matching technique(s): Propensity score matching (several versions) • Nearest neighbor (1, 5 and 10 neighbours) • Kernel regression (bandwidth 0.1 and 0.2) • Local linear regression (bandwidth 0.05 and 0.1) Matching based on characteristics collected at endline. Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Monthly earnings At 12 months Outcomes reported for: • Male youth • Female youth • Adult males (outside age range) • Adult females (outside age range)

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Mensch et al. (2004) This study examined the effects of a livelihoods intervention for adolescent girls in slum areas of India.	Quasi-experiment Individuals residing in the intervention area slums were compared with individuals of the same age residing in control area slums. Two comparable wards were purposively selected and one was randomly assigned as the experimental site and the other to control. Five slum areas within the experimental ward were randomly selected as project sites and nine slum areas within the control ward were selected for comparison purposes. Authors had no control over which adolescents residing in the experimental area participated in the intervention. Girls in the experimental ward self-selected into the programme Groups created ex-post, after both surveys had taken place (see opposite).	3199 interviewed at baseline; 6148 interviewed at endline (of whom 1887 had been interviewed at baseline) However, whilst all adolescents who were living in the slum areas were specified for inclusion in the surveys (only small number of whom actually participated in the intervention – see below). Single-sex (female) participants only Numbers used in analysis: Analysis restricted to the intervention participants and respondents for the control areas who were interviewed in both survey rounds: beneficiaries 122; non-beneficiaries 381 (following matching, number of controls is reduced to 117)	Two surveys: one at baseline and another approximately 22 months later (approximately 6-9 months after the end of training). Evaluation period: 2001-2003 Calculation of treatment effects combined matching with standard regression-based approaches. First, propensity score matching techniques (nearest neighbor, without replacement) were used to create more comparable samples of beneficiaries and non-beneficiaries. Matching based on characteristics collected at baseline. Second, to estimate the effects of participating in the programme: • unadjusted comparison of mean differences • a series of regression models was estimated for each outcome variable (depending on the type of outcome, the statistical model varied; for the time-use variables, a Tobit model was employed) Impact estimator(s): • Cross-sectional Evaluation parameter(s): • Average treatment on the treated (ATT)	• Weekly hours (spent in labour market work: defined as paid/unpaid employment, vocational training) At 6-9 months Results presented for the whole sample.

Study	Design	Sample	Data collection and analysis	Outcomes of interest
Ñopo et al. (2007) This study evaluates ProJoven, the Peruvian training programme. Companion publications: Ñopo, Robles, & Saavedra (2002); Ñopo and Saavedra (2003) ProJoven programme also evaluated by: • Chong and Galdo (2006) • Chong et al. (2008) • Díaz and Jaramillo (2006) • Espinoza (2010) • Jaramillo et al. (2007)	Quasi-experiment (ex-post re-evaluation of previously collected evaluation data) Beneficiaries and controls were selected ex-ante by ProJoven as part of the original evaluation design. The beneficiary sub-samples are selected from a stratified random sample of the population of participants. ProJoven constructs its official control group by selecting 3 comparable individuals per trainee, who are required to be eligible and satisfy enrolment requirements. The searches for potential controls take place in the same neighborhoods as the actual trainees. Matching of individuals is based on individual information at baseline, using a small set of characteristics (gender, age, poverty, education and employment status). In this study, Ñopo and colleagues redefine this comparison group through the use of matching techniques. This new control group is a sub-sample of the official control group.	Dataset consists of panel data from the sixth class of the programme Pre-matched samples: beneficiaries (n=1014), controls (n=1534) Average age not reported, but individuals were initially sampled according to age: 16-20 years and 21-25 years (NB: other variables were included in the stratified sampling). Numbers used in analysis: not reported	Four measurements in total: surveys conducted at baseline and 6, 12 and 18 months after the sample class had completed its internships. Evaluation period: 1999- 2002 Matching technique(s): A two-step matching procedure was used (firstly, based on propensity scores, and secondly, based on gender and average hourly wages). • Nearest neighbour (0.05 caliper) Matching based on characteristics collected at baseline. Impact estimator(s): • Difference-in-differences Evaluation parameter(s): • Average treatment on the treated (ATT)	• Overall employment • Hours worked • Monthly earnings • Hourly rate of pay • Occupational segregation, by gender At 6, 12 and 18 months Results reported for whole sample and disaggregated by: • Gender

Study	Design	Sample	Data collection and analysis	Outcomes of interest
van Gameren (2010) This paper evaluates several programmes of the Employment Support Program (PAE) associated with the National Employment Services of Mexico. This review is only concerned with the evaluation of the Capacitación en la Práctica Laboral sub-component of the Becas de Capacitación para el Trabajo (Bécate) programme. Bécate has also been evaluated by: • Analítica Consultores (2006)	Quasi-experiment (ex-post) Groups were created after the end of the programme. The sample of beneficiaries was selected according to demographic and socio-economic variables and stratified by region. The comparison group was selected through propensity score matching of unemployed counterparts who took part in the ENOE survey.	Data corresponds to three time points for three cohorts of participants and comparison groups (a total of 9 data points). The Practica Laboral sub-programme of Bécate targeted individuals aged 16-29. Average age of participants in the matched sample: First semester 2008: participants 22.73 years; comparison group 22.44 years; First semester 2009: participants 21.55 years; comparison group 21.53 years. Numbers used in analysis for 3 samples: 2008 (first quarter): - beneficiaries: 2,138 - comparison: 1,825 2008 (second quarter): - beneficiaries: 2,010 - comparison: 1,711 2009 (first quarter): - beneficiaries: 2,530 - comparison: 1,536	Two different data sources (surveys) used. ENCOPE (for treatment group): Encuesta Nacional de Colocación y Permanencia del Empleo - data for a sample of individuals who participated in training programs conducted by the National Employment Services. ENOE (for control group): Encuesta Nacional de Ocupacion y Empleo- this survey drops and recruits 20% of participants each iteration. As a result, only a portion of total respondents had 3 time points available. ENCOPE – A survey of the level of placement and tenure of employment of the beneficiary group. This survey was administered once and it gathered data on 3 time points retrospectively. That is, it asked individuals to think back to the time directly after the treatment, 13 weeks after the treatment, and 26 weeks after the treatment. ENOE – A national survey of occupation and employment given to a national sample. This survey was administered 3 different times (at 'baseline', and then after 13 weeks and 26 weeks). Authors assume that it is unlikely that very many people from the ENCOPE survey took part in the ENOE survey Matching techniques: Propensity score matching (nearest neighbor, without caliper). Matching based on characteristics collected after treatment. Impact estimator(s): • Cross-sectional • Differences-in-difference Evaluation parameter(s): • Not reported	• Overall employment • Formal employment (job with work benefits) • Formal employment (job with health benefits) • Monthly earnings Outcomes immediately after treatment, 13 weeks after the training course ended, and then at 26 weeks. Results reported for whole sample at each time point and disaggregated by: • Gender

9.10 INCLUDED INTERVENTIONS (N=20)

Name, setting, dates	Funding, implementation	Main features	Other features
1. Entra21 *Country, coverage and dates of operation:* • Entra21 operates in 18 different countries in Latin America and the Caribbean • Projects focused on for this review: Fundación SES (Sustentabilidad- Educación - Solidaridad) which provided training in 5 regions of Argentina; Agencia para el Desarollo Económico de la Ciudad de Córdoba (ADEC) which trained students from the area of Rio Segundo in Argentina; Centro de Ensino Profissionalizante Rotary (CEPRO), a Brazilian project in the city of Cotia; Instituto de Hospitalidade (IH) which operates in northeast Brazil. • 2001 - ongoing	*Funded by*: • Multilateral Investment Fund, part of IDB • International Youth Foundation (IYF) • USAID • Other international donors (including Microsoft and Nokia) *Developed and implemented by:* IYF	*Type of TVET programme:* programme comprised of two sequential phases involving classroom-based training followed by practical on-the-job internships. In addition, each NGO committed to insert at least 40% of their graduates in the labour market. *Target groups:* Young people (aged 16-24) living in a poor household, who face difficulties in terms of both finding a job and remaining employed over a long term period. *Training delivered by:* for entra21 as a whole, 35 private sector NGOs across 18 countries executed the programme; each NGO was entirely responsible for eligibility criteria, obtaining the internships, defining the course contents and duration of courses and internships, etc *Duration:* entra21 as a whole: an average of 430 hours of training (the shortest course lasting 210 hours and the longest 730) and an average of 217 hours of on-the-job training (range 80 to 412 hours) *Frequency*: no details	*Other features:* The training is concentrated in information and communications technology (ICT). ICT-related activities ranged from simple computing operations to more sophisticated courses, such as webpage design. In addition, all courses had to provide training in life skills (e.g., communication skills) and job seeking skills. Job placement services were provided. Training was oriented to meet labour market needs. All the NGOs selected had experience either in training and/or in working with vulnerable populations, but for most of them the idea of supplying specific short-term training to meet labour market needs was new. Thus, they had to develop creative ways to read labour demand and to place beneficiaries in internships. The internship component was a requirement. Most of the projects provide some sort of stipend.

Name, setting, dates	Funding, implementation	Main features	Other features
2. Proyecto Joven *Country, coverage and dates of operation:* • Argentina (distribution of the training activities at the national level was determined in accordance with regional populations) • 1993/4 – 2001 (information not consistent between reports)	*Funders*: • Inter-American Development Bank • Ministry of Labor *Developed and implemented by:* Ministry of Labor (in each province a local office was established to manage the programme; however, the main decisions about the programme are taken at federal level)	*Type of TVET programme:* sequential two-phase programme involving classroom-based (technical) vocational training followed by practical on-the-job internships Target group(s): young people aged 16-29 years living in a poor household, who had attained less than secondary education, had little work experience, and were under/unemployed or otherwise out of the labour force *Delivered by*: Instituciones de Capacitación (ICAP) - selected through an international bidding process. The ICAPs were responsible for obtaining firms willing to accept interns. The criteria for selecting firms for internships were: general characteristics of the firm; tasks to be performed by the trainees; personnel involved in similar positions in the firm; equipment; and supplies and infrastructure. *Duration:* varied from 14 to 20 weeks (6-12 weeks for the first phase of the programme; 8 weeks for the internship) *Frequency:* on average, 200 hours of training was provided (implied that daily, but number of hours not reported); no details for the internship phase	*Other features*: During the first training phase, participants were taught knowledge and technical skills for a particular occupation. Training also focused on the development of work habits and sought to strengthen the skills that would be required by potential employees on a day-to-day basis: for example, reading and writing, mathematical operations, problem resolution, logical reasoning, team work, interaction with peers and superiors, understanding instructions, etc. Courses were demand-driven. Most of the training offered in the fifth round of the programme was concentrated in the following sectors: agriculture, industry, services, and construction. No course had more than 20 participants. The programme provided transportation expenses, medical checkups, books, training materials, clothes, and a subsidy for mothers with young children. During the internship phase, the beneficiaries applied what they learned in the first phase, acquired experience in a real work environment, and engaged in the main tasks of the occupation for which they have been trained. Not reported if interns were paid a wage.

Name, setting, dates	Funding, implementation	Main features	Other features
3. Emergency Demobilization and Reintegration Project (EDRP) *Country, coverage and dates of operation:* • 2 political entities within Bosnia and Herzegovina: the Federation of Bosnia and Herzegovina (FBH), and the Republika Srpska (RS); each managed and implemented its own project resources independently • national programme implemented in each of the 10 cantons of FBH and 6 regions of RS • 1996 – 1999	*Funded by*: • World Bank (International Development Association) • USAID (Support for Eastern Democracy funds) • Dutch Trust Fund *Developed and implemented by*: the US International Labor Affairs Bureau (ILAB) provided technical assistance in several areas, including development of institutional management structures, counseling services and data systems)	*Type of TVET programme:* multi-component project combining education and training, counseling services, labour market information database and management assistance. In the implementation of the project, the education and training component was combined with the counseling services component and these form the focus of the impact evaluation. The vast majority of training services provided under the EDRP training programmes were provided on-the-job (i.e., at the workplace), with the remaining training taking place at educational institutions. In addition, training providers were committed to inserting 60-80% of graduates into the labour market. *Target groups:* Primary: demobilised soldiers Secondary: included refugees, war victims, disabled persons, widows, general unemployed *Training delivered by:* the vast majority of contracts were granted to enterprises that agreed to provide on-the-job training and to hire 80% of the training participants; a smaller number of contracts were granted to (i) educational institutions that agreed to provide training and to provide employment for 60% of the training participants, and (ii) employment counseling providers. *Duration*: no details *Frequency*: no details	*Other features*: demand-driven education and training services (including small business training). No further details reported.

Name, setting, dates	Funding, implementation	Main features	Other features
4. **Rural Skills Development Project** *Country, coverage and dates of operation:* • Bhutan (all subdisticts across the three rural districts of Haa, Trashigan and Bumthang) • 2007 - 2010	*Funders***:** • Royal Government of Bhutan • Asian Development Bank *Developed and implemented by:* as above	*Type of TVET programme:* sequential classroom-based (theoretical) instruction/training plus off-the-job practical demonstration (i.e., at the training institution). *Target group(s):* rural residents (all ages) living in poverty *Delivered by*: not reported *Duration:* the first two stages took place over 3 months; stage 3 over a period of approximately 4–6 months *Frequency:* seven hours each day for all stages	*Other features*: training was provided in three stages. Stage 1 was comprised of theory lessons where participants reviewed basic concepts related to the skills. Stage 2 involved practical demonstration where participants were introduced to building and construction of a toilet structure. Stage 3 provided on-the-job training where participants were involved in school hostel construction. NB: only stages 1 and 2 were evaluated The programme provided training in four basic construction skills types: carpentry, masonry, plumbing, and house wiring. In addition, training in hair-dressing was also offered. Implicit that programme was demand-driven. Stipend compensation for the first two stages was approximately $4 per day, while stage 3 participants received about $6 per day.

Name, setting, dates	Funding, implementation	Main features	Other features
5. Chile Joven *Country, coverage and dates of operation:* • Chile • 1991 –2002	*Funded by:* • Government of Chile • Inter-American Development Bank *Implemented by:* Government of Chile (Ministry of Labor and Social Security) And FOSOS (Fund for Solidarity and Social Investment))	*Type of TVET programme:* two-phase programme involving classroom-based training followed by practical on-the-job internships *Target groups:* young people (16 – 24) of low income status, who were unemployed, underemployed, inactive, or looking for a job for the first time and who are not pursuing any formal educational except in evening classes. *Training delivered by:* private training institutions (*Organismos Técnicos de Capacitación* [OTECs]), selected through a competitive bidding process. In order to participate, OTECs had to be properly incorporated into the records of SENCE (the National Training and Employment Service). Depending on the sub-programme, the proposals of these agencies had to include "commitments of participation" from companies that committed them to receive the students participating in the program *Duration:* about 200 hours training and 3-6 months internship full time. *Frequency:* not reported	*Other features:* courses were demand-driven (private institutions were required to identify labour market opportunities). Courses were free for young beneficiaries, who also received a transportation and food subsidy during the theoretical and practical phases of the course, unless the latter involved a contract. Trainees also received personal accident insurance. The training institutions were responsible for the placement of the trainees. To guarantee that the content of the course was effectively demanded by the private sector, the course proposals submitted for this sub-programme must be accompanied by letters of intent or some other equivalent written commitments with businesses or business associations that cover at least 80 percent of the course enrolment.

Name, setting, dates	Funding, implementation	Main features	Other features
6. Jóvenes al Bicentenario *Country, coverage and dates of operation:* • Chile • 2008 –	*Funded by:* Government of Chile *Developed and implemented by:* Government of Chile (National Training Fund)	*Type of TVET programme:* two-phase programme involving classroom-based training followed by practical on-the-job internships *Target groups:* young people (18 – 29) who achieved less than higher education; priority was given to those residing in vulnerable communities (as of the Vulnerability Index by Ministry of Interior), heads of household, unemployed and low educational attainment. *Training delivered by:* government accredited training institutions, Organismos Técnicos de Capacitación (OTEC). *Duration:* 52% of training courses lasted less than 250 hour, 46% were between 250 hours, and 2% were between 400 and 773 hours. 360 hours of internship. *Frequency:* no details	*Other features:* programme beneficiaries received transportation subsidy, food subsidy, and one of every six participants received job search assistance. Courses were demand-driven and participants could choose the training content most aligned with their interests

Name, setting, dates	Funding, implementation	Main features	Other features
7. Retraining programmes for laid-off workers *Country, coverage and dates of operation:* • China • piloted in 30 municipalities in 1994 and expanded to 200 cities by 1996 (training in the cities of Shenyang and Wuhan is evaluated)	*Funded by*: • municipal governments in the cities of Shenyang and Wuhan • state-owned enterprises *Developed and implemented by:* • municipal governments (e.g., the municipal labour bureaus); • Wuhan: the responsibility for re-training of laid-off workers rested mostly with the enterprises (they could approach the industry bureau and labour bureau for assistance if they faced financial or technical difficulties)	*Type of TVET programme:* retraining programmes for laid-off workers (predominantly two-phase programmes involving classroom-based instruction combined with practical training; however, a number of training institutions provided only theoretical instruction without any practical training in their courses) *Target group(s):* unemployed former employees of Chinese state-owned enterprises in the cities of Shenyang and Wuhan *Delivered by*: Shenyang: the city's re-emplo.ent training centre administered the programme, which was executed by training organisations under the district labour bureaus. Training institutions could be disqualified if they did not meet the performance standards set. Wuhan: training was conducted by the labour bureau training organisations (such as the city employment training centre and district employment training centres); other organisations that satisfied the qualifications requirements also undertook this training, for which they were compensated to cover part of their expenses. *Duration*: Shenyang: one-month (132 hours of study) Wuhan: one to six months (typically, 2-3 months of full-time study); in 1998, the average number of course hours was 255, of which 55% were practical *Frequency*: unclear (full-time for at least some of the participants)	*Other features*: Shenyang: training courses with a minimum duration of one month were eligible for the government subsidy of 100 yuan per trainee. Training was free of charge. The quality of training varied widely across training institutions. Training institutions differed greatly in capacity, space, classroom set-up, workshop facilities and laboratory and mechanical equipment. There were no minimal standards governing the content of curricula and the qualifications of instructors. Wuhan: only courses of 2-3 months were eligible for the government subsidy. Trainees were charged part of the training costs (they were expected to purchase textbooks and practice materials). Predominantly demand-driven, however some of the courses did not provide skills demanded in the local labour market. Trainees selected from a list of available courses. In Shenyang, these included computer training courses, cooking courses, beauty courses, massage and hair cutting courses, sewing courses and toy making courses. In Wuhan, these included computer training course, management courses, repair training courses, and driver's education courses.

Name, setting, dates	Funding, implementation	Main features	Other features
8. Jóvenes en Acción *Country, coverage and dates of operation:* • Colombia; offered in seven largest cities of the country: Barranquilla, Bogota, Bucaramanga, Cali, Cartagena, Manizales and Medellin. • first cohort received training in 2002 and the last one in 2005 This was one of three social programs introduced by the Columbian government to help those hardest hit by the 1998 recession.	*Funded by:* • World Bank • Inter-American Development Bank *Developed and implemented by:* Colombian government	*Type of TVET programme:* sequential two-phase programme involving classroom-based training followed by practical on-the-job internships *Target groups:* Young people between the ages of 18 and 25, living in urban areas in the two lowest socio-economic strata of the population *Training delivered by*: classroom training was provided by private and public training institutions, which had to participate in a bidding process to be able to participate in the program. The training institutions were selected based on the following criteria: legal registration, economic solvency, quality of teaching, and ability to place trainees after the classroom phase into internships with registered employers. On-the-job training was provided by legally registered companies. *Duration:* 6 months (3 months classroom training, followed by 3 months on-the-job training) *Frequency:* average number of hours of classroom training was 7.56 per day; internships offered an average of 5.19 daily hours of on-the-job training	*Other features:* Training courses provided vocational skills in a diverse number of occupations. Both manual and administrative courses were on offer. The private training institutions played a fundamental role in determining what courses were offered, how they were marketed and how they were designed Training institutions were paid according to market prices and were paid conditional on completion of training by the participants of the programme. Average class had 27 students. The programme provided cash transfers to cover transportation and lunch, as well as childcare expenses for women with children less than 7 years of age. Internships were unpaid. The companies participating in the programme operated in manufacturing (textiles, food and beverages, pharmaceuticals, and electricity), retail and trade, and services (including security, transportation, restaurants, health, childcare, and recreation).

Name, setting, dates	Funding, implementation	Main features	Other features
9. Servicio Nacional de Aprendizaje (SENA) job training programme *Country, coverage and dates of operation:* • Colombia (national; there are 20 regional offices in the main cities) • 1957 – ongoing SENA is an agency of the Colombian government. It aims to promote efficiency in the labour market through a variety of services. These services are provided through programmes for firms and workers, and include advice to employers, job search workshops and job training. This study evaluates the job training programme (i.e., one aspect of SENA provision) which offers short and long courses to train people for work.	*Funded by*: • Colombian government (NB: SENA's revenue comes from mandatory levies on employers and provision of technological services to firms; employers must pay 2% of their payroll to SENA) *Developed and implemented by:* Colombian government	*Type of TVET programme:* vocational training (assumed classroom-based) Target group(s): although in theory the entire population is eligible for the courses, people on the lowest incomes are targeted; thus, applicants are asked to provide information that enables assessment of their socioeconomic status *Delivered by*: instructors in Centres of Vocational Training (CVT) *Duration:* both short and long courses Female youth: average hours of training = 533 Male youth: average hours of training= 506 *Frequency:* not reported	*Other features*: SENA's courses make up a large share of the job training provided by Colombia's public sector. There were a total of 111 CVT centres by 1997, providing training in four areas: agriculture, industry, commerce and services, and other sectors. Implicit that programme is demand-driven. The job training programme provides a range of different length courses. SENA also offers counselling services to help individuals determine which courses best serve their interests. About 60% of trained public sector workers attend SENA courses. Although the main purpose of the SENA programme is to prepare people for the labour force, many people enrol to improve their job prospects. Training is delivered to both the disadvantaged (who may not be employed) and individuals in employment who are seeking to update their skills. Most programmes offered by SENA require candidates to have at least nine years of schooling.

Name, setting, dates	Funding, implementation	Main features	Other features
10. Public sector vocational training (including training provided by SENA) *Country, coverage and dates of operation:* • Colombia (national) • not reported	*Funded by*: not reported *Developed and implemented by:* not reported	*Type of TVET programme:* vocational training (assumed classroom-based) Target group(s): not reported *Delivered by*: public vocational training institutions, other than those provided by SENA (no further details) *Duration:* unclear Female youth: average hours of training = 333 Male youth: average hours of training = 463 *Frequency:* not reported	*Other features*: Implicit that programme is supply-driven.
11. Private sector vocational training *Country, coverage and dates of operation:* • Colombia (national; though concentrated in the main cities) • not reported	*Funded by*: not reported *Developed and implemented by:* not reported	*Type of TVET programme:* vocational training (assumed classroom-based) Target group(s): unclear (reported that different to those for public courses, because of differences in the subjects taught, the education level of the courses, their national coverage, and so on) *Delivered by*: private vocational training institutions (no further details) *Duration:* unclear Female youth: average hours of training = 152 Male youth: average hours of training = 186 *Frequency:* not reported	*Other features*: private institutions offer more specialised courses than public organisations; unlike public agencies, such as SENA (see above), they do not offer free basic skills courses. They tend to be concentrated in the main cities. Implicit that programme is supply-driven.

Name, setting, dates	Funding, implementation	Main features	Other features
12. Juventud y Empleo *Country, coverage and dates of operation:* • Dominican Republic • 1999 – ongoing	*Funded by:* • Inter-American Development Bank *Developed and implemented by:* Government of the Dominican Republic (Ministry of Labor)	*Type of TVET programme:* sequential two-phase programme involving classroom-based training followed by practical on-the-job internships *Target groups:* young people (16 – 29) of low income status, who are both unemployed and who achieved less than a secondary education (and who are not enrolled in regular schooling); a special emphasis was placed on enrolling women *Training delivered by:* private training institutions (instituciones de capacitacion [ICAPs]), selected through a competitive bidding process. All potential training providers were required to present training proposals for the courses they would offer. The proposals were evaluated and revised by the National Institute of Technical and Professional Training (Instituto Nacional de Formacion Tecnica Profesional [INFOTEP]). INFOTEP was also contracted to inspect the selected ICAPs before any training took place and during the training courses. Much less frequently, ICAP personnel also visited some of the firms that were providing internships. *Duration: c*ourses with a maximum duration of 350 hours and two-month internships *Frequency:* no details	*Other features:* the training courses were split into two parts: basic skills training and technical/vocational training. Basic skills training was intended to strengthen trainees' self-esteem and work habits, while vocational training was customised to the needs of local employers (i.e., demand-driven). Trainees were not paid during the classroom component of the program, but they did receive partial reimbursement for their transportation costs and meals. The stipend was well below the typical level of earnings for members of the control group who were working in the follow-up survey. The programme also provided trainees with insurance against workplace accidents.

Name, setting, dates	Funding, implementation	Main features	Other features
13. Livelihoods Intervention *Country, coverage and dates of operation:* - India (city of Allahabad) - Pilot project starting in 2001 and lasting approximately one year	*Funded by***:** - Population Council *Developed and implemented by:* CARE India, one of the largest NGOs working in the country	*Type of TVET programme:* multi-component programme offering participants reproductive health training sessions, vocational counselling & vocational training, savings formation information, and follow-up support from a peer educator. *Target group(s):* adolescent (14-19 year old) female slum residents in the Indian city of Allahabad *Training delivered by:* "..literate girls who had their parents' permission were identified and trained to be peer educators. The peer educators attended a six-day reproductive health training course and a two-day peer-education training course to help them become more effective communicators and facilitators for discussions about vocational training opportunities and savings account formation. Emphasis was placed on providing information about reproductive health and livelihoods and on communication and group-formation skills. Although the vocational counseling and savings components were integrated with the reproductive health lessons, the vocational training sessions were held after the completion of the reproductive health classes and were open only to those participants who had maintained good attendance" (Mensch et al., 2004: 6). No details reported about personnel delivering the government-run courses. *Duration:* unclear *Frequency:* groups met once a week in the home of the peer educator; frequency of government-run training courses attended by participants is not reported	*Other features*: The vocational training courses offered was based on the number of girls interested (i.e., the programme followed a supply-driven approach). Courses arranged by the project included tailoring, mehndi, creative painting, dhari, mending and embroidery, candle making, silver ornament and link making, pot decoration, crochet, jute doll making, basic cooking, personal grooming, and fabric painting. The project set a limit of five courses per girl. Almost all of the intervention participants attended a class in either mehndi or creative painting, held in the home of the peer educator. Assumed that these classes involved practical training. Subsequent training sessions for other courses were held at a central training center that required a brief commute. The project also made arrangements for older girls (18 years and older) to attend government-run courses. Participants contributed a small amount toward the purchase of the raw materials required for their courses. After the training, some materials were required for the production of handicrafts at home. Initially the project provided some of this capital investment (for example, handlooms for dhari weaving classes and materials used for sewing).

Name, setting, dates	Funding, implementation	Main features	Other features
14. Technical and Vocational Vouchers Program (TVVP) *Country, coverage and dates of operation:* • Kenya (target area: Busia, in Western Kenya) • 2008 - ongoing	*Funded by***:** • World Bank (through the Bank-Netherlands Partnership Program) *Developed and implemented by:*	*Type of TVET programme:* technical and vocational education financed directly through tuition vouchers which are provided through the programme (assumed that the training is predominantly comprised of classroom-based theoretical instruction) *Target group(s):* young people (18-30 years old) not in education *Delivered by: public and private institutions (t*he programme targeted all the major government village polytechnics and technical training institutes in the study area, as well as a large cross-section of available private institutions) *Duration*: unclear (the majority of voucher winners chose courses that lasted 2 years or more, while roughly 20% chose courses that lasted at most one year *Frequency*: not reported	*Other features*: programme provides vocational education tuition vouchers, which cover the tuition costs for most private vocational education programmes and government-run rural village polytechnics or technical training institutes. Of the voucher winners, a random half were awarded a voucher that could only be used in public (government) institutions, while the other half received a voucher that could be used in either private or public institutions. Beneficiaries were then provided with the opportunity to apply to the vocational education institution of their choosing. A wide variety of courses and institution-types were available to choose from. The programme is designed to be open to students who have already received some vocational training but want to further their skills. Programme entails a supply-driven approach.

Name, setting, dates	Funding, implementation	Main features	Other features
15. Occupational training programmes for the unemployed *Country, coverage and dates of operation:* • Latvia • Early 1990s-ongoing	*Funded by***:** • Latvian government *Developed and implemented by:* Ministry of Welfare through the State Employment Agency of Latvia (SEAL)	*Type of TVET programme:* vocational training The design of the programme allows either obtaining a new profession (vocational training and re-qualification involves 75% of participants in occupational training) or upgrading skills in a current occupation (raising of qualifications involves 25% of participants). *Target group(s):* job seekers *Delivered by:* not reported *Duration: average* 4-6 months *Frequency:* not reported	*Other features*: Educational programmes are selected by SEAL according to the demand in the labour market (inquired through employer surveys).

Name, setting, dates	Funding, implementation	Main features	Other features
16. Becas de Capacitación para el Trabajo (Bécate) and the sub-component, Capacitación en la Práctica Laboral *Country, coverage and dates of operation:* • Mexico • 2004 – ongoing Bécate replaced PROBECAT-SICAT in 2004. First developed in 1984, PROBECAT (Programa de Becas de Capacitación para Trabajadores Desempleados) changed its name in 2001 to SICAT (Sistema de Capacitación para el Trabajo). Capacitación en la Práctica Laboral - a sub-component of Bécate - is targeted specifically at youth.	*Funded by:* Inter-American Development Bank (since 1996). *Developed and implemented by:* Government of Mexico (National Employment Services) - Bécate is a programme of the PAE, the Program for Employment Assistance.	*Type of TVET programme:* on-the-job internships *Target groups:* young people (16 – 29) who are unemployed or underemployed and who lack work experience. *Training delivered by:* a monitor/instructor (representative of Bécate) working conjointly with a local business to provide training opportunities. During the (on-the-job) training process, beneficiaries received personalised advice from an instructor in charge of arranging training activities with businesses. The instructor supported the advising of groups of up to 25 recipients in at least 5 companies, to promote the placement of beneficiaries and to monitor the training process. *Duration:* short-term job placements (1-3 months) *Frequency:* 30-48 hours of work weekly (5-8 hours daily)	*Other features:* Wages were at least one to three times the regional minimum wage, depending on the nature of the work. Support for transportation costs was provided (up to 350 Mexican Pesos per month). In addition, accident insurance was provided for the duration of the training course.

Name, setting, dates	Funding, implementation	Main features	Other features
17. College of Professional Technical Education (CONALEP) system *Location and period of operation:* • Mexico (by 1986, all 31 states in Mexico had CONALEP schools; although the distribution of students by state is uneven, with large numbers attending schools within the metropolitan zone of Mexico City) • 1978 - ongoing	*Funded by:* Mexican government *Developed and implemented by:* Mexican government	*Type of TVET programme:* technical education (assumed that this is predominantly comprised of classroom-based theoretical instruction) *Target group(s):* young people at the upper-secondary school level, and from a lower socio-economic status *Delivered by:* teachers in CONALEP schools (implicit) *Duration:* not reported *Frequency:* daily (implicit)	*Features:* CONALEP was created as a public decentralised body of the Ministry of Public Education. In its first year of operation (1979) it offered training in seven careers (five of which focused on manufacturing, and the remaining two on medical assistant and nursing professions). Since 1983, it has also offered courses for industry. The number of careers has expanded rapidly (146 in the early 1990s, but dropping to 29 in 1997). This growth has coincided with a shift towards white-collar occupations in commerce, administration, computing, and accounting. In 1991-1992, a modular programme was introduced and in 1994 a competency-based model for nine careers (the latter was introduced as part of the Education Modernisation Project financed by the World Bank). CONALEP offers the opportunity for students to gain access to higher education as they can opt to take more courses per semester and to take a separate high school diploma exam.

Name, setting, dates	Funding, implementation	Main features	Other features
18. PROBECAT-SICAT First developed in 1984, PROBECAT (Programa de Becas de Capacitación para Trabajadores Desempleados) changed its name in 2001 to SICAT (Sistema de Capacitación para el Trabajo). More recently, SICAT was replaced by Bécate (Becas a la Capacitación para el Trabajo). *Country, coverage and dates of operation:* • Mexico (national: about 4.74 million workers have been trained between 1984 and 2005) • 1984 - 2004 (ongoing in the form of Bécate)	*Funded by:* Mexican Government (for example, for the so-called 'mixta', the SNE (Sistema Nacional de Empleo) paid for the workers scholarships, the SEE (Servicios Estatales de Empleo) paid for the operative costs and the firm financed the training itself: see below) *Developed and implemented by:* The SNE is the main institution in charge of organising and implementing the programme, with the aid of the regional offices of SEE.	*Type of TVET programme:* (i) school-based vocational training (escolarizada); once training over, workers would look for a job using the placement services available at the SNE and SEE); (ii) on-the-job training (the so-called 'mixta'); training was done at the plant or workshop; those workers not hired by the firm would look for a job using the SNE placement services; (iii) training for the self-employed (capacitacion para el autoempleo); not specified whether this is classroom-based. The 'escolarizada' sub-programme was terminated in 2001. Since 2002, the 'mixta' and 'autoempleo' types have dominated the training activities, accounting for about 60 % and 30 % of the trainees, respectively. *Target groups:* individuals characterised by low levels of schooling, low wages, high unemployment, low share of qualified labour, and high level of informality in the labour market. *Training delivered by:* unclear *Duration:* (i) school-based training involved three months of attending classes; (ii) length of on-the-job 'mixta' training not reported); (iii) autoempleo training length also not reported *Frequency:* not reported	*Other features:* There is a large difference between both types of training activities; while the 'escolarizada' offered a general type of education, the 'mixta' offered a specific type of training. It is not clear whether unemployed workers could choose between one of these two activities or if they were just assigned to them by SEE clerks. There is some evidence, however, that the SEE distinguished between workers with and without previous experience, between qualified and unqualified workers, and between temporary unemployed workers and self-employed informal workers. While the SEE decide the type of training activities to be offered as well as the capabilities and abilities that the trainees should developed during their training, the SNE is in charge of providing the funding for these activities. Funding channeled by the SNE covers the workers' scholarships and all the costs associated with the training activities.

Name, setting, dates	Funding, implementation	Main features	Other features
19. PROCAJOVEN (2 sub-programmes) *Country, coverage and dates of operation:* • Panama • approved in 2002; unclear if still ongoing PROCAJOVEN formed one of three components of the PN0125 programme (Programa de Apoyo para un Sistema Panameno de Capacitacion y Empleo). Two sub-programmes of PROCAJOVEN (see opposite) were financed. A third was considered, but did not take place.	*Funded by***:** • Inter-American Development Bank • Government of Panama *Developed and implemented by:* the Ministry of Labor and Workforce Development (MITRADEL)	*Type of TVET programme:*): 1. Insertion Modality (IM): sequential two-phase programme involving classroom-based training (which was comprised of a job readiness course and technical vocational training) followed by a period of on-the-job internship 2. Transition Modality (TM): on-the-job internship (preceded by a short job readiness course) *Target group(s):* IM: low-income unemployed youths aged 18-29 TM: first-time job seekers with complete secondary education aged 16-23 *Delivered by*: training institutions (known as OCAs), who were also in charge of the detection and selection of potential beneficiaries. *Duration:* IM: 120 and 150 hours classroom training (job readiness skills and technical vocational training, respectively), followed by 172 hours of internship TM: 344 hours of internship *Frequency:* no details	*Other features*: OCAs were required to provide job orientation and job placement activities in addition to demand-driven training. Competitive public bids were done periodically, and training institutions were required to provide a letter from a training firm interested in providing internships to ensure relevance of the courses. Not reported if trainees were paid during the internship phase.

Name, setting, dates	Funding, implementation	Main features	Other features
20. ProJoven (Programa de Capacitacion Laboral Juvenil) *Country, coverage and dates of operation:* • Peru (originally intended to be implemented nationwide, but this abandoned: so, for example, in the first four years of the programme youth were trained in three cities) • 1996 - ongoing	*Funded by*: • Inter-American Development Bank • Peruvian Ministry of Labor • other diverse sources of funding (e.g., German government) Over time, the investment in ProJoven has been considerable, but it has never enjoyed long-term funding. *Developed and implemented by:* the International Labour Organization (ILO) proposed the idea and design of ProJoven and provided technical assistance for its implementation; the programme is run by a coordinating unit within the Ministry of Labor	*Type of TVET programme:* sequential two-phase programme involving classroom-based training followed by practical on-the-job internships *Target group(s):* young adults (16-24 years) who are either unemployed or underemployed, have low educational levels, and come from low-income families *Delivered by:* services are provided by private and public training institutions (Entidades de Capacitación – [ECAP]), which compete to obtain funding (from ProJoven) for the courses they are offering. The assignment of government funds to any training institution, public or private, was motivated by the idea that competition would be translated into improved quality. It was stipulated that all training centres must present, as part of their offers, formal agreements with private manufacturing firms that guaranteed paid on-the-job training for each beneficiary. Responsibility for the completion of both phases of training falls solely on the training institutions. A system of conditional payments provided the incentives to train only for those occupations with assured labour demand. *Duration:* 6 months (3 months classroom training, followed by 3 months on-the-job training) *Frequency:* Classroom-based training was roughly five hours per day for three months (300 hours total in formal classes). No details given for on-the-job-training.	*Other features*: ProJoven finances the training and provides a monthly stipend for trainees to cover transportation, meals and medical insurance. Additionally, to encourage their participation, women with children under the age of five receive a double stipend. The courses must follow the job competence approach stated by ProJoven in its manual. A job competence involves skills and aptitudes to solve problems and fabricate products (or perform services) within a particular occupation. Course design must be based on the training needed for a specific trade. Classroom phase completion is not conditional on passing an examination. All participants who complete the classroom instruction, with attendance being the only requirement, are entitled in principle to start the on-the-job training phase. In the second phase, training institutions place trainees into a paid, on-the-job training experience in private manufacturing firms. Training firms must hire trainees under a youth training contract (although not all firms comply with this requirement). The wage paid should not be lower than the minimum wage. The trainees should be under the supervision of a tutor. The assigned internship must consist of activities that complement the training received during the classroom phase. The ECAPs are responsible for the compatibility of course content with the internship. Participants are covered by basic health insurance during this phase of the programme.

9.11   EFFECT SIZES

Sample	Outcome	Follow-up	SMD	SE (SMD)
Acero et al. (2011) - Jóvenes al Bicentenario				
all	employment	6 months	0.2962	0.0774
all	monthly earnings	6 months	0.0376	0.0916
all	weekly hours	6 months	0.087	0.0916
all	hourly wages	6 months	0.0376	0.0916
Aedo and Nuñez (2004) - Proyecto Joven				
females	employment	12 months	0.00462	0.07030
males	employment	12 months	0.06830	0.10149
females	earnings	12 months	0.17556	0.07666
males	earnings	12 months	0.14545	0.07435
Aedo and Pizarro (2004) - Chile Joven				
all	employment	12 months	0.3664	0.067
all	formal employment	12 months	0.372	0.067
all	monthly earnings	12 months	0.5027	0.111
Attanasio et al. (2011) - Jóvenes en Acción				
all	employment	12 months	0.08486	0.04139
all	weekly hours	12 months	0.02668	0.04230
all	earnings	12 months	0.12155	0.04082
all	formal employment	12 months	0.11906	0.04268
females	employment	12 months	0.13758	0.04653
females	weekly hours	12 months	0.10195	0.04973
females	earnings	12 months	0.17274	0.04855
females	formal employment	12 months	0.16379	0.04748
males	employment	12 months	0.02893	0.06454
males	weekly hours	12 months	-0.09075	0.05881
males	earnings	12 months	0.06241	0.05844
males	formal employment	12 months	0.11926	0.05346
Card et al. (2011) - Juventud y Empleo				

Sample	Outcome	Follow-up	SMD	SE (SMD)
all	employment	12 months	0.0329	0.0615
males	employment	12 months	0.0027	0.1005
females	employment	12 months	0.0329	0.0816
all	earnings	12 months	0.1107	0.0553
males	earnings	12 months	0.1273	0.0834
females	earnings	12 months	0.0741	0.0738
all	formal employment	12 months	0.1131	0.0790
all	hourly wage*	12 months	0.1141	0.0744
all	weekly hours*	12 months	0.0097	0.0738
all	earnings*	12 months	0.1319	0.0744

Elías et al. (2004) - Proyecto Joven

Sample	Outcome	Follow-up	SMD	SE (SMD)
all	earnings	12 months	0.02356	0.01823

Espinoza (2010) - ProJoven

Sample	Outcome	Follow-up	SMD	SE (SMD)
all	employment	6 months	0.18563	0.08975
all	employment	12 months	-0.02856	0.15490
all	employment	18 months	-0.06935	0.16836
all	formal employment	6 months	0.35201	0.28247
all	formal employment	12 months	0.30855	0.46241
all	formal employment	18 months	0.04346	0.47051
all	monthly wages	6 months	0.49059	0.12204
all	monthly wages	12 months	0.33298	0.19291
all	monthly wages	18 months	0.37632	0.19654

Hicks et al. (2011) - TVVP

Sample	Outcome	Follow-up	SMD	SE (SMD)
all	employment	6 months	-0.1494	0.204
males	employment	6 months	-0.1074	0.2736
females	employment	6 months	-0.2138	0.3327
all	self-employment	6 months	0.2007	0.1942
males	self-employment	6 months	0.1339	0.2938
females	self-employment	6 months	0.2616	0.2608
all	weekly hours*	6 months	-0.146	0.335

Sample	Outcome	Follow-up	SMD	SE (SMD)
males	weekly hours*	6 months	-0.169	0.414
females	weekly hours*	6 months	-0.257	0.574
all	earnings	6 months	0.0852	0.335
males	earnings	6 months	-0.0716	0.413
females	earnings	6 months	0.466	0.579
all	weekly hours**	6 months	-0.121	0.314
males	weekly hours**	6 months	-0.175	0.464
females	weekly hours**	6 months	-0.154	0.428
all	job search	6 months	-0.194	0.335
males	job search	6 months	-0.314	0.415
females	job search	6 months	-0.154	0.573
all	self-employment profits	6 months	0.116	0.314
males	self-employment profits	6 months	0.344	0.467
females	self-employment profits	6 months	-0.0903	0.427
all	months worked*	6 months	-0.034	0.335
males	months worked*	6 months	-0.0185	0.413
females	months worked*	6 months	-0.0905	0.572

Ibarraran and Rosas-Shady (2006) – PROCAJOVEN (MI)

Sample	Outcome	Follow-up	SMD	SE (SMD)
all	employment	12 months	0.1295	0.113
males	employment	12 months	-0.076	0.182
females	employment	12 months	0.2627	0.146
all	monthly earnings	12 months	0.1266	0.1021
males	monthly earnings	12 months	-0.013	0.1651
females	monthly earnings	12 months	0.2182	0.1301
all	weekly hours	12 months	0.1276	0.1021
males	weekly hours	12 months	-0.0736	0.1651
females	weekly hours	12 months	0.276	0.1303

Ibarraran and Rosas-Shady (2006) – PROCAJOVEN (MT)

Sample	Outcome	Follow-up	SMD	SE (SMD)
all	employment	12 months	0.0715	0.126
males	employment	12 months	-0.291	0.201
females	employment	12 months	0.4024	0.182
all	monthly earnings	12 months	0.0604	0.1134
males	monthly earnings	12 months	-0.064	0.1755

Sample	Outcome	Follow-up	SMD	SE (SMD)
females	monthly earnings	12 months	0.1946	0.149
all	weekly hours	12 months	0.095	0.1134
males	weekly hours	12 months	-0.129	0.1757
females	weekly hours	12 months	0.3184	0.1495
Ibarraran and Rosas-Shady (2006) – PROCAJOVEN (MI and MT)				
all	formal employment*	12 months	0.0739	0.489
Medina and Nuñez (2005) - SENA				
males	monthly earnings	12 months	-0.05937	0.17298
females	monthly earnings	12 months	0.16057	0.30578
Medina and Nuñez (2005) – Public training				
males	monthly earnings	12 months	-0.05270	0.12832
females	monthly earnings	12 months	0.06771	0.20731
Medina and Nuñez (2005) – Private training				
males	monthly earnings	12 months	0.35301	0.20164
females	monthly earnings	12 months	-0.08537	0.13282

*among employed
**among self-employed

9.12 MAIN SUB-GROUP ANALYSES CONDUCTED IN INCLUDED STUDIES

Variable	Employment	Formal employment	earnings	weekly hours
Gender	Aedo & Nunez (2004); Aedo & Pizarro (2004); Alzuá & Brassioli (2006); Attanasio et al. (2011); Card et al. (2011); Chong et al. (2008); Delajara et al. (2006); Díaz & Jaramillo (2006); Elías et al. (2004); Espinoza (2010); Hicks et al. (2011); Ibarraran & Rosas-Shady (2006); Jaramillo et al. (2007); Nopo et al. (2007); van Gameren (2010)	Aedo & Pizarro (2004); Alzuá & Brassioli (2006); Attanasio et al. (2011); Chong et al. (2008); Díaz & Jaramillo (2006); Espinoza (2010); Ibarraran & Rosas-Shady (2006); van Gameren (2010)	Aedo & Nunez (2004); Aedo & Pizarro (2004); Alzuá & Brassioli (2006); Attanasio et al. (2011); Card et al. (2011); Chong & Galdo (2006); Delajara et al. (2006); Díaz & Jaramillo (2006); Elías et al. (2004); Espinoza (2010); Hicks et al. (2011); Ibarraran & Rosas-Shady (2006); Jaramillo et al. (2007); Medina & Nunez (2005); Nopo et al. (2007); van Gameren (2010)	Attanasio et al. (2011); Díaz & Jaramillo (2006); Hicks et al. (2011); Ibarraran & Rosas-Shady (2006); Nopo et al. (2007)
Age[34]	Card et al. (2011); Díaz & Jaramillo (2006)	Díaz & Jaramillo (2006)	Card et al. (2011); Díaz & Jaramillo (2006)	Díaz & Jaramillo (2006)
Education	Card et al. (2011); Delajara et al. (2006); Ibarraran & Rosas-Shady (2006)	Ibarraran & Rosas-Shady (2006)	Card et al. (2011); Delajara et al. (2006); Ibarraran & Rosas-Shady (2006)	Ibarraran & Rosas-Shady (2006)
Work experience	Espinoza (2010)	Espinoza (2010)	Espinoza (2010)	
Region	Card et al. (2011); Delajara et al. (2006); Ibarraran & Rosas-Shady (2006)	Ibarraran & Rosas-Shady (2006)	Card et al. (2011); Delajara et al. (2006); Elías et al. (2004); Ibarraran & Rosas-Shady (2006)	Ibarraran & Rosas-Shady (2006)

---

[34] Only those studies where a sample of young people up to 25 years of age is split into sub-groups (e.g., 16-20 years and 21-25 years) are considered here. A number of other studies disaggregated results by age, but these considered slightly older adults and typically split their sample into groups aged <25 years and 25 years plus.

Variable	Employment	Formal employment	earnings	weekly hours
Income	Espinoza (2010); Jaramillo et al. (2007)	Espinoza (2010)	Espinoza (2010); Jaramillo et al. (2007)	

9.13 MODERATOR ANALYSIS RESULTS: OVERALL PAID EMPLOYMENT

	Q_t	df	Q_w	Q_b	p	Q_{groups}	df	Mean ES	95% CI
Study quality	11.56	7	5.07	6.49	0.0108				
Group 1: Medium						1.93	3	0.06	-0.01, 0.12
Group 2: Low						3.13	3	0.25	0.12, 0.38
Intervention type	7.14	7	5.71	1.43	0.231				
Group 1: Two-phase						4.86	5	0.16	0.04, 0.28
Group 2: Other models						0.849	1	0.01	-0.2, 0.22
Length of follow-up	7.36	8	7.09	0.273	0.601				
Group 1: Medium-term						4.44	5	0.12	0, 0.24
Group 2: Short-term						2.65	2	0.18	0, 0.36
Gender	10.86	11	8.76	2.1	0.147				
Group 1: Female						5.7	5	0.1	0, 0.2
Group 2: Male						3.06	5	0.01	-0.08, 0.09

9.14 MODERATOR ANALYSIS RESULTS: FORMAL EMPLOYMENT

	Q_t	df	Q_w	Q_b	p	Q_{groups}	df	Mean ES	95% CI
Study quality	11.14	4	0.538	10.6	0.0012				
Group 1: Medium						0.174	2	0.12	0.05, 0.19
Group 2: Low						0.365	1	0.37	0.24, 0.5

9.15 MODERATOR ANALYSIS RESULTS: MONTHLY EARNINGS

	Q_t	df	Q_w	Q_b	p	Q_{groups}	df	Mean ES	95% CI
Study quality	5.83	8	5.63	0.204	0.652				
Group 1: Medium						0.0344		0.12	0.05, 0.18
Group 2: Low						5.59		0.15	0.01, 0.3
Intervention type	9.22	8	8.82	0.397	0.529				
Group 1: Two-phase						8.82		0.14	0.04, 0.23
Group 2: Other models						0.0049		0.06	-0.15, 0.27
Length of follow-up	9.97	9	9.78	0.186	0.666				
Group 1: Medium-term						8.4	6	0.14	0.05, 0.24
Group 2: Short-term						1.39	2	0.22	-0.13, 0.58
Gender	11.14	15	9.88	1.26	0.262				
Group 1: Female						5.15	7	0.14	0.08, 0.21
Group 2: Male						4.73	7	0.09	0.02, 0.16

9.16 MODERATOR ANALYSIS RESULTS: SELF-EMPLOYMENT EARNINGS

	Q_t	df	Q_w	Q_b	p	Q_{groups}	df	Mean ES	95% CI
Gender	2.12	3	0.845	1.27	0.259				
Group 1: Female						0.0764	1	0.03	-0.08, 0.13
Group 2: Male						0.768	1	-0.06	-0.18, 0.05

9.17 MODERATOR ANALYSIS RESULTS: WEEKLY HOURS WORKED IN PAID EMPLOYMENT

	Q_t	df	Qw	Q_b	p	Q_{groups}	df	Mean ES	95% CI
Study quality	1.79	5	0.383	1.41	0.234				
Group 1: Medium						0.29	2	0.02	-0.05, 0.09
Group 2: Low						0.0931	2	0.1	-0.01, 0.22
Intervention type	1.80	5	1.73	0.068	0.795				
Group 1: Two-phase						1.27	3	0.04	-0.02, 0.1
Group 2: Other models						0.464	1	0.07	-0.14, 0.28
Length of follow-up	1.80	5	1.69	0.109	0.741				
Group 1: Medium-term						1.24	3	0.04	-0.03, 0.1
Group 2: Short-term						0.45	1	0.07	-0.1, 0.24
Gender	12.96	7	2.86	10.1	0.0015				
Group 1: Female						2.77	3	0.16	0.04, 0.28
Group 2: Male						0.0915	3	-0.09	-0.2, 0.01

9.18 OVERALL PAID EMPLOYMENT: DIRECTIONS OF EFFECT

Study	Intervention		Direction of effect		
			Short term	Med term	Long term
Acero et al. (2011)	Vocational training followed by on-the-job training (6. Jóvenes al Bicentenario)		+**		
Aedo & Nuñez (2004)	Vocational training followed by on-the-job training (2. Proyecto Joven)	5th wave: females		+	
		5th wave: males		+	
Aedo & Pizarro (2004)	Vocational training followed by on-the-job training (5. Chile Joven)			+***	
Alzuá & Brassiolo (2006)	Vocational training followed by on-the-job training (2. Proyecto Joven)	2nd/3rd call		-	+
		5th call		+	
Alzuá et al. (2007)	Vocational training followed by on-the-job training (1. Entra 21)	Arg II-03	+***		
		Arg II-04	+***		
		Arg I-05	+		
		ArgII-05	-		
		Bra Cepro	- ***		
		Bra IH	- *		
Analítica Consultores (2006)	On-the-job training (16. Bécate, and the programme sub-component Capacitación en la Práctica Laboral)	Bécate <20 years		-	
		Practica Laboral		+	
Attanasio et al. (2011)	Vocational training followed by on-the-job training (8. Jóvenes en Acción)			+**	
Benus et al 2001	On-the-job training (majority) (3. Emergency Demobilization and Reintegration Project)				+***
Bidani et al. (2002)	Vocational training followed by on-the-job training (majority) (7. Retraining programmes)	Shenyang			+
		Wuhan			-
Card et al. (2011)	Vocational training followed by on-the-job training (12. Juventud y Empleo)			+	
Chong et al. (2008)	Vocational training followed by on-the-job training (20. ProJoven)			+	
Delajara et al.	Vocational training (18. PROBECAT-	1999	-***		

(2006)	SINAT)	2000	-***			
		2001	-***			
	Vocational training followed by on-the-job training (18. PROBECAT-SINAT)	1999	+***			
		2000	+***			
		2001	+***			
		2002	+***			
		2003	+***			
	On-the-job training (18. PROBECAT-SINAT)	1999	-***			
		2000	-***			
		2001	-***			
		2002	+***			
		2003	+***			
	Self-employment training (18. PROBECAT-SINAT)	1999	-***			
		2000	-***			
		2001	-***			
		2002	-			
		2003	-***			
		2004	-***			
Díaz & Jaramillo (2006)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 1	+***			+***
		Cohort 2	+	+***	+	
		Cohort 4	+	+	+	
		Cohort 6	+***	+	+***	
		Cohort 8	+**	+**		
Dmitrijeva (2009)	Vocational training (15. Occupational training programmes)		+*** [35]			
Elías et al. (2004)	Vocational training followed by on-the-job training (2. Proyecto Joven)	5th call	+			
Espinoza (2010)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 6	+**	-		- *
Hicks et al. (2011)	Technical and vocational education (14. Technical and Vocational Vouchers Program)		**-**			
Ibarraran & Rosas-Shady (2006)	Vocational training followed by on-the-job training (19. Procajoven – Insertion Modality)				+	
	On-the-job training (19. Procajoven – Transition Modality)				+	

---

[35] Same result, regardless of sub-sample

Jaramillo et al. (2007)	Vocational training followed by on-the-job training (20. ProJoven)	Pooled-poorest	+	**_**	+
		Pooled - poor	+**	+**	+**
		Pooled – less poor	**_**	+	+*
López-Acevedo (2003)	Technical education (17. CONALEP)				**_** [36]
Ñopo et al. (2007)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 6	**_**	**_**	+
van Gameren (2010)	On-the-job training (16. Capacitación en la Práctica Laboral – a sub-component of Bécate programme)	Cohort 1	+		
		Cohort 2	+*		
		Cohort 3	+		

*p<0.10; **p<0.5; ***p<0.1

9.19 MONTHLY EARNINGS: DIRECTIONS OF EFFECT

Study	Intervention		Income		
			Short term	**Med term**	**Long term**
Acero et al. (2011)	Vocational training followed by on-the-job training (6. Jóvenes al Bicentenario)		+		
Aedo & Nuñez (2004)	Vocational training followed by on-the-job training (2. Proyecto Joven)	5th wave: females		+**	
		5th wave: males		+**	
Aedo & Pizarro (2004)	Vocational training followed by on-the-job training (5. Chile Joven)			+***	
Alzuá & Brassiolo (2006)	Vocational training followed by on-the-job training (2. Proyecto Joven)	2nd/3rd call		+	+
		5th call		+	
Alzuá et al. (2007)	Vocational training followed by on-the-job training (1. Entra 21)	AR II-03	+***		
		AR II-04	+***		
		AR I-05	+***		
		AR II-05	+***		
		BR Cepro	-		

---

[36] Same result, regardless of cohort: 1990-93; 1991-94; 1992-95; 1993-96; 1994-97

		BR IH	- ***		
Analítica Consultores (2006)	On-the-job training (16. Bécate, and the programme sub-component Capacitación en la Práctica Laboral)	Bécate < 20 years	+		
		Practica Laboral	+		
Attanasio et al. (2011)	Vocational training followed by on-the-job training (8. Jóvenes en Acción)			+***	
Benus et al. (2001)	On-the-job training (majority) (3. Emergency Demobilization and Reintegration Project)				+***
Bidani et al. (2002)	Vocational training followed by on-the-job training (majority) (7. Retraining programmes)	Shenyang			-
		Wuhan			-
Card et al. (2011)	Vocational training followed by on-the-job training (12. Juventud y Empleo)			+**	
Chong & Galdo (2006)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 1	+**	+	+*
		Cohort 2	+*	+**	+*
		Cohort 4	+*	+	+
		Cohort 6	+*	-	+
		Cohort 8	+***	+***	
		Pooled	+	+	+
Chun & Watanabe (2011)[37]	Vocational training followed by on-the-job training (4. Rural Skills Development Project)			+	
Delajara et al. (2006)	Vocational training (18. PROBECAT-SINAT)	2000	+***		
		2001	-		
	Vocational training followed by on-the-job training (18. PROBECAT-SINAT)	2000	+***		
		2001	+		
		2002	+***		
		2003	+***		
	On-the-job training (18. PROBECAT-SINAT)	2000	+		
		2001	+		
		2002	+*		
		2003	+***		
Díaz & Jaramillo (2006)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 1	+***	+***	+***
		Cohort 2	+***	+**	+**
		Cohort 4	+***	+**	+***

---

[37] Overall household per capita income

Study	Program	Subgroup			
		Cohort 6	+***	+***	+***
		Cohort 8	+***	+***	
Elías et al. (2004)	Vocational training followed by on-the-job training (2. Proyecto Joven)	5th call		+	
Espinoza (2010)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 6	+***	+	+
Hicks et al. (2011)	Technical and vocational education (14. Technical and Vocational Vouchers Program)		+		
Ibarraran & Rosas-Shady (2006)	Vocational training followed by on-the-job training (19. Procajoven – Insertion Modality)			+	
	On-the-job training (19. Procajoven – Transition Modality)			+	
Jaramillo et al. (2007)	Vocational training followed by on-the-job training (20. ProJoven)	Pooled-poorest	+**	+**	+**
		Pooled - poor	+**	+**	+**
		Pooled – less poor	+**	+	+**
López-Acevedo (2003)	Technical education (17. CONALEP)				+ [38]
Medina & Nuñez (2005)	Vocational training (9. SENA - Servicio Nacional de Aprendizaje)	Females		+	
		Males		-	
	Vocational training (10. Public sector vocational training)	Females		+	
		Males		-	
	Vocational training (11. Private sector vocational training)	Females		-	
		Males		+*	
Ñopo et al. (2007)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 6	+	+	+
van Gameren (2010)	On-the-job training (16. Capacitación en la Práctica Laboral – a sub-component of Bécate programme)	Cohort 1	+***		
		Cohort 2	+***		
		Cohort 3	+***		

*p<0.10; **p<0.5; ***p<0.1

[38] Non-significant positive result for the initial four cohorts (1990-93; 1991-94; 1992-95; 1993-96), followed by a non-significant negative result for the final cohort (1994-97)

9.20 WEEKLY HOURS: DIRECTIONS OF EFFECT

Study	Intervention		Weekly hours		
			Short term	Med term	Long term
Acero et al. (2011)	Vocational training followed by on-the-job training (6. Jóvenes al Bicentenario)		+		
Attanasio et al. (2011)	Vocational training followed by on-the-job training (8. Jóvenes en Acción)			+	
Card et al. (2011)	Vocational training followed by on-the-job training (12. Juventud y Empleo)			+	
Díaz & Jaramillo (2006)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 1	+***	+**	+***
		Cohort 2	+	+	-
		Cohort 4	+	+	+**
		Cohort 6	+	+**	+
		Cohort 8	+***	+**	
Hicks et al. (2011)	Technical and vocational education (14. Technical and Vocational Vouchers Program)		-		
Ibarraran & Rosas-Shady (2006)	Vocational training followed by on-the-job training (19. Procajoven – Insertion Modality)			+	
	On-the-job training (19. Procajoven – Transition Modality)			+	
López-Acevedo (2003)	Technical education (17. CONALEP)				+ [39]
Mensch et al. (2004)[40]	Vocational training, as part of a health-focused multi-component livelihoods intervention (13. livelihoods intervention)		-		
Ñopo et al. (2007)	Vocational training followed by on-the-job training (20. ProJoven)	Cohort 6	+	+	+

*p<0.10; **p<0.5; ***p<0.1

---

[39] Same result, regardless of cohort: 1990-93; 1991-94; 1992-95; 1993-96; 1994-97

[40] Hours spent in labour market work (defined in the study as paid employment, unpaid employment, vocational training)

9.21 SYNTHESIS INTERPRETATION FRAMEWORK

Evidence grade	Definition	Intervention category
Evidence of effectiveness (positive or negative)	For a given outcome, we can conclude there is evidence of an intervention's effectiveness if all the following conditions are met: • there are meta-analytical results • at least two studies in the meta-analysis were assessed as medium or greater quality • the 95% confidence intervals do not cross the line of 'no effect' • the preponderance of the evidence not included in the meta-analysis (because effect sizes could not be calculated) indicates the same direction of effect (positive or negative)	**What works/does not work** *These programmes demonstrate evidence of beneficial/non-beneficial impacts on participants compared to an alternative.*
Weak evidence of effectiveness (positive or negative)	For a given outcome, we can conclude there is weak evidence of an intervention's effectiveness if: • there are meta-analytical results • at least two studies in the meta-analysis were assessed as medium or greater quality • the 95% confidence intervals cross the line of 'no effect' • the preponderance of the evidence not included in the meta-analysis (because effect sizes could not be calculated) indicates the same direction of effect (positive or negative)	**What is promising/not promising** *These are programmes where the level of certainty from available evidence is too low to support causal or generalisable conclusions, but where there is some empirical basis for predicting that further research could support such conclusions.*
No evidence of effectiveness	Where any of the above categories do not apply.	**What is unknown** These programmes are defined as having unknown effects.

Originally adapted from www.bestevidence.org